Elicitation of Subjective Probability Distributions

Thesis

UNRESTRICTED-

Elicitation of Subjective Probability Distributions

By

Fadlalla Ghaly Hassan Mohamed Elfadaly

BSc., Cairo University. MSc., Cairo University.

A thesis submitted for the Degree of Doctor of Philosophy in Statistics

Department of Mathematics and Statistics

The Open University, UK

April 2012

Date of Submission: 23 April 2012

Date of Award: 3 July 2012

Acknowledgements

Faithful gratitude, sincere thanks, and appreciation are due to Prof. Paul Garthwaite, The Open University, UK, for suggesting the research topic, his supervision, guidance, valuable advice, encouragement, kindness, deep interest and continuous help during the preparation of this thesis. I would like also to thank my co-supervisor Dr. Robin Laney, The Open University, UK, for his advices, directions and his continuous willingness to help. I would like to express my deepest gratitude to my viva examination panel, Prof. Jim Smith, Warwick University, UK, Prof. Kevin McConway and Dr. Karen Vines, The Open University, UK, for their valuable comments, constructive criticism and helpful suggestions.

Also, many thanks are due to the experts whose opinions were quantified in the examples of this thesis. I am very grateful to Dr. Neville Calleja, Ministry of Health, the Elderly and Community Care, Malta, for quantifying his opinion in the obesity misclassification example, and to Dr. Stephen Burnley and Dr. James Warren, The Open University, UK, for quantifying their opinions in the waste collection and transport preferences examples, respectively.

I wish to thank all members of the Statistics Group, The Open University, UK. They all helped me a lot in a very cooperative and supportive research environment that leads to continuous progress and achievement. Special gratitude to the previous PhD students, Dr. Yoseph Araya, Dr. David Jenkinson, Dr. Swarup De, Dr. Youssef Elaziz, Dr. Steffen Unkel, Dr. Angela Noufaily and Dr. Doyo Gragn and also to the current PhD students, Mr. Osvaldo Anacleto-Junior, Mr. Yonas Weldeselassie, Mr. Alexandre Santos, Miss. Sofia Villers. They formed a great academic and social atmosphere for effective work.

True gratitude and deep appreciation are due to Prof. Abdel-Hamid Nigm, Cairo University, Egypt, for suggesting, encouraging, and making fruitful efforts to help me undertake my PhD in the UK. I am also very grateful to Prof. Sanaa El Gayar, Cairo University, Egypt, for her faithful guidance and support during my studies for BSc and MSc degrees. She really guided my first steps on an academic career. I am highly indebted to Dr. Osama Saleh, Cairo University, Egypt, for being such a sincere, supportive and helpful friend.

Heartily and earnest thankfulness to the soul of my late father Mr. Ghaly Elfadaly, my caring mother Mrs. Aziza Belal and my two kind sisters Dr. Hanan and Mrs. Fadila Elfadaly, for their faithful wishes and prayers. I am truly and heartily grateful to my beloved wife, Mrs. Nehal Marghany for her steady love, care and support, and to our son, Master. Malek Elfadaly who lightened up our life with cheer, happiness and innocence.

Abstract

To incorporate expert opinion into a Bayesian analysis, it must be quantified as a prior distribution through an elicitation process that asks the expert meaningful questions whose answers determine this distribution. The aim of this thesis is to fill some gaps in the available techniques for eliciting prior distributions for Generalized Linear Models (GLMs) and multinomial models.

A general method for quantifying opinion about GLMs was developed in Garthwaite and Al-Awadhi (2006). They model the relationship between each continuous predictor and the dependant variable as a piecewise-linear function with a regression coefficient at each of its dividing points. However, coefficients were assumed *a priori* independent if associated with different predictors. We relax this simplifying assumption and propose three new methods for eliciting positive-definite variance-covariance matrices of a multivariate normal prior distribution. In addition, we extend the method of Garthwaite and Dickey (1988) for eliciting an inverse chi-squared conjugate prior for the error variance in normal linear models. We also propose a novel method for eliciting a lognormal prior distribution for the scale parameter of a gamma GLM.

For multinomial models, novel methods are proposed that quantify expert opinion about a conjugate Dirichlet distribution and, additionally, about three more general and flexible prior distributions. First, an elicitation method is proposed for the generalized Dirichlet distribution that was introduced by Connor and Mosimann (1969). Second, a method is developed for eliciting the Gaussian copula as a multivariate distribution with marginal beta priors. Third, a further novel method is constructed that quantifies expert opinion about the most flexible alternate prior, the logistic normal distribution (Aitchison, 1986). This third method is extended to the case of multinomial models with explanatory covariates.

All proposed methods in this thesis are designed to be used with interactive Prior Elicitation Graphical Software (PEGS) that is freely available at http://statistics.open.ac.uk/elicitation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In many situations there is a substantial amount of information that is only recorded in the experience and knowledge of experts. To efficiently use this knowledge as an input to a statistical analysis, the experts must be asked meaningful questions whose answers determine a probability distribution. This process is referred to as *elicitation* and different forms of probability model require different elicitation methods.

Bayesian statistics offers an approach in which data and expert opinion are combined at the modelling stage, yielding probabilities that are a synthesis of the survey data and the expert's opinion. To incorporate expert opinion into a Bayesian analysis, it must be quantified as a prior distribution. This should be accomplished through an elicitation process that asks the expert to perform various assessment tasks. These tasks include questions that the expert is able to comprehend and answer accurately according to her prior knowledge, without needing to know about mathematical and statistical coherence that is required in her assessments.

The elicitation of prior beliefs has been studied extensively in the statistical, psychological, decision and risk analysis literature. Elicitation techniques have been proposed for many probabilistic models including both univariate and multivariate probability distributions. However, achieving accurate elicitation is not an easy task, even for single events or univariate distributions. The difficulty increases for multivariate distributions in which many constraints must be imposed on the expert's assessments to be statistically coherent. Due to this complexity, relatively little literature deals with elicitation techniques for multivariate distributions. O'Hagan *et al.* (2006) argued that the lack of elicitation methods for multivariate models and the lack of user-friendly elicitation software to implement them constitute remarkable deficiencies in the existing elicitation research.

The aim of this thesis is to fill some gaps in the available techniques for eliciting prior distributions for multivariate models. We are mainly interested in eliciting prior distributions for the parameters of Generalized Linear Models (GLMs) and multinomial models. We extend some of the available methods of prior elicitation for GLMs parameters and propose some original novel methods for eliciting different prior distributions for the parameters of multinomial models. All proposed methods in this thesis are designed to be used with interactive graphical software that is written in Java and tailored to the specific requirements of each method. These pieces of software are freely available as Prior Elicitation Graphical Software (PEGS) at http://statistics.open.ac.uk/elicitation.

The elicitation methods for GLMs that are available in the literature focus mainly on logistic regression. A more general elicitation method for quantifying opinion about a logistic regression model was developed in Garthwaite and Al-Awadhi (2006). The method is very general and flexible and can be generalized to GLMs with any link function. The same authors proposed this generalization in an unpublished paper, Garthwaite and Al-Awadhi (2011). In their method, the relationship between each continuous predictor and the dependant variable is modeled as a piecewise-linear function and each of its dividing points is accompanied with a regression coefficient. However, a simplifying assumption was made regarding independence between these coefficients, in the sense that regression coefficients were *a priori* independent if associated with different predictors. One of the main purposes of this thesis is to relax the independence assumption between coefficients of different variables. Then the variance-covariance matrix of the prior distribution is no longer block-diagonal. Different elicitation methods for this more complex case are proposed and it is shown that the resulting variance-covariance matrix is positive-definite. The method of Garthwaite and Al-Awadhi (2006) was designed to be used with the aid of interactive graphical software. It has been used in practical case studies to quantify the opinions of ecologists and medical doctors (Al-Awadhi and Garthwaite (2006); Garthwaite *et al.* (2008)). The software is revised and extended further in this thesis to handle the case of GLM with correlated pairs of covariates.

Available methods of prior elicitation for GLMs all concentrate on the task of quantifying opinion about regression coefficients. For some GLMs, such as logistic regression, this determines the prior distribution completely. But with some other common GLMs, such as the normal linear model and gamma GLMs, prior opinion about an extra parameter must also be quantified in order to obtain a prior distribution for all model parameters. For this reason, we extend the method of Garthwaite and Dickey (1988) for eliciting an inverse chi-squared conjugate prior for the error variance in normal linear models. We also propose a novel method for eliciting the scale parameter of a gamma GLM.

The other multivariate model for which we develop original elicitation methods in this thesis is the multinomial model. Multinomial models consist of items that belong to a number of complementary and mutually exclusive categories. These models arise in many scientific disciplines and industrial applications. The multinomial data are well described using the multinomial distribution, say with parameter vector $\underline{p}$. In Bayesian analysis of multinomial models, an important assessment task is to elicit an informative joint prior distribution for the multinomial probabilities $\underline{p}$. It is well-known that the Dirichlet distribution is a conjugate prior for the parameters of multinomial models. A limited number of attempts have been made to introduce elicitation methods for Dirichlet parameters. However, the Dirichlet distribution has been criticized as insufficiently flexible to represent 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. A k-variate Dirichlet distribution is specified by just k parameters that determine all means, variances and covariances. Dirichlet variates are always negatively correlated, which may not represent prior belief.

Several authors have been interested in constructing new families of sampling distributions to model proportions. Some of these distributions can be used as prior distributions for the probabilities of multinomial models. See, for example, Forster and Skene (1994) and Wong (1998). However, elicitation methods that give these more flexible families as prior distributions for multinomial models have not been proposed. It is tricky, in the case of multinomial models, to elicit assessments that satisfy all the necessary constraints. 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.

Partly because of these difficulties, no doubt, elicitation methods and software for multinomial sampling seem to have been constructed only for modelling opinion by a Dirichlet distribution. In this thesis, we propose novel methods that quantify expert opinion about a Dirichlet distribution and additionally about three more general and flexible prior distributions. First, an elicitation method is proposed for a generalized Dirichlet distribution as a more flexible prior distribution. The generalized Dirichlet distribution, introduced by Connor and Mosimann (1969), has a more general covariance structure than the standard Dirichlet distribution and a larger number of parameters. Second, another method elicits the Gaussian copula as a multivariate prior that expresses the dependence structure between the marginal beta priors of multinomial probabilities using a multivariate normal distribution. Third, a further novel method quantifies expert opinion about the most flexible alternate prior, the logistic normal distribution, Aitchison (1986). With this distribution, the multinomial probabilities are transformed to variables that (by assumption) follow a multivariate normal distribution, using a multivariate form of the logistic transformation. These different elicitation methods are each implemented in interactive graphical software.

The logistic normal distribution has a large number of parameters and gives a prior distribution with a much more flexible dependence structure. Moreover, assuming a logistic normal prior for multinomial models enables us to extend the elicitation method to the case of multinomial models with explanatory covariates. For these models, we proposed a method for eliciting a multivariate normal prior distribution for the regression coefficients based on the multivariate logistic transformation.

The assessment tasks and the task structure implemented in all the proposed methods lead to coherent assessments without the expert having to be conscious of coherence constraints. Using the interactive software, the expert is only required to assess conditional and/or unconditional medians and quartiles for the elements of the probability vector $\underline{p}$. For each of the available prior distributions, 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.

This thesis consists of 10 chapters. After this introductory chapter, Chapter 2 first gives a brief review of the main findings and considerations from psychological literature that should influence the construction of elicitation methods. Then the most relevant methods of eliciting prior distribution for normal linear models and GLMs are reviewed and discussed. Interactive computer software for these purposes is also listed with some of the different applications for which they have been used. In addition, the limited literature of prior elicitation methods for multinomial models is also reviewed, together with its implementing software. We also discuss some recent interactive graphical computer programs that have been reported in the literature for some other problems.

In Chapter 3, the piecewise-linear model of Garthwaite and Al-Awadhi (2006), for eliciting multivariate normal priors for regression coefficients in GLMs is reviewed in detail and the assessment tasks that the expert performs to quantify her opinion are discussed. Also, we describe the software that implements it and detail improvements to the implementation that were made by the author of this thesis.

As mentioned earlier, the elicitation method of Garthwaite and Al-Awadhi (2006) makes the simplifying assumption that the regression coefficients associated with different predictors are independent in the prior distribution. In Chapter 4, we propose 3 new methods for eliciting positive-definite variance-covariance matrices of a multivariate normal prior for regression coefficients that do not require this simplifying assumption. Each method is a trade-off between flexibility and the number of assessments that must be made by the expert.

The first method proposed in Chapter 4 is an extension to the method of Garthwaite and Al-Awadhi (2006). It is the most flexible of the methods but it needs a large number of assessments. The second method requires fewer assessments but assumes a restricted correlation pattern between regression coefficients. The third method first uses one of the other two methods to obtain the correlations between the regression coefficients of two predictors. Then all other correlations are induced through some assessed weights that reflect the magnitude of correlations relative to each other. The expert assesses these weights and then an implementing software presents interactive graphs that help her review and revise assessments to her satisfaction.

In Chapter 5, we introduce two elicitation methods that aim to complete the prior structure of the normal and gamma GLMs. The methods quantify expert opinion about prior distributions for the extra parameters of these models. The first proposed methods elicits a conjugate inverted chi-squared prior distribution for the error variance in normal models. Our proposed method is based on the expert's assessments of medians and conditional medians of the absolute difference between two observed values of the response variable at the same design point. It extends the method of Garthwaite and Dickey (1988) by using more than one data set of hypothetical future samples.

The second proposed method in Chapter 5 is a novel method for eliciting a lognormal prior distribution for the scale parameter of gamma GLMs. Given the mean value of a gamma distributed response variable, the method is based on conditional quartile assessments. It can also be used to quantify an expert's opinion about the prior distribution for the shape parameter of any gamma random variable, if the mean of the distribution has been elicited or is assumed to be known.

Chapter 6 proposes two methods for eliciting a standard Dirichlet prior distribution for multinomial probabilities, using either a marginal or a conditional approach. The main difference between the two proposed approaches is in the assessment tasks that they require. In the marginal approach, the expert assesses unconditional medians and quartiles for each multinomial probability p_i. Then we use these quartiles to obtain a marginal beta distribution for each p_i. The parameters of these marginal betas are reconciled to form a standard Dirichlet distribution. Three different forms of reconciliation are used, each based on least-squares optimizations. For each optimization method, the medians and quartiles of the consequent

Dirichlet distribution are computed and graphically presented to the expert, who chooses which of the Dirichlet distributions best represents her opinion. She is also offered the option to change the medians and quartiles if none of the offered sets is an adequate representation of her opinions.

The other approach proposed in Chapter 6 is the conditional approach. Using this approach, the expert is asked to assess the median and quartiles of the first probability. For each of the remaining probabilities, she assesses conditional medians and quartiles, where the conditions state values for the preceding probabilities that the expert should treat as correct when making her assessments. These conditional assessments are then used to form conditional beta distributions that are also reconciled into a standard Dirichlet distribution.

New elicitation methods for two more general prior distributions for multinomial models are proposed in Chapter 7. The first method uses the same conditional assessments, as obtained in Chapter 6, to elicit a flexible generalized Dirichlet prior, a Connor-Mosimann distribution, through its conditional beta distributions. The flexibility of the generalized Dirichlet distribution means that the elicited parameters of these conditional betas are exactly the same hyperparameters of the elicited generalized Dirichlet prior; no reconciliation is required. This elicitation method and the elicitation methods proposed in Chapter 6 are compared in an example in Section 7.3. In the example, a prominent medical expert in Malta quantified his prior opinions about obesity misclassification in health surveys in Malta.

The second proposed method in Chapter 7 elicits a Gaussian copula prior for the multinomial probabilities. To do this, marginal beta distributions for the multinomial probabilities are obtained from their assessed unconditional medians and quartiles. Then the correlations between the multinomial probabilities are elicited using extra sets of assessments of their conditional medians and quartiles. The proposed Gaussian copula prior assumes that the dependence structure between the multinomial probabilities can be represented by a multivariate normal distribution, where the marginal prior distribution of each multinomial probability is still expressed as a beta distribution. In Section 7.5, the proposed elicitation method and its implementing software are used by an environmental engineering expert to quantify his opinion about the fuel used by waste collection vehicles in the UK.

In Chapter 8, a novel method is proposed for eliciting a logistic normal prior distribution for the probabilities of a multinomial distribution. The method requires conditional medians and quartiles of multinomial probabilities to be assessed. No beta distribution is elicited, instead, a monotonic multivariate logistic transformation is used to transform these assessments into medians and quartiles of a multivariate normal vector. Then a mean vector and a positive-definite covariance matrix of the multivariate normal are determined using the transformed quartiles. The adopted structural method of getting assessments guarantees that the elicited variance-covariance matrix is positive-definite. Chapter 8 also gives an illustrative example in which prior knowledge of a transport expert is quantified to elicit a logistic normal prior distribution for a multinomial model about a transportation problem.

The elicitation method proposed in Chapter 8 for logistic normal priors of multinomial distributions is extended further in Chapter 9 to handle multinomial models that contain explanatory covariates. Our extended method in Chapter 9 elicits a multivariate normal prior distribution for the regression coefficients associated with different covariates in a form of the base-line multinomial logit model. For k categories and m covariates, the model that contains a constant term has exactly $(k-1)(m+1)$ free parameters. In Chapter 9, we show that the same assessment tasks of Chapter 8 can be repeated for each covariate to elicit a mean vector and a positive-definite variance-covariance matrix of a multivariate normal prior distribution for the $(k-1)(m+1)$ regression coefficients.

Concluding comments are given in Chapter 10 where some directions for future research are also considered.

Chapter 2

Literature review

2.1 Introduction

Relatively recent comprehensive reviews of eliciting probability distributions in its theory, methods, techniques, software, applications and case studies are found in Garthwaite *et al.* (2005), O'Hagan *et al.* (2006) and Jenkinson (2007). The aim of this chapter is to review the recent literature on quantifying expert opinion that is most relevant to eliciting prior distributions for Bayesian GLMs and multinomial models. The emphasize here is on the different statistical formulations of elicitation models as well as on the design of the software pieces available in the literature as elicitation tools.

A brief review of some important elicitation topics, ideas and psychological aspects is given in Section 2.2. The important elicitation method of Kadane *et al.* (1980) for normal linear models is reviewed in Section 2.3, where some other elicitation methods for these models are also reviewed briefly. Important and recent elicitation methods and software tools available in the literature for the prior distributions of Bayesian GLMs are reviewed in Section 2.4. However, most of these methods and their accompanying computer programs were devoted to prior elicitation of the Bayesian logistic regression models with anticipated extensions to the more general family of GLMs. Section 2.5 reviews available methods and computer programs for quantifying expert's opinion about priors for multinomial models. As expected, the majority of these methods and tools are quantifying opinions about the simple conjugate prior, the Dirichlet distribution. Some of the recent graphical interactive software that quantifies expert opinion about different problems other than GLMs and multinomial priors are reviewed in Section 2.6.

2.2 Psychological aspects in eliciting opinion

Psychological research on human performance in assessing probabilities dates back to the 1960's. Peterson and Beach (1967) in their paper "Man as an Intuitive Statistician" studied human statistical inference for estimating proportions, means, variances and correlations.

Their results conclude that man can use probability theory and statistics intuitively in performing these inferential tasks. In the same year, Winkler (1967) stated that, in assessing prior distribution for Bayesian analysis, the expert has no 'true' built-in prior distribution that can be elicited. Instead, an elicitation process only "helps to draw out an assessment of a prior distribution from the prior knowledge". This prior distribution is affected by both the assessor and the assessment techniques.

Garthwaite *et al.* (2005) reviewed a body of psychological literature about some of the main mental operations, *heuristics*, that an expert may perform in his mind to give a specific numeric assessment and *biases* that may influence these operations. A recent comprehensive review of psychological research on assessing probabilities including heuristics and biases is given by Kynn (2008). She also provided some guidelines for eliciting expert knowledge based on human biases and inadequacies in assessing probabilities given in the psychological literature. Other useful discussions on psychological aspects in the elicitation context may be found in Hogarth (1975), Wallsten and Budescu (1983) and O'Hagan *et al.* (2006).

The main interest of this thesis is to elicit multivariate probability distributions. Multivariate distributions require more quantities to be elicited than univariate distributions. Beside the usual summaries of each random variable, the dependence structure between all variables must be also assessed. In the rest of this section, we briefly review psychological aspects involved in assessing quantities required for multivariate distributions.

As a measure of central tendency for each random variable, we have decided to elicit its median value from the expert. Experimental work in the literature reveals that people are better at eliciting medians rather than means, especially for skewed distributions. See Garthwaite *et al.* (2005) and references therein. The median value can be assessed through one step of the bisection method, see for example Winkler (1967), Staël von Holstein (1971) and Pratt *et al.* (1995). The expert is asked to determine her median as the value that the random variable is equally likely to be less than or greater than. For more discussion about bisection tasks and their usage, see for example Garthwaite and Dickey (1985), Hora *et al.*

(1992) and Fischer (2001).

To elicit variances, we have chosen to assess the two quartile values of each univariate distribution. By assuming a smooth unimodal distribution, such as the normal or approximate normal distribution, quartiles are transformed to elicit the variances. Quartiles can be easily assessed using the bisection method, which is also called the successive subdivision method, as follows. The upper quartile is assessed by asking the expert to assume that the random variable is above her assessed median value. She is then asked to assess her upper quartile as the value that the random variable is equally likely to be less than or greater than. Similarly, the lower quartile is assessed as the value that divides the range below the median into two equally likely ranges.

The assessed quartiles represent a central 50% credible interval. People can perform the task of assessing credible intervals reasonably well. However, there is a clear tendency for people to be overconfidence in assessing central credible intervals; they tend to give shorter intervals [Garthwaite *et al.* (2005)]. Some other quantiles were found to reduce the degree of overconfidence, such as the 33 and 67 percentiles. O'Hagan (1998) suggested using the central 66% interval, and mentioned that experimental work about different quantile assessments had not revealed any single choice to be the best in all cases. For more details, see Hora *et al.* (1992), Garthwaite and O'Hagan (2000) and Kynn (2005, 2006).

To complete the elicitation process of a multivariate distribution for dependent variables, summaries of dependence structure must be elicited. Typically, determining correlations is the trickiest part in a multivariate elicitation, especially when there are more than two random variables and a variance-covariance matrix must be assessed. Such a matrix must be positive-definite for mathematical coherence. We will make extensive use of the method of Kadane *et al.* (1980) to elicit positive-definite variance-covariance matrices. The method is described in the next section. It relies on assessing conditional medians and quartiles to compute conditional variances and covariances. Conditional quartiles are assessed in a structural way that guarantees positive-definiteness.

Assessing conditional quartiles is not, however, the only way to elicit correlations. Other methods were suggested in Clemen and Reilly (1999) and Clemen *et al.* (2000). These methods include direct assessment of a correlation coefficient, and assessing conditional percentiles or probabilities of one variable given percentiles or probabilities of the other variable, either for one or two items from the population. These assessments were used to calculate Pearson, Spearman and Kendall's τ correlation coefficients. Although Clemen and Reilly (1999) discussed building copula functions as joint distributions, that can be elicited using marginal distributions and elicited correlations, they did not attempt to obtain a positive-definite variance-covariance matrix for multivariate distributions.

In summary, in building our proposed elicitation methods throughout this thesis, we take into account the following considerations. These were mentioned by Kadane and Wolfson (1998) as the points of agreement among most of the statistical literature on how elicitation should be carried out.

1. Expert opinion is the most worthwhile to elicit.
2. Experts should be asked to assess only observable quantities, conditioning only on covariates (which are also observable) or other observable quantities.
3. Experts should not be asked to estimate moments of a distribution (except possibly the first moment); they should be asked to assess quantiles or probabilities of the predictive distribution.
4. Frequent feed-back should be given to the expert during the elicitation process.
5. Experts should be asked to give assessments both unconditionally and conditionally on hypothetical observed data.

2.3 Prior elicitation for normal linear models

Although it was introduced as an elicitation method for the parameters of a normal linear model, the work of Kadane *et al.* (1980) has been an important step towards eliciting prior distributions for GLMs, and even for eliciting many other multivariate distributions. See, for example, Dickey *et al.* (1986) Al-Awadhi and Garthwaite (1998), Garthwaite and Al-Awadhi (2001, 2006). The ideas of Kadane *et al.* (1980) are utilized, modified and implemented extensively throughout this thesis. A detailed review of their elicitation method is given below.

Suppose the normal linear model is given by

$$\underline{Y} = \underline{X}'\underline{\beta} + \varepsilon, \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2), \tag{2.1}$$

where $\underline{X} = (X_1, \quad \cdots, \quad X_r)'$ is a vector of r explanatory variables, and $\underline{\beta} = (\beta_1, \quad \cdots, \quad \beta_r)'$ is the vector of regression coefficients. Kadane *et al.* (1980) introduced an elicitation method for the natural conjugate prior distribution structure of the parameters in model (2.1) as

$$(\underline{\beta}|\sigma^2) \sim \mathrm{N}(\underline{b}, \frac{\sigma^2 R^{-1}}{\delta + r}), \tag{2.2}$$

$$\frac{w\delta}{\sigma^2} \sim \chi^2_\delta. \tag{2.3}$$

The hyperparameters to be elicited are thus a mean vector $\underline{b}$, the two positive scalars δ, w and a positive-definite matrix R. The expert cannot be asked about these quantities directly as they are not observable. Instead, the prior distributions are induced from expert assessments about the response variable Y, which is an observable quantity, at some given values of the explanatory variables. Hence, a number of m realizations $\underline{X}_1, \cdots, \underline{X}_m$ is selected. Kadane and Wolfson (1998) discussed how these design points can be selected efficiently.

At each design point $\underline{X}_i$, $i = 1, \cdots, m$, the expert assesses a median value $y_{i,0.5}$, an upper quartile $y_{i,0.75}$ and a 0.9375 quantile $y_{i,0.9375}$ of the explanatory variable Y_i. The quantile $y_{i,0.9375}$ can be obtained using two bisection iterations above $y_{i,0.75}$. These assessments were used by Kadane *et al.* (1980) to elicit $\underline{b}$ and δ as follows.

To elicit the mean vector $\underline{b}$, the assessed medians were treated as observations of Y, and $\underline{b}$ was elicited as the least-squares estimate

$$\underline{b} = (X'X)^{-1}X'\underline{y}_{0.5}, \tag{2.4}$$

where $\underline{y}_{0.5} = (y_{1,0.5}, y_{2,0.5}, \cdots, y_{m,0.5})'$, and X is the design matrix, which is given by $X = (\underline{X}'_1, \underline{X}'_2, \cdots, \underline{X}'_m)'$.

Under the prior structure in (2.2) and (2.3), the predictive distribution of $(Y|X)$ is a multivariate t distribution with δ degrees of freedom. To elicit δ, Kadane *et al.* (1980) pointed out that the ratios

$$a_i(\underline{X}'_i) = \frac{y_{i,0.9375} - y_{i,0.5}}{y_{i,0.75} - y_{i,0.5}}, \tag{2.5}$$

depend only on δ as a measure of the thickness of the distribution tails. Since the standard normal distribution has the minimum value of this ratio as 2.27, Kadane *et al.* (1980) used a_i^* instead of a_i to elicit δ, where $a_i^*(\underline{X}'_i) = \max\{a_i(\underline{X}'_i), 2.27\}$. Then δ was elicited as the nearest value of degrees of freedom that gives the closest ratio $t_\delta(0.9375)/t_\delta(0.75)$ to

$$\bar{a}_i^* = \frac{\sum_{i=1}^{m} a_i^*(\underline{X}'_i)}{m}. \tag{2.6}$$

We propose a different method for eliciting a degrees of freedom hyperparameter in Chapter 5 of this thesis. Our proposed method is an extension of the approach given by Garthwaite and Dickey (1988), which is described later in Chapter 5.

Although the method of Kadane *et al.* (1980), for eliciting a positive definite matrix R and a value for w, is complicated and requires substantial mathematical notation and details, we review it here because its structural elicitation approach is essential in our proposed methods for eliciting positive-definite matrices throughout this thesis.

The method is based on the properties of the multivariate t distribution. The center and spread of the distribution are defined as follows. For any constant vector $\underline{a}$, and any constant matrix B, if $\underline{Y}$ has a standard multivariate t distribution, then the center of the vector $\underline{Z} = \underline{a} + B\underline{Y}$ is defined as $C(\underline{Z}) = \underline{a}$. The spread of $\underline{Z}$ is defined as $S(\underline{Z}) = BB'$. If $\delta > 1$, then the mean exists and $E(\underline{Z}) = C(\underline{Z})$. If $\delta > 2$, then the variance exists, and is given by $\text{Var}(\underline{Z}) = \frac{\delta}{\delta-2}S(\underline{Z})$. Expert's assessments were used to compute centers and spreads to elicit R and w, as detailed below.

The conditional elicitation structure suggested by Kadane *et al.* (1980), for $i = 2, \cdots, m$, involved assessing conditional medians and upper quartiles of Y_i given sequences of hypothetical values $y_1^0, \cdots, y_{i-1}^0$. The conditions that were imposed on these hypothetical values insured discrepancy between conditional and unconditional centers, in the sense that

$$y_1^0 \neq C(Y_1), \tag{2.7}$$

$$y_i^0 \neq C(Y_i|y_1^0, \cdots, y_{i-1}^0), \quad i = 2, \cdots, m. \tag{2.8}$$

These conditions guarantee the existence of the elicited positive-definiteness matrix R, as will be shown later.

Centers and conditional centers were assessed using medians and conditional medians. For example, $C(Y_1)$ was taken as the unconditional median assessment $y_{1,0.5}$. For $j < i$, $C(Y_i|y_1^0, \cdots, y_j^0)$ were taken as the conditional medians of Y_i given that $Y_1 = y_1^0, \cdots, Y_j = y_j^0$, which are denoted by $(y_{i,0.5}|y_1^0, \cdots, y_j^0)$. Similarly, conditional upper quartiles of Y_i given $y_1^0, \cdots, y_j^0$ are denoted by $(y_{i,0.75}|y_1^0, \cdots, y_j^0)$. Spreads and conditional spreads were computed by dividing the assessed semi-interquartile range by the corresponding semi-interquartile range $t(\delta, 0.75)$ of a standard multivariate t distribution with δ degrees of freedom. This gives

$$S(Y_1) = \left[\frac{y_{1,0.75} - y_{1,0.5}}{t(\delta, 0.75)}\right]^2, \tag{2.9}$$

and, for $i = 1, 2, \cdots, m-1$,

$$S(Y_{i+1}|y_1^0, \cdots, y_i^0) = \left[\frac{(y_{i+1,0.75}|y_1^0, \cdots, y_i^0) - (y_{i+1,0.5}|y_1^0, \cdots, y_i^0)}{t(\delta + i, 0.75)}\right]^2. \tag{2.10}$$

To elicit a positive-definite matrix R, the approach of Kadane *et al.* (1980) is to successively elicit the spread matrices U_i of $(Y_1, \cdots, Y_i)$ in a way that guarantees the positive-definiteness of the final matrix, U_m. The value of U_1 equals $S(Y_1) > 0$ as given in (2.9). Then, supposing that U_i has been estimated as a positive-definite matrix, the aim now is to elicit U_{i+1}, and show it is positive-definite. U_{i+1} is partitioned as

$$U_{i+1} = \begin{bmatrix} U_i & U_i \underline{g}_{i+1} \\ \underline{g}'_{i+1} U_i & S(Y_{i+1}) \end{bmatrix}. \tag{2.11}$$

Conditional median assessments were used to estimate $\underline{g}_{i+1}$ as follows. The partition in (2.11), with the properties of the multivariate t distribution, gives

$$C(Y_{i+1}|y_1^0, \cdots, y_i^0) - C(Y_{i+1}) = (y_1^0 - C(Y_1), \cdots, y_i^0 - C(Y_i))\, \underline{g}_{i+1}. \tag{2.12}$$

Moreover, for $j \leq i$, taking the center of both sides of (2.12) given that $Y_1 = y_1^0, \cdots, Y_j = y_j^0$, gives

$$C(Y_{i+1}|y_1^0, \cdots, y_j^0) - C(Y_{i+1}) = \begin{bmatrix} y_1^0 - C(Y_1) \\ \vdots \\ y_j^0 - C(Y_j) \\ C(Y_{j+1}|y_1^0, \cdots, y_j^0) - C(Y_{j+1}) \\ \vdots \\ C(Y_i|y_1^0, \cdots, y_j^0) - C(Y_i) \end{bmatrix}' \underline{g}_{i+1}. \tag{2.13}$$

Since $j = 1, 2, \cdots, i$, Kadane *et al.* (1980) ended up with a system of i equations of the form

$$\underline{h}_{i+1} = M_{i+1} \underline{g}_{i+1}, \tag{2.14}$$

where

$$h_{i+1} = \begin{bmatrix} C(Y_{i+1}|y_1^0) - C(Y_{i+1}) \\ C(Y_{i+1}|y_1^0, y_2^0) - C(Y_{i+1}) \\ \vdots \\ C(Y_{i+1}|y_1^0, \cdots, y_i^0) - C(Y_{i+1}) \end{bmatrix}, \tag{2.15}$$

and

$$M_{i+1} = \begin{bmatrix} y_1^0 - C(Y_1) & C(Y_2|y_1^0) - C(Y_2) & \cdots & C(Y_i|y_1^0) - C(Y_i) \\ y_1^0 - C(Y_1) & y_2^0 - C(Y_2) & \cdots & C(Y_i|y_1^0, y_2^0) - C(Y_i) \\ \vdots & \vdots & \ddots & \vdots \\ y_1^0 - C(Y_1) & y_2^0 - C(Y_2) & \cdots & y_i^0 - C(Y_i) \end{bmatrix}. \tag{2.16}$$

Multiplying both sides of (2.14) from the left by the matrix

$$Q_{i+1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}, \tag{2.17}$$

gives an upper diagonal system that can be solved for $\underline{g}_{i+1}$ as follows,

$$\underline{g}_{i+1} = \begin{bmatrix} y_1^0 - C(Y_1) & C(Y_2|y_1^0) - C(Y_2) & \cdots & C(Y_i|y_1^0) - C(Y_i) \\ 0 & y_2^0 - C(Y_2|y_1^0) & \cdots & C(Y_i|y_1^0, y_2^0) - C(Y_i|y_1^0) \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & y_i^0 - C(Y_i|y_1^0, \cdots, y_{i-1}^0) \end{bmatrix}^{-1} \underline{\theta}_{i+1}, \tag{2.18}$$

where $\underline{\theta}_{i+1} = Q_{i+1}\underline{h}_{i+1}$.

Under conditions (2.7) and (2.8), the upper diagonal matrix in (2.18) is nonsingular and hence a unique solution for $\underline{g}_{i+1}$ exists. It remains now to elicit the value of the spread $S(Y_{i+1})$ in (2.11). Kadane *et al.* (1980) used the elicited conditional spread, with the properties of the conditional spread of multivariate t distribution, to get a formula for $S(Y_{i+1})$ as follows,

$$S(Y_{i+1}) = \frac{S(Y_{i+1}|y_1^0, \cdots, y_i^0)[1 + i/\delta]}{1 + \delta^{-1}H_i} + \underline{g}'_{i+1}\, U_i\, \underline{g}_{i+1}, \tag{2.19}$$

where

$$H_i = (y_1^0 - C(Y_1), \ \cdots, \ y_i^0 - C(Y_i))\, U_i^{-1}\, (y_1^0 - C(Y_1), \ \cdots, \ y_i^0 - C(Y_i))'.$$

Using Schurr complement, the matrix U_{i+1} as partitioned in (2.11), is positive-definite if and only if U_i is positive-definite and

$$S(Y_{i+1}) - \underline{g}'_{i+1} U_i \underline{g}_{i+1} > 0, \tag{2.20}$$

which is guaranteed from (2.19). Then, using mathematical induction, the final matrix U_m is positive-definite.

To elicit R using U_m, properties of the multivariate t distribution were used to yield the following formula

$$R^{-1} = \frac{\delta + r}{w}(X'X)^{-1}X'(U_m - wI_m)X(X'X)^{-1}, \tag{2.21}$$

where I_m is the identity matrix of order m. See Kadane *et al.* (1980) for details.

The formula requires w to be elicited first. To elicit w, the expert is asked to suppose that two independent observations Y_i and Y_i^* are taken at the same design point $\underline{X} = \underline{X}_i$. Given $y_1^0, \cdots, y_{i-1}^0$, the expert assesses the median of Y_i which is used to estimate $C(Y_i|y_1^0, \cdots, y_{i-1}^0)$. Then the expert is given a hypothetical value y_i^0 for Y_i and is asked to assess the conditional median of Y_i^* given $y_1^0, \cdots, y_i^0$ to be used as an estimate of $C(Y_i^*|y_1^0, \cdots, y_i^0)$. The conditional distribution of the two observations is a bivariate t, and its properties were used to elicit w_i as

$$w_i = [S(Y_i|y_1^0, \cdots, y_{i-1}^0) - K_i]\frac{\delta + i - 1}{\delta + L_i}, \tag{2.22}$$

where

$$K_i = [C(Y_i^*|y_1^0, \cdots, y_i^0) - C(Y_i|y_1^0, \cdots, y_{i-1}^0)]\frac{S(Y_i|y_1^0, \cdots, y_{i-1}^0)}{y_i^0 - C(Y_i|y_1^0, \cdots, y_{i-1}^0)},$$

and

$$L_i = (y_1^0 - \underline{X}_1'\underline{b}, \quad \cdots, \quad y_{i-1}^0 - \underline{X}_{i-1}'\underline{b})\, U_{i-1}^{-1}\, (y_1^0 - \underline{X}_1'\underline{b}, \quad \cdots, \quad y_{i-1}^0 - \underline{X}_{i-1}'\underline{b})'.$$

Different values $w_1, \cdots, w_m$, were then averaged to get a final elicited value w. Our extension of the method suggested by Garthwaite and Dickey (1988) for eliciting w, as proposed in Chapter 5, makes the same assumption of getting two independent observations at the same design point. But we require a median assessment of the difference between the two observations, which is due only to the random variation.

The method of Kadane *et al.* (1980) has been extensively reviewed in the literature. See for example Kadane and Wolfson (1998) and Daneshkhah and Oakley (2010), where two extra examples for its implementation were also discussed. Two drawbacks of the method were mentioned by Garthwaite *et al.* (2005). The assessments it uses are likely to be biased by conservatism as the expert is asked to revise her opinion based on hypothetical data.

Eliciting the spread using the median and upper quartile may not reflect both halves of the distribution, hence masking any asymmetry of expert opinion.

Some other alternate methods for eliciting the parameters of normal linear models are available in the literature. See, for example, Oman (1985), Garthwaite and Dickey (1988, 1992) and Ibrahim and Laud (1994). Oman (1985) used empirical Bayes methods to estimate both δ and R instead of eliciting them from the expert. The method of Garthwaite and Dickey (1988) is similar to that of Kadane *et al.* (1980) in that both of them make use of repeated assessments that are reconciled and utilize a structural set of conditional questions to guarantee the positive-definiteness of the covariance matrix.

However, instead of asking about Y_i, Garthwaite and Dickey (1988) suggested asking the expert about the mean $\bar{Y}_i$ of Y that may be observed in a large number of experiments at the design point $\underline{X}_i$. In this way, the expert's assessments do not include random variation. On the other hand, the design points that are used in Garthwaite and Dickey (1988) are to be selected by the expert. This enabled the method to be extended to the variable selection problem in linear models, see Garthwaite and Dickey (1992). Nevertheless, the method of Kadane *et al.* (1980) is more flexible than that of Garthwaite and Dickey (1988). The latter is not designed to handle categorical explanatory variables nor polynomial regression models that contain interactions between explanatory variables. A more detailed review of normal linear models elicitation can be found in Garthwaite *et al.* (2005) or O'Hagan *et al.* (2006).

2.4 Prior elicitation for GLMs

Starting from the idea that it is more efficient and easier to elicit expert opinion about observable quantities, rather than about parameter values, Bedrick *et al.* (1996) were the first to elicit priors for some arbitrary generalized linear models. Their work switched from normal linear regression elicitation (Kadane *et al.* (1980); Garthwaite and Dickey (1988); Garthwaite and Dickey (1992)) into GLM. Their specification of informative prior distributions for the regression coefficients of a GLM is based on expanding the idea of conditional means priors

(CMP).

The idea of the CMP is that the expert is asked to give his assessment of the mean of potential observations conditional on given values at some carefully chosen points in the explanatory variable space. This information is used to specify a prior distribution at each location point. These priors are conveniently assumed to be independent for the various locations. A prior distribution for the regression coefficient vector is then induced from the CMP.

To clarify this idea, consider for example the binomial GLMs, with n independent observations Y_i, each with a corresponding vector $\underline{X}_i$ of p explanatory variables. Let $N_iY_i|\underline{X}_i \sim$ Binomial(N_i, μ_i), hence $\mu_i = E(Y_i|\underline{X}_i)$. The probability of success μ is related to the vector $\underline{X}$ through a monotonic increasing link function $g(.)$ as

$$g(\mu) = \underline{X}'\underline{\beta}, \tag{2.23}$$

where $\underline{\beta}$ is a p vector of regression coefficients. Common choices for the link function $g(.)$ yield logistic, probit and complementary log-log regressions. The likelihood function for β is given by

$$L(\underline{\beta}) \propto \prod_{i=1}^{n} g^{-1}(\underline{X}_i'\underline{\beta})^{N_iY_i}[1 - g^{-1}(\underline{X}_i'\underline{\beta})]^{N_i(1-Y_i)}. \tag{2.24}$$

Bedrick *et al.* (1996) induced the prior on $\underline{\beta}$ from a CMP on $\tilde{\mu}_i = E(\tilde{Y}_i|\underline{\tilde{X}}_i)$, the success probability for a "potentially observable" response $\tilde{Y}_i$ at the vector $\underline{\tilde{X}}_i$ of explanatory variables. They assume that the p vectors $\underline{\tilde{X}}_i$ are linearly independent and assume that

$$\tilde{\mu}_i \sim \text{beta}(a_{1,i}, a_{2,i}). \tag{2.25}$$

Hence, from independence, the prior on $\underline{\tilde{\mu}}$ is given by

$$\pi(\underline{\tilde{\mu}}) \propto \prod_{i=1}^{p} \tilde{\mu}_i^{a_{1,i}-1}(1 - \tilde{\mu}_i)^{a_{2,i}-1}. \tag{2.26}$$

Under the independence assumption and from (2.23), (2.26), they gave the induced prior on $\underline{\beta}$ as

$$\pi(\underline{\beta}) \propto \prod_{i=1}^{p} g^{-1}(\underline{\tilde{X}}_i'\underline{\beta})^{a_{1,i}-1}[1 - g^{-1}(\underline{\tilde{X}}_i'\underline{\beta})]^{a_{2,i}-1} d[g^{-1}(\underline{\tilde{X}}_i'\underline{\beta})]. \tag{2.27}$$

Although the above example is only valid for binomial GLMs, Bedrick *et al.* (1996) gave generalization and examples where their method is applicable to common GLMs including Poisson and exponential regression. However, for normal and gamma regression models they were only interested in eliciting priors on the regression coefficients $\underline{\beta}$ assuming that the dispersion parameters of these models are known.

The power of this approach as they stated is that "it is much easier to elicit information about success probabilities such as $E(Y|\underline{X}) = \mu$, which are on the same scale as the data, than to attempt the extremely difficult task of eliciting prior knowledge about $\underline{\beta}$."

In their work, the use of data augmentation priors (DAP) was also proposed to induce priors on $\underline{\beta}$. They showed that DAP's are closely related to CMP's and can be induced by particular cases of CMP's. A DAP on $\underline{\beta}$ has the same functional form of the likelihood and can be obtained by specifying "prior observations" and their weights. These prior observations must be taken at specific locations in the predictor space. Hence, a DAP also needs some locations in the predictor space to be specified as in the case of a CMP.

The good choice of the predictor space location should be in the expected range of $\underline{X}$, spread enough so that the corresponding probabilities can be reasonably assumed to be independent and they should also be accepted by the expert. It is straightforward, however, to let the field expert choose these locations. Bedrick *et al.* (1996) noted that the independence in CMP's does not mean that the component of the $\underline{\beta}$ vector will be independent too.

After selecting a proper $\underline{\tilde{X}}_i$, $i = 1, \cdots, p$, to determine the value of $\tilde{Y}_i$ in a DAP, it can be thought of as a typical prior observation associated with $\underline{\tilde{X}}_i$. For example, in binomial GLMs, it can be thought of as a prior estimate of the mean number of successes at $\underline{\tilde{X}}_i$. If the beta prior in (2.25) is reparameterized such that

$$a_{1,i} = \tilde{w}_i \tilde{Y}_i \qquad \text{and} \qquad a_{2,i} = \tilde{w}_i (1 - \tilde{Y}_i), \tag{2.28}$$

then, for the logistic model, the CMP in (2.27) is exactly a DAP since it takes the same functional form of the likelihood in (2.24). The CMP in (2.27) induces a DAP for the logistic model as the logit link function is such that

$$d[g^{-1}(\tilde{\underline{X}}'\underline{\beta})] = g^{-1}(\tilde{\underline{X}}'\underline{\beta})[1 - g^{-1}(\tilde{\underline{X}}'\underline{\beta})]. \tag{2.29}$$

The induced DAP in (2.27), using (2.28) and (2.29), is proportional to a likelihood based on the "prior observations" $(\tilde{Y}_i, \tilde{\underline{X}}_i, \tilde{w}_i, : i = 1, \cdots, p)$. The weight parameter $\tilde{w}_i$ in (2.28) can be interpreted as the prior number of observations associated with $\tilde{Y}_i$. Consequently, large values of $\tilde{w}_i$ reflect more confidence in the prior belief which means that the prior is relatively more informative. However, these extra parameters need to be quantified, the matter which may make the CMP easier to be elicited.

Although the resulting priors are not necessarily members of any specific family of distributions, Bedrick *et al.* (1996) argued that the CMP and DAP priors lead to tractable posteriors for GLMs through importance sampling and Gibbs sampling techniques.

Another approach for eliciting different classes of priors for GLM parameters started with the work of Ibrahim and Laud (1994) for normal linear models. Their work was then extended to prior elicitation and variable selection for logistic regression models by Chen *et al.* (1999). A further extension to GLMs was given by Chen *et al.* (2000), who proposed the class of power priors for GLMs.

The main idea of the above series of papers is that a prior prediction vector $\underline{Y}_0$ can be specified for the response vector $\underline{Y}$, either using historical data or an expert's opinion. A scalar $0 \leq a_0 \leq 1$ needs also to be elicited to quantify the expert's confidence about her best guess $\underline{Y}_0$ relative to the actual data. Hence the scalar a_0 reflects the contribution of the prior information in the posterior relative to the information given by the current experiment. Together with the design matrix X, $\underline{Y}_0$ and a_0 are used to specify an informative prior for regression coefficients.

In the class of power priors, the prior density is raised to the power a_0, which is considered as a precision parameter that controls the heaviness of the tails of the prior distribution. For a random a_0, a beta distribution was assumed by Chen *et al.* (2000) as a prior for a_0. Although the class of power priors cannot be expressed in a closed form, Chen *et al.* (2000) discussed its theoretical properties and propriety together with its required computations.

Different extensions to this class of priors have been proposed in the literature. For example, based on the same ideas, Chen and Ibrahim (2003) proposed a class of conjugate priors for GLMs and discussed its elicitation. Moreover, Chen *et al.* (2003) introduced an informative class of priors for generalized linear mixed models. Extensions to variable selection were suggested by Meyer and Laud (2002), Chen and Dey (2003) and Chen *et al.* (2008).

Garthwaite and Al-Awadhi (2006) developed an elicitation method for piecewise-linear logistic regression. The method is also valid for other GLMs and Garthwaite and Al-Awadhi (2011) extends the idea to GLMs with any link function. They assumed a multivariate normal distribution for the regression coefficients; its parameters can be determined from the expert assessments. One of the main aims of this current thesis is to extend this piecewise-linear elicitation method in the context of GLMs to treat the case of correlated regression coefficients. The method is reviewed in detail in Chapter 3 and the proposed extensions are given in Chapters 4 and 5.

The piecewise-linear elicitation method was designed to be used with the aid of interactive graphical software written for this purpose. Older prototypes of the software were used in practical case studies for threatened species in Garthwaite (1998) and Al-Awadhi and Garthwaite (2006). A more recent version of the software has been written by Jenkinson (2007), this version of the software has been reviewed, modified and extended further in Chapters 3, 4 and 5 of the current thesis.

Another prototype of the interactive graphical software was given by Kynn (2005, 2006) to elicit expert opinion for the Bayesian logistic regression model. The software is called ELICITOR and appeared as an add-on to WinBUGS. Kynn extended the program written by Garthwaite (1998) and rewrote it in a more robust programming language. The software was originally developed as a user friendly tool for quantifying environmental experts' knowledge while studying the presence or absence of endangered species. It adopted the same approach of Al-Awadhi and Garthwaite (2006).

Following Garthwaite (1998), the elicitation scheme adopted in ELICITOR is based on the logistic regression model in which the probability of the presence of an endangered species is represented by a Bernoulli distribution and can be related to a number of environmental variables via a logit function. The expert is asked to give conditional probability assessments at the preferred or optimum site of species presence as the intercept. Then assessments are made at other sub-optimum levels of each other covariate.

The choice of the "optimum" value or level of each covariate to be its intercept, also called the reference value, is made by Garthwaite (1998) and thoroughly justified in Kynn (2006). She discussed that it is psychologically meaningful to the expert to be asked about conditional probabilities given that all or all except one covariate are at their optimum level. In this case, conditioning on all other covariates can be translated in the expert's mind as conditioning on one event where everything is optimal. Kynn mentioned also some ecological concerns that make the optimum point a good selection, a noticeable concern is that the species responses distribution is usually considered to be unimodal. However, in our extensions to the piecewise-linear model, the expert freely chooses the reference level, although she is advised to select the optimum one.

While categorical covariates are related to the probability of presence, or generally of success, through a bar chart in both ELICITOR and the prototype and its extensions, representing continuous covariates is clearly different. ELICITOR does not only assume a piecewise-linear relation between continuous covariates and the presence probability, but it also offers the options of linear and quadratic functions to model this relation. Nevertheless, Kynn (2006) stated that the fully linear form is not realistic and that the quadratic form can be too restrictive. We believe that the piecewise-linear relation is a very general form that can model many other forms as special cases.

The main critical point in the statistical model of ELICITOR is that the regression coefficients are assumed to be independent *a priori*, an assumption that may not be true in many situations. Thus, only univariate normal priors were elicited and no attempt was made to elicit covariances even for the coefficients at the dividing points of the same piecewise-linear curve or at the different levels of each single categorical covariate.

The idea of successive sub-division, also called the bisection method, as a technique to assess the three quartiles from an expert, has been generally accepted as a comparatively easy task for the expert to perform. The prototype software in Garthwaite (1998) and its extensions apply the bisection method to obtain expert's assessments. However, Kynn (2006) has a detailed discussion about available alternatives to assess percentiles, and cites results of studies comparing these methods. But in designing ELICITOR, she decided to use a quite different technique by letting the expert give her two boundaries of a credible interval, then give the probability of this interval. Despite being easy to perform, this method does not seem to be efficiently tested or justified.

Rather than assessing probabilities as numbers, the users of ELICITOR have more interactive visualizations for estimating probabilities. These include a probability wheel, a probability bar and other visualizations to help experts assess probabilities closer to their knowledge. The feedback provided after the assessment process are alternative credible intervals and probability distribution functions for the intercept and categorical variables.

ELICITOR was intended to be extended to encompass other GLMs, with flexible options of the link functions and prior distributions, not only the logistic regression. The software documentation mentioned that this and other extensions were being tested, but we do not know of any version of the software where these extensions have been implemented. For more details on ELICITOR see Kynn (2005); Kynn (2006) and O'Leary *et al.* (2009), although the software and its documentation no longer seem to exist as an open source on the web.

Denham and Mengersen (2007) introduced a method and developed software to elicit expert opinion based on maps and geographic data for logistic regression models. Eliciting information on observable quantities, such as values of the dependant variable at given values of the predictors, (referred to as the predictive procedure) is usually preferred and easier than direct assessment of the regression parameters (structural procedure). However, they argued that each procedure is more convenient for a specific type of experts. For example, they considered two types of ecological experts: the 'physiologist', who has a good understanding of the physical requirements of each species, is more likely to respond well to a structural elicitation. The 'field ecologist', who has more knowledge about the places of existence for each species, may be better at responding to a predictive elicitation. Denham and Mengersen (2007) proposed a new approach that combines both strategies. In their combination approach, the expert may use either method or the two methods simultaneously with each variable, according to his preference and background.

They adopted the usual logistic regression for species modelling,

$$Y_i \sim \text{Binomial}(n_i, \mu_i),$$

with the logit link function $Y_i = g(\mu_i) = \log(\mu_i/(1-\mu_i))$, and $\underline{Y} = \mathbf{X}\underline{\beta}$, where Y_i is the number of observations of a species at site i, and $\mathbf{X}$ is the matrix of explanatory variables. The aim is to quantify the expert's opinion about the prior distribution of $\underline{\beta}$ in the form

$$\underline{\beta} \sim \text{MVN}(\underline{b}, \Sigma).$$

They stated that the methods of Kadane *et al.* (1980) and Garthwaite and Dickey (1988, 1992) can be used in this context to estimate the hyperparameters $\underline{b}$ and Σ by asking the expert to assess some quantile information for the value of $\underline{Y}$ at particular values of $\mathbf{X}$. However, they referred to the difficulty of this predictive elicitation procedure for the 'field ecologists' who may have knowledge about the presence of a specific species at a located site map rather than the explanatory variables affecting this presence.

To help this type of experts, Denham and Mengersen (2007) suggested two alternatives. The method of Kadane *et al.* (1980) can be used, with the expert choosing the design points based on location, without specific reference to explanatory variables. Or, instead, the design points could be selected as in the method of Kadane *et al.* (1980), and then transformed to map locations that are displayed on the map for the expert.

Their proposed combination approach as an elicitation method is not only a hybrid approach that combines both the predictive and structural procedures together, but it also offers the opportunity to use either of the two procedures simultaneously for each single variable. The basis of their method is to use the standard elicitation method with maps as discussed above, to derive a "first pass" elicitation of $\underline{b}$. A structural elicitation procedure is then applied. The latter is implemented by presenting a univariate graph for each of the p explanatory variables. In each graph, they fix all the other $p-1$ variables at their mean or median value, i.e. for the jth variable, $j = 1, \cdots, p$, they display the graph of

$$Y = b_0 + b_j X_j + \sum_{k=1, k\neq j}^{p} b_k \overline{X}_k.$$

These univariate graphs are automatically updated once the expert updates the map by adding new points or editing values. Moreover, the expert can directly manipulate the graphs, which cause the map to automatically change as well. The expert is meant to keep changing the map and/or the graphs until they all represent her prior knowledge. To elicit Σ, The expert is asked to provide a 95% "envelope" around the displayed regression lines by assessing upper and lower 95% quantiles.

To apply this approach, Denham and Mengersen (2007) developed elicitation software under a Geographic Information System (GIS), in which design points were actual location on interactive maps. They listed the benefits of the elicitation procedure using the software with interactive maps over the usual elicitation with paper maps. The new procedure is more flexible, it allows the expert to access information at any point in a convenient manner. The scale dependency of the hard copy maps could be removed by using the feature of zoom in and out. Using the software allows the visualization of the responses and provide feedback to the expert. In which case, the expert can revisit and/or modify any previous assessment on the interactive map.

Denham and Mengersen (2007) implemented their software in two case studies for modeling the median house prices in an Australian city and for predicting the distribution of an endangered species in Queensland. In their first case study, they modelled the median house prices using a piecewise-linear regression to attain flexibility and maintain the simplicity of the linear regression. They chose the dividing knots of the piecewise-linear relations as the 0.33 and 0.66 quantiles of each explanatory variable. Their model takes the form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X'_{i1} + \beta_3 X''_{i1} + \beta_4 X_{i2} + \beta_5 X'_{i2} + \beta_6 X''_{i2},$$

where X_1 is the distance from city center in kilometers and X_2 is the distance from the river in kilometers. For $j = 1, 2$, they defined X'_{ij} and X''_{ij} as

$$X'_{ij} = \begin{cases} X_{ij} - X_{0.33j} & \text{if } X_{ij} > X_{0.33j}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$X''_{ij} = \begin{cases} X_{ij} - X_{0.66j} & \text{if } X_{ij} > X_{0.66j}, \\ 0 & \text{otherwise,} \end{cases}$$

where $X_{0.33j}$ and $X_{0.66j}$ are the 0.33 and 0.66 quantiles of X_j, respectively.

They meant to simplify the Bayesian prior structure of the model compared to that of Kadane *et al.* (1980) or Garthwaite and Dickey (1992), to be of the form

$$\begin{aligned} Y|\mathbf{X}, \underline{\beta}, \sigma^2 &\sim \mathrm{N}(\mathbf{X}'\underline{\beta}, \sigma^2), \\ \underline{\beta} &\sim \mathrm{MVN}(\underline{b}, \Sigma), \\ \sigma^2 &\sim \text{Inverted Gamma}(\nu_0/2, \nu_0 S_0/2), \end{aligned}$$

In this case study, they specify a prior for the regression parameters $\underline{\beta}$. However, it does not seem that they implemented any procedure to elicit the two extra hyperparameters ν_0 and S_0. The results suggested that the experts managed to elicit quantifications of their opinions of the house prices in the city that were consistent with the actual house prices. The priors appeared to be relatively consistent. All participant experts in this case study reported that they preferred the combined approach over the map or the standard approach. Most experts elicited slightly different priors under the different elicitation methods they used.

The second case study in Denham and Mengersen (2007) was devoted to eliciting two experts' opinion about the distribution of the brush-tailed rock-wallaby in Queensland. The explanatory variables were chosen by one of the experts to be X_1: a measure of terrain, X_2: a moisture index, X_3: aspect and X_4: a 4-category variable representing the rock type. They were interested in the following logistic model

$$
\begin{aligned}
Y_i &\sim \text{Bernoulli}(p_i), \\
\text{logit}(p_i) &\sim \text{N}(\mu_i, \sigma^2), \\
\mu_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{2i}^2 + \beta_4 X_{3,i} \\
&\quad + \beta_5 X_{3,i}^2 + \beta_6 X_{4,1i} + \beta_7 X_{4,2i} + \beta_8 X_{4,3i}, \\
\underline{\beta} &\sim \text{MVN}(\underline{b}, \Sigma).
\end{aligned}
$$

They aimed to elicit the multivariate normal prior of $\underline{\beta}$. The experts were allowed to choose the design points. The expert chooses a design point by clicking on a map, then an interactive dialogue pops up giving a plot of a beta distribution of the probability of presence at the selected design point. The given plot has three adjustable points at the median and the 0.05 and 0.95 quantiles. The expert is asked to adjust the three quantiles, or the computed beta parameters, until the presented beta curve is the best representation of the expert's belief about the probability of the specie presence at the selected point. This procedure is repeated for a number of design points.

Once the expert has selected a minimum number of points, a logistic regression model is fitted by the software at each design point. Then the univariate relation between the probability of presence and each of the explanatory variables is presented to the expert in a separate graph, a response curve. Each curve is drawn assuming that the other variables are kept fixed at their means. The categorical variable X_4 is represented by box-plots rather than a curve. The expert can review and modify the design points to get the automatic impact on the response curves. The elicited beta distribution at each design points could be used to elicit the multivariate normal distribution of the regression parameters $\underline{\beta}$ through weighted logistic regression or a simulation based approach, see Denham and Mengersen (2007) for more details. They stated that the priors elicited from the experts were reasonably informative, with corresponding posteriors that are clearly different from those posteriors obtained from a uniform improper prior.

Although the software is specially designed for geographical data elicitation of a logistic regression model, they indicated that the concepts can be generalized to any GLM. However, Denham and Mengersen (2007) wrote the software explicitly for each of the two case studies separately, tailored for the given cases and sets of explanatory variables. In its present form their software is thus limited and cannot be used as a general elicitation tool. Moreover, they used the *R* language to code statistical functions, with Visual Basic and other software for interactive graphs embedded in the GIS system. The latter limits the usability of their software.

Jenkinson (2007) re-wrote the software of Garthwaite and Al-Awadhi (2006) in Java to provide a more transportable and stable version. He gave a detailed description and documentation of both the software and the piecewise-linear theoretical model behind it [Jenkinson (2007), p.215-251]. Further modifications of the theoretical model and the software are given in this current thesis in Chapters 3, 4 and 5.

An important medical application of the GLM elicitation software is given in a case study reported in Garthwaite *et al.* (2008). Aiming to estimate the costs and benefits of current and alternate bowel cancer service in England, a pathway model was developed, whose transition parameters depend on covariates such as patient characteristics. Data to estimate some parameters were lacking and expert opinion was elicited for these parameters, using the indicated software and under the assumption that the quantity of interest was related to covariates by the generalized piecewise-linear model given by Garthwaite and Al-Awadhi (2006). The assessments were used to determine a multivariate normal distribution to represent the expert's opinions about the regression coefficients of that model. One conclusion of this work was that quantifying and using expert judgement can be acceptable in real problems of practical importance, provided that the elicitation is carefully conducted and reported in detail.

A thorough detailed comparison has been conducted by O'Leary *et al.* (2009) for three relatively recent elicitation tools for logistic regression. The comparison included the interactive graphical tool of Kynn (2005) and Kynn (2006), the geographically assisted tool under GIS of Denham and Mengersen (2007) and a third simple direct questionnaire tool with no software. These tools were compared in an elicitation workshop (see O'Leary *et al.* (2009) for more details on the third method). The paper discusses and gives a detailed description for each of the three methods used, showing advantages and disadvantages of each of them. Methods were compared according to their differences in the type of elicitation, the proposed prior model, the elicitation tool and the requirement of a facilitator to help the expert. Prior knowledge of two experts was elicited to model the habitat suitability of the endangered Australian brush-tailed rock-wallaby. The comparison revealed that the elicitation method influences the expert-based prior, to the extent that the three methods gave substantially different priors for one of the experts. Some guidelines were also given for proper selection of the elicitation method. This work of O'Leary *et al.* (2009) is part of a large body of applied research which shows the importance of eliciting expert knowledge when modeling rare event data, see also Kynn (2005); Al-Awadhi and Garthwaite (2006); Low Choy *et al.* (2009) and Low Choy *et al.* (2010).

Although they are interested mainly in designing the elicitation process for ecological applications, Low Choy *et al.* (2009) give a framework for statistical design of expert elicitation processes for informative priors which may be valid for Bayesian modeling in any field. The proposed design consists of six steps, namely, determining the purpose and motivation for using prior information; specifying the relevant expert knowledge available; formulating the statistical model; designing effective and efficient numerical encoding; managing uncertainty; and designing a practical elicitation protocol. Other important stages in the elicitation process may be found in Garthwaite *et al.* (2005), Jenkinson (2007) and Kynn (2008). Low Choy *et al.* (2009) validated these six steps in a detailed discussion and comparison of five case studies, revisiting the principles of successful elicitation in a modern context.

The recent work of James *et al.* (2010) is very interesting and important in the current review for two aspects. First, it introduces and describes a general elicitation tool for quantifying opinion in logistic regression using interactive graphical stand-alone software, called Elicitator. Second, the software is based on a novel statistical methodology to elicit a normal prior distribution for regression parameters.

Their work is an extension to that of Denham and Mengersen (2007) as applied on normal prior elicitation for logistic regression in a geographically-based ecological context. As mentioned before, Denham and Mengersen (2007) did not introduce a general purpose tool; their software was tailored to the requirements of specific case studies. Motivated by that, James *et al.* (2010) developed the Elicitator software as a stand-alone elicitation tool that can be used for a wide range of applications.

Although the Elicitator software is based on the same interface and protocol as its prototype in Denham and Mengersen (2007), the statistical method adopted to transform assessed values into elicited priors is a novel one inspired from the CMP ideas of Bedrick *et al.* (1996). James *et al.* (2010) argued that the CMP is more tractable and more applicable in general compared to the predictive approach used by Kadane *et al.* (1980) and Denham and Mengersen (2007). The novel modification in the Elicitator design to the approach of Bedrick *et al.* (1996) is that it relaxes the assumption that the number of chosen points at which the expert assesses her priors is exactly equal to the number p of explanatory variables in the logistic model. This is the assumption that leads to the induced prior on $\underline{\beta}$ as in (2.27).

Relaxing this assumption allows the number of elicitation points, say k, to exceed the number p of explanatory variables, the situation that is commonly encountered. Although the prior on $\underline{\beta}$ can no longer be induced as in (2.27), James *et al.* (2010) proposed a measurement error model in which elicitation points represent data in a beta regression model. In this sense, increasing the number k of elicitation points will lead to a more accurate prior.

Specifically, they assume a standard logistic regression model with a Bernoulli distribution and a logit link function as used by Bedrick *et al.* (1996). A main criticism is that they assume that the explanatory variables are independent *a priori*, in the sense that independent univariate normal priors were assumed for $\underline{\beta}$, i.e.

$$\beta_j \sim \mathrm{N}(b_j, \sigma_j^2), \qquad j = 1, \cdots, p. \tag{2.30}$$

Although they mentioned the possibility of assuming a multivariate normal prior distribution, no attempt has been made for its implementation in Elicitator.

For $i = 1, \cdots, k$, the expert assesses information about the probability of success μ_i at a geographical site i, selected by the expert, with a known combination of the explanatory variables $X_{1,i}, X_{2,i}, \cdots, X_{p,i}$. For example, the expert may assess information about the probability of presence of a species at a known combination of environmental predictors at site i. Following Bedrick *et al.* (1996), expert's assessments are used to elicit a beta prior on μ_i as in (2.25). However, in situations where $k > p$, a beta prior on μ_i would not help induce the normal prior for $\underline{\beta}$. Instead, James *et al.* (2010) assumed a beta prior on the expert's probability of success, say Z_i, which is different from the actual probability μ_i. As in a measurement error model, μ_i is the conditional expectation of Z_i in the sense that

$$\begin{aligned} \mathrm{logit}(\mu_i) &= \underline{X}_i' \underline{\beta} \\ Z_i | \mu_i &\sim \mathrm{beta}(a_{1,i}, a_{2,i}), \\ E(Z_i | \mu_i) &= \mu_i. \end{aligned} \tag{2.31}$$

James *et al.* (2010) discussed the expert's assessments about Z_i that are required to elicit beta distributions as in (2.31). They argued that the required best estimate of the probability Z_i in the measurement error model is the arithmetic mean, however it is difficult to assess. They were also against the idea of assessing the median, claiming that it needs more effort from the expert to assess. Hence, Elicitator requires the mode of Z_i as its best estimate. Then, following the well-established practice of assessing several quantiles for beta elicitation, Elicitator requires the four bounds of the 50% and 95% credible intervals. Although two assessments are mathematically sufficient for eliciting the two beta parameters, it is better to elicit more assessments and reconcile them, especially for skewed distributions.

A simple numerical procedure is used to elicit beta parameters from the mode and either two or four assessed quantiles.

To elicit the hyperparameters b_j and σ_j^2, $j = 1, \cdots, p$ in (2.30) using the elicited beta parameters $a_{1,i}$ and $a_{2,i}$, $i = 1, \cdots, k$ in (2.31), James *et al.* (2010) proceed as follows. In principal, the beta regression in (2.31) is performed using the expert's data on Z_i and the known values of the explanatory variables to provide the expert-defined estimates of $\underline{\beta}$. However, due to difficulties in implementing any beta regression package in Elicitator, the beta regression problem has been approximated by its discrete version, a binomial regression. An R software package is used to perform the binomial regression, where point estimates $\widehat{\beta}_j$ and their corresponding standard errors s.e.$(\widehat{\beta}_j)$ are obtained. The prior distributions in (2.30) are finally elicited using these estimates as

$$\beta_j \sim \mathrm{N}(\widehat{\beta}_j, \mathrm{s.e.}(\widehat{\beta}_j)^2), \qquad j = 1, \cdots, p. \tag{2.32}$$

Two criticisms of the proposed measurement error model in this context are as follows. First, it adds additional sources of uncertainty, namely, the discrepancy between the expert's probability Z_i and the conceptual probability μ_i. Second, it imposes difficulties in computation and implementation in the software, requiring a binomial regression approximation. However, these criticisms do not seem to be a high price compared to the increased accuracy gained by increasing the number of elicitation points of CMPs. Moreover, the use of beta or binomial regression make it easy to represent standard regression diagnostics to the expert as feedback.

Interactive graphs that are given by Elicitator to the expert as feedback fall in three main groups. The first group includes a box-plot, a pdf curve and some numeric statistics of the elicited beta prior at each site. These are all interactive in the sense that they are automatically modified if the expert changes her assessments of the mode value or the credible interval bounds of the probability of success at each site.

The second group involves the univariate graphs that highlight the main effect of each explanatory variable associated with each of the elicitation sites. These graphs plot the elicited probability against the value of the site predictor with a standard regression fit. The categorical predictors are drawn as bars to emphasize their discrete nature. Various regression diagnostics graphs are given in the third group. These graphs help the expert consider how the estimated prior model elicited from her assessments corresponds to her knowledge overall.

The Elicitator software is written in Java and uses open source libraries. It does not require a commercial GIS, in contrast to the prototype of Denham and Mengersen (2007). All statistical calculations are performed using the R statistical package. Elicitator uses a Java package to communicate with R, without needing to run an actual instance of the R software. This greatly increases the generality and flexibility of Elicitator as a stand-alone tool that can be used by a wide range of experts with different backgrounds.

According to James *et al.* (2010), Elicitator is highly extensible and one of the main extensions they are willing to handle is the ability to implement more GLMs rather than only the logistic regression model. But they did not mention or discuss how this can be done for other distributions and link functions under their proposed model for measurement error.

2.5 Prior elicitation for multinomial models

An early attempt to elicit a Dirichlet prior distribution for multinomial parameters was suggested by Bunn (1978). He argued that the usual fractile assessment procedure that has been used for eliciting beta priors may be difficult and tedious to be applied on their multivariate extensions, the Dirichlet priors, when more conditions and restrictions must be taken into consideration. As will be shown on Chapter 6 of this thesis, developments in computing techniques and tools make it easy to implement fractile procedures in user-friendly software that assess quartiles and elicit Dirichlet priors effectively and interactively.

However, the approach suggested by Bunn (1978) as an alternative to the fractile method for Dirichlet elicitation was the method of 'imaginary results'. He used two versions of this method, namely, the Equivalent Prior Samples (EPS) and the Hypothetical Future Sample

(HFS), to quantify opinions about a Dirichlet prior. Specifically, let $\underline{p} = (p_1, \; p_2, \; \cdots, \; p_k)$, be the vector of multinomial probabilities, with a Dirichlet prior distribution of the form

$$f(\underline{p}) = \frac{\Gamma(\sum_{i=1}^{k} a_i) \prod_{i=1}^{k} p_i^{a_i - 1}}{\prod_{i=1}^{k} \Gamma(a_i)}, \qquad \sum_{i=1}^{k} p_i = 1, \quad a_i > 0. \tag{2.33}$$

It can be shown that the posterior mean of p_i, say $\bar{p}_i$, after sampling N data is given by

$$\bar{p}_i = \frac{a_i + n_i}{N + \sum_{i=1}^{k} a_i}, \tag{2.34}$$

where n_i is the number of items, out of N, that falls in category i.

In the EPS method, the expert is asked to assess a set of prior means $\bar{p}_i^*, i = 1, 2, \cdots, k$. She also assesses the equivalent sample size of her subjective belief that would empirically give this set of probabilities. This sample size gives direct information on $\sum_{i=1}^{k} a_i$. Thus, the prior hyperparameters can be elicited as

$$a_i = \bar{p}_i^* \sum_{i=1}^{k} a_i. \tag{2.35}$$

The main criticism to the usage of the EPS method here is that the expert cannot easily give an assessment for $\sum_{i=1}^{k} a_i$ directly. The assessed value does not necessarily represent her opinion accurately and may contain sources of assessment bias. Therefore, Bunn (1978) proposed the alternate HFS method, in which the expert also assesses the set of prior expectations $\bar{p}_i^*, i = 1, 2, \cdots, k$, but, in addition, she is asked to assess her posterior expectations, say $\bar{p}_i^{**}, i = 1, 2, \cdots, k$, given that a hypothetical future sample of size M has resulted in a number of m_i items in category i, where $1 \leq m_i \leq M$. Hence, the hyperparameters can be elicited, using (2.34) and (2.35), as

$$a_i = \bar{p}_i^* \frac{m_i - M\bar{p}_i^{**}}{\bar{p}_i^{**} - \bar{p}_i^*}. \tag{2.36}$$

The main source of bias in the HFS method is 'conservatism'; the expert tends to revise her probabilistic beliefs from prior expectations to posterior expectations as a result of the new data 'insufficiently' if compared with the revision indicated by Bayes theorem. The strong assumptions of the HFS method, that the expert can be an 'intuitive Bayesian' and can modify her prior beliefs in the light of new data sets, turned out to be poorly satisfied in the case study of Bunn (1978) and other studies mentioned therein. For example, in eliciting beta priors, Winkler (1967) found that the methods of imaginary results gave greater bias than the usual fractile methods.

Another problem with the two methods suggested by Bunn (1978) is that probability means are directly elicited from the expert. We believe that medians are easier to assess and, by using the bisection method, the expert will represent her beliefs more accurately. Although the unit sum of the probability assessments can be directly fulfilled by assessments of means (the means must sum to one), median assessments of these probabilities can be elicited for beta marginal or conditional distributions. Methods for reconciliation of beta elicited distributions into a Dirichlet prior are proposed in Chapter 6.

In the HFS method of Bunn (1978), he did not give any suggestion about the selection of the hypothetical sample. Instead, in a case study, he used an actual sample based on a survey, and called his method an Actual Future Sample (AFS). To investigate the feasibility of this method and its possible biases and subjective inconsistencies, the AFS method was implemented in a case study reported in Bunn (1978). In this study, a publisher quantified his opinion about the expected market attitudes towards a new product. Different possible attitude events were summarized in three categories, for which he assessed their expected prior probabilities as

$$\bar{p}_1^* = 0.20,\ \ \bar{p}_2^* = 0.30,\ \ \bar{p}_3^* = 0.50. \tag{2.37}$$

From his EPS assessment, $\sum_{i=1}^{3} a_i$ was set equal to 10.

Then, a survey of 20 customers revealed that the number of customers in each category were 6, 7, 7, respectively. Based on this survey, the publisher was asked to revise his prior probability expectations. He gave the following posterior expectations

$$\bar{p}_1^{**}(A) = 0.25,\ \ \bar{p}_2^{**}(A) = 0.30,\ \ \bar{p}_3^{**}(A) = 0.45. \tag{2.38}$$

To investigate the conservatism of the publisher, the posterior expected probabilities were computed as in (2.34). Since, $a_1 = 2$, $a_2 = 3$, $a_3 = 5$, the computed posterior expectations given by Bayes theorem are

$$\bar{p}_1^{**}(C) = 0.27,\ \ \bar{p}_2^{**}(C) = 0.33,\ \ \bar{p}_3^{**}(C) = 0.40. \tag{2.39}$$

Comparing the assessed posterior probabilities $\bar{p}_i^{**}(A)$ in (2.38) to the computed ones in $\bar{p}_i^{**}(C)$ in (2.39) reveals the conservatism of the publisher, who did not revise his prior probabilities by as much as Bayes theorem would revise them.

Bunn (1978) discussed the possible reasons of the revealed bias and inconsistency in using the methods of imaginary results for eliciting a Dirichlet prior. He argued that the expert should complete several iterations with these methods to achieve consistent results. However, he did not discuss how this might be done through feedback given to the expert, nor did he suggest any method of reconciliation. These drawbacks of the imaginary results methods suggest that a fractile method is to be preferred, especially in multivariate cases where more inconsistency can be expected.

Using the same idea as the HFS method, and consequently the same forms of equation as in Bunn (1978), Dickey *et al.* (1983) reintroduced the elicitation method with a different case study. The mathematical formulation of the two methods is identical. However, two main differences in the elicitation process can be identified.

In assessing the expected prior probabilities $\bar{p}_i^*, i = 1, 2, \cdots, k$, Bunn (1978) assumed that the expert is coherently aware that these assessed expected probability must sum to one. In contrast, in the work of Dickey *et al.* (1983), the expert was free to assess the expected probabilities without being conscious of any probabilistic constraints. Instead, Dickey *et al.* (1983) suggested normalizing the initial assessed probabilities to get the following normalized set

$$\frac{\bar{p}_i^*}{\sum_{i=1}^{k} \bar{p}_i^*}, \tag{2.40}$$

that is guaranteed to add up to one. We use this simple normalization procedure extensively for our proposed logistic normal prior in Chapters 8 and 9. An important property of a good elicitation method is that the expert is not overly conscious of the mathematical constraints on her assessments. Methods that include normalization and reconciliation procedures are generally better than those that ask the expert to make assessments that meet specified constraints.

The second difference between the elicitation procedure of Bunn (1978) and that of Dickey *et al.* (1983) regards the reconciliation of an expert's assessments. As mentioned before, given the hypothetical sample, one expected posterior probability suffices to elicit the full vector of the Dirichlet hyperparameters. But it is usually better to assess several posterior probabilities and then reconcile the different results. Bunn (1978) regarded discrepancies in results as inconsistency on the part of the expert and suggested asking the expert to resolve inconsistency by doing many iterations of the elicitation process. On the other hand, Dickey *et al.* (1983) suggested reconciling different hyperparameter values by averaging them. They also advised that large discrepancies may indicate that the Dirichlet distribution is not a suitable prior.

The case study in Dickey *et al.* (1983) quantified a social psychologist's opinion about the attitudes of potential jurors in law trials where the death penalty was available. Their attitudes were classified into 4 categories, and the psychologist's assessments of the prior probabilities of the categories were:

$$\bar{p}_1^* = 0.02, \;\; \bar{p}_2^* = 0.08, \;\; \bar{p}_3^* = 0.15, \;\; \bar{p}_4^* = 0.75. \tag{2.41}$$

The psychologist was then told that a hypothetical sample of 200 potential jurors had been distributed between the four categories as 16, 20, 32, 132. Given this information, the expert revised her prior probabilities and gave the following expected posterior probabilities:

$$\bar{p}_1^{**} = 0.05, \;\; \bar{p}_2^{**} = 0.09, \;\; \bar{p}_3^{**} = 0.16, \;\; \bar{p}_4^{**} = 0.70. \tag{2.42}$$

Using each of these values in (2.36) gives an initial value of a_i, which can then be used in (2.35), together with the corresponding prior probability, to get an estimate of $\sum_{i=1}^{4} a_i$. These estimates were averaged in Dickey *et al.* (1983) and gave a value of 140. This gives the final hyperparameter elicited values, again from (2.35), as $a_1 = 2.8$, $a_2 = 11.2$, $a_3 = 21$, $a_4 = 105$.

In contrast to the case study in Bunn (1978), the expert here was not conservative; her posterior probabilities were closer to the relative frequency of the hypothetical data, 0.08, 0.10, 0.16, 0.66, rather than to her prior probabilities. A lack of conservatism is also shown by the small value of $\sum_{i=1}^{4} a_i = 140$, compared to the hypothetical sample size of $N = 200$. Using (2.35), the posterior probabilities in (2.34) can be considered as a weighted average of the prior probabilities and the relative frequency of the hypothetical sample, since

$$\bar{p}_i^{**} = \frac{N}{N + \sum_{i=1}^{k} a_i} \frac{n_i}{N} + \frac{\sum_{i=1}^{k} a_i}{N + \sum_{i=1}^{k} a_i} \bar{p}_i^{*}. \tag{2.43}$$

If the expert assesses $\sum_{i=1}^{k} a_i$ to be less than the hypothetical sample size N, then she gives more weight to the relative frequency of the hypothetical sample. If $\sum_{i=1}^{k} a_i = N$, then the expert has given her prior opinion and the data equal weight. As in Bunn (1978), Dickey *et al.* (1983) did not suggest a way to generate the hypothetical sample.

Another method for eliciting a Dirichlet prior distribution was developed by Chaloner and Duncan (1987) as an extension of their method for eliciting beta distributions (Chaloner and Duncan, 1983). Their approach relied on assessing the mode vector for the predictive distribution, and some probabilities for other vectors around the mode. These assessments were used to elicit a Dirichlet-multinomial predictive distribution that was then used to induce a Dirichlet prior distribution for multinomial sampling. The approach thus differs from other Dirichlet elicitation methods in using mode assessments and in utilizing the predictive distribution rather than the prior distribution.

The predictive distribution of a multinomial likelihood and a conjugate Dirichlet prior is a Dirichlet mixture of multinomial distributions. This distribution is referred to as a Dirichlet-multinomial distribution and its probability mass function takes the form

$$f(x_1, x_2, \cdots, x_k) = \frac{\Gamma(n+1)\Gamma(N)\left[\prod_{i=1}^{k} \Gamma(x_i + a_i)\right]}{\Gamma(n+N)\left[\prod_{i=1}^{k} \Gamma(x_i+1)\right]\left[\prod_{i=1}^{k} \Gamma(a_i)\right]}, \tag{2.44}$$

$$x_i > 0, \quad \sum_{i=1}^{k} x_i = n, \quad a_i > 0, \quad \sum_{i=1}^{k} a_i = N.$$

Chaloner and Duncan (1987) proved that the Dirichlet-multinomial predictive distribution in (2.44) is a unimodal distribution for large values of n. They also gave sufficient conditions under which a vector, with components greater than or equal to one, is the unique mode of the Dirichlet-multinomial distribution. These conditions are mainly related to the probabilities of a set of vectors that are coordinate adjacent to the mode vector. Moreover, the identifiability of the Dirichlet prior distribution from the Dirichlet-multinomial predictive distribution was also proved.

The above results were used in an elicitation scheme that was implemented in a computer program, in Chaloner and Duncan (1987), as follows. The expert specifies a large value of n as the sample size. Then she specifies a mode vector $\underline{m} = (m_1, m_2, \cdots, m_k)$ that satisfies $\sum_{i=1}^{k} m_i = n$ and $m_i \geq 1$. The computer program then uses a multinomial probability vector of $n^{-1}\underline{m}$ to compute probabilities at some points that are component adjacent to the mode vector. These probabilities are presented to the expert and she is given the option of changing them if they do not represent her opinion adequately. The modified set of probabilities, together with the mode vector $\underline{m}$, determine an initial value for the parameter vector $\underline{a}$ of the Dirichlet-multinomial predictive distribution. This is also taken as the elicited parameter vector for the Dirichlet prior distribution.

The elicitation scheme of Chaloner and Duncan (1987) does not stop there. Instead, they chose to use the initially elicited vector $\underline{a}$ to compute the Dirichlet-multinomial probabilities at the same points where assessments had been elicited and give them as feedback to the expert, offering her the possibility of revising them to more closely represent her opinion. Moreover, Chaloner and Duncan (1987) believed that more replications were required. Therefore, the expert was to repeat the whole process again for a number of S different sample sizes n_1, n_2, $\cdots$, n_S. The resulting parameter vectors $\underline{a}_1$, $\underline{a}_2$, $\cdots$, $\underline{a}_S$ were to be reconciled to give one final elicited vector of parameters. Chaloner and Duncan (1987) argued that it might be "dangerous" to use an automatic specific reconciliation method, instead, they recommended that the expert should examine the inconsistencies and "reconcile them introspectively".

However, the method requires direct assessment of the sample size n, this might lead to improper representation of an expert's opinion and incur more bias [Bunn (1978)]. On the other hand, Chaloner and Duncan (1987) did not mention how large the assessed value n should be, neither did they discuss whether the expert should keep in mind the constraint $\sum_{i=1}^{k} m_i = n$, on the mode vector $\underline{m}$, or whether it may be corrected by the program if necessary. Nevertheless, it seems from their reluctance to apply any reconciliation that they preferred to leave it to the expert to make sure that the constraints were satisfied. Repeating the elicitation process for S different sample sizes may constitute an extra burden on the expert, especially if she is responsible for the final reconciliation. Unfortunately, the computer program implementing their method does not seem to be available for reviewing and testing.

Instead of using means or modes, van Dorp and Mazzuchi (2000, 2003, 2004) introduced a numerical algorithm and software to specify the parameters of a beta distribution and its Dirichlet extensions using quantiles. The motivation for their work was to quantify expert opinion as beta and Dirichlet distributions for subjective Bayesian analyses. They favored assessing quantiles rather than means or modes, as betting strategies can be used by the expert to make their assessments. They started by solving for the two parameters of a beta distribution using two quantiles, as follows.

First, to ease the generalization to Dirichlet extensions, the beta distribution with two parameters a and b was reparameterized in terms of a location parameter $\mu = a/(a+b)$, and a shape parameter $N = a+b$. Given the values of any two quantiles, say L and U, $L < U$, the two parameters μ and N can be obtained, although solving for these two parameters involves the use of the incomplete beta function, so that no closed form solution can be obtained. van Dorp and Mazzuchi (2000) utilized the limiting forms of a beta distribution as N tends to 0 and ∞ to prove the existence of at least one solution for the beta parameters in terms of any two quantiles.

They gave a numerical algorithm to determine the beta parameters using a bisection method as a numerical search procedure. If multiple solutions were found, the algorithm selects the solution with the lowest value of N, i.e. with the highest level of uncertainty. The algorithm was implemented in software called BETA-CALCULATOR that inputs any two beta quantiles to output the corresponding values of the beta parameters.

To extend the numerical algorithm to Dirichlet parameters, van Dorp and Mazzuchi (2003, 2004) used quantiles that were assessed through direct specification of marginal beta distributions. A Dirichlet distribution as given in (2.33) was also reparameterized in terms of its mean values $\mu_i = a_i/N$, as location parameters, and $N = \sum_{i=1}^{k} a_i$ as a shape parameter. The extended algorithm was designed to use two quantiles for one of the Dirichlet variates, say L_i and U_i, $L_i < U_i$, for the ith variate, and just one quantile for each of the remaining variates, say Q_j, $j \neq i$. Hence, the number k of quantile equations that they had is exactly equal to the number of required parameters.

Following similar lines to their arguments for the beta distribution, van Dorp and Mazzuchi (2003, 2004) showed theoretically that at least one solution of the resulting system of equations always exists. The two quartiles L_i and U_i were first used to elicit the marginal distribution of the ith Dirichlet variate as $X_i \sim \text{beta}(\mu_i, N)$. The value of N is then used with the quantiles Q_j to elicit the remaining beta marginal distributions as $X_j \sim \text{beta}(\mu_j, N)$, $j \neq i$. If more than one solution exists, they decided to choose the solution with the smallest N, which is again the solution with maximum Dirichlet variance, hence giving the highest level of uncertainty. In addition to the Dirichlet distribution, they also gave another numerical algorithm for the ordered Dirichlet distribution, which differs from the Dirichlet in the domain of its variates, see Wilks (1962).

A criticism of the algorithm regards the selection of the Dirichlet variate for which two quantiles are assessed. No comment regarding the selection of this special variate was given in the published paper. The importance of its choice is that it determines the value of N for all other variates and hence determines the variances of the Dirichlet distribution. If substantial bias is made in assessing these two quartiles, all elicited parameters will be highly affected as a result.

In addition, to get a better representation of an expert's opinion in the elicitation context, it is better to use over-fitting (Kadane and Wolfson (1998)). We believe that it is preferable to assess more quantiles than the minimum necessary and then apply a reconciliation technique to estimate parameters. The expert may then be given feedback and questioned as to whether the feedback corresponds to her opinion, with re-assessment made when necessary.

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 defined as a function that represents a multivariate cumulative distribution in terms of one-dimensional marginal cumulative distribution functions. Hence, it joins marginal distributions into a multivariate distribution that has those marginals. The importance of the copula function is due to Sklar's Theorem, 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) and Nelsen (1999).

The use of copula functions to elicit multivariate distributions has been considered in the literature, see Jouini and Clemen (1996), Clemen and Reilly (1999) and Kurowicka and Cooke (2006), among others. 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 and Frank's families of copulae to aggregate multiple experts' opinions about a random quantity. However, the simplest and most intuitive family of copulae is the inversion copula [Nelsen (1999)], of the form

$$C[G_1(x_1), \cdots, G_k(x_k)] = F_{(1,\cdots,k)}\left\{F_1^{-1}[G_1(x_1)], \cdots, F_k^{-1}[G_k(x_k)]\right\}, \tag{2.45}$$

where G_i are the known marginal distribution functions, $F_{(1,\cdots,k)}$ is the assumed multivariate distribution function and its marginals are F_i. Hence, the marginal functions G_i's are coupled through $F_{(1,\cdots,k)}$ into a new multivariate distribution given by the copula function C.

The distribution $F_{(1,\cdots,k)}$ is usually selected as a multivariate normal distribution, which gives a Gaussian copula [Clemen and Reilly (1999)]. It has also been taken as a multivariate t distribution, [Demarta and McNeil (2005)], or even as a Dirichlet distribution [Lewandowski (2008)]. The Gaussian copula function is given by

$$C[G_1(x_1), \cdots, G_k(x_k)] = \Phi_{k,R}\left\{\Phi^{-1}[G_1(x_1)], \cdots, \Phi^{-1}[G_k(x_k)]\right\}. \tag{2.46}$$

where $\Phi_{k,R}$ is the cdf of a k-variate normal distribution with zero means, unit variances, and a correlation matrix R that reflects the desired dependence structure. Φ is the standard univariate normal cdf.

For eliciting a multivariate distribution, the Gaussian copula is the most appealing, see Clemen and Reilly (1999), as it is parameterized by the correlation matrix R of the multivariate normal distribution; hence it only requires pairwise correlations among the variables. To elicit the Gaussian copula, any assessed positive-definite correlation matrix R can be used together with the elicited marginal distributions $G_1(x_1), \cdots, G_k(x_k)$. As with any other inversion copula, any univariate distributions are allowed as marginal distributions G_i's in the Gaussian copula.

To elicit R, Clemen and Reilly (1999) suggested 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:

$$r_{i,j} = 2\sin(\pi\rho_{i,j}/6), \qquad \text{or} \qquad r_{i,j} = \sin(\pi\tau_{i,j}/2). \tag{2.47}$$

Then the product-moment correlation matrix R is formed from the elements $r_{i,j}$.

Clemen and Reilly (1999) suggested that only rank-order correlations should be elicited, not product-moment Pearson correlation, as the latter cannot necessarily be transformed through the function $\Phi^{-1}[G_i(.)]$ - while rank-order correlations transform regardless of the choice of the marginal distribution function $G_i(.)$. To elicit these correlations, Clemen and Reilly (1999) mentioned three methods that can be used either separately or together. The first method involved the direct assessment of the correlation coefficient. Although people are not good at such direct assessment (Kadane and Wolfson, 1998), experimental evidence in Clemen *et al.* (2000) suggested that it can be a reasonable approach. The other two methods were based on assessed conditional probabilities or conditional quantiles that can be used to compute Kendall's τ or Spearman's ρ correlation coefficients, respectively.

The method proposed by Clemen and Reilly (1999) for eliciting a correlation matrix is not guaranteed to yield a positive-definite matrix. They cited two other studies in which dependence measures were assessed in a hierarchical way using dependence trees that require a fewer number of assessments. These studies use entropy maximization to guarantee the positive-definiteness of the resulting correlation matrix. However, Clemen and Reilly (1999) criticized this approach for the relatively constrained nature of its dependence structure modelling. Instead, they suggested that the expert should be asked to revise her assessments if the resulting correlation matrix is not positive-definite. For large problems with many variables, this revision method would generally be very tedious and confusing.

In Chapter 7, we propose a method for eliciting a Gaussian copula function, as a prior distribution for multinomial models. Our approach overcomes two problems of the method of Clemen and Reilly (1999) simultaneously. First, we transform the assessed conditional quartiles of X_i and X_j, through $\Phi^{-1}[G_i(.)]$, then product-moment correlations can be computed on the normal scale with no need for the 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.

Copula functions were used extensively in the literature for building multivariate distributions based on known marginals. This includes, of course, building joint prior distributions for Bayesian analysis using copulae. For example, Yi and Bier (1998) utilized some copula families to construct a joint prior distribution that reflects inter-system dependencies between accident precursors in a Bayesian study to estimate accident frequencies. A Gaussian copula has not been widely used in the literature as a prior distribution for multinomial models.

However, the need for a flexible joint prior distribution that effectively combines the marginal beta prior distributions of multinomial probabilities makes the Gaussian copula an attractive choice as it gives a more general dependence structure than the usual Dirichlet distribution. An applied Bayesian study by Palomo *et al.* (2007) used a Gaussian copula to model external risk in project management. In one of their adopted scenarios, they assumed that any of k potential disruptive events might occur, one at a time, according to a multinomial distribution. The multinomial probabilities were assigned beta marginals, and a Gaussian copula function was used as a multivariate distribution to parameterize the dependence structure between these probabilities.

2.6 Other general graphical elicitation software

This section reviews other interactive graphical elicitation software that has been reported in the literature. Software projects that are reviewed below cover general elicitation problems apart from those for GLMs and multinomial models. These have already been reviewed in Sections 2.4 and 2.5.

Chaloner *et al.* (1993) aimed to quantify experts' opinion in the form of a prior distribution about regression coefficients in a proportional hazards regression model. In a clinical trial, prior distributions from five AIDS experts were elicited. To compare two treatments with a placebo, experts were asked to elicit the joint and marginal distributions of the survival probability under each treatment. This could be done by assessing some probabilities and quantiles to elicit a joint extreme value prior distribution for the proportional hazards model parameters.

For this purpose, they developed an interactive computer program that uses interactive graphs to elicit experts' opinion and give them feedback. The curves of the two marginal distribution and the contour representing the joint distribution were presented to the experts. This feedback was given in the form of dynamic graphical displays of probability distributions that can be adjusted freehand.

Some of the main "lessons" learned about this elicitation process, as stated by Chaloner *et al.* (1993), can be summarized as follows. They stressed the importance of the dynamic graphical displays in helping experts to visualize probability distributions and in giving useful instant feedback. They also noted that it is necessary to have a clear well-defined outline and explanation of the questions that will be addressed to the expert. In cases where an expert had to assess her best guess of a specific probability, they wanted her also to report her uncertainty about it. In assessing approximate bounds, experts found extreme percentiles easier to think about than quartiles. However, there is substantial empirical evidence that people are poor at assessing extreme quantiles [e.g. Winkler (1967); Hora *et al.* (1992)] and we believe that quartiles provide a more faithful representation of an expert's opinion, especially if they are assessed using the bisection method.

A comparatively simple elicitation computer program was developed by Kadane *et al.* (2006) for the generalized Poisson distribution. In their paper, they explored the properties of the Conway-Maxwell-Poisson (COM-Poisson) distribution, in particular, the conjugate family of prior distributions associated with it. A computer application has been created to elicit the hyperparameters of the conjugate prior distribution of the COM-Poisson parameters.

The COM-Poisson distribution is a two parameter generalization of the Poisson distribution that allows for over- and under-dispersion. It has the following probability function

$$\Pr\{X = x|\lambda, \nu\} = \frac{\lambda^x}{(x!)^\nu} \cdot \frac{1}{Z(\lambda, \nu)}, \quad x = 0, 1, 2, \ldots,$$

where

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}.$$

The distribution indicates over-dispersion (under-dispersion) if ν is less (greater) than 1. It is the usual Poisson distribution if $\nu = 1$. Since the COM-Poisson distribution is a member of the exponential family, it has a conjugate prior of the form

$$h(\lambda, \nu) = \lambda^{a-1} e^{-\nu b} Z(\lambda, \nu)^{-c} k(a, b, c), \qquad \lambda > 0, \nu \geq 0,$$

where $k(a, b, c)$ is the integration constant.

The computer program, available at http://www.stat.cmu.edu/COM-Poisson/, is designed to elicit the values of the hyperparameters a, b and c from the field expert. It computes and plots the histogram of the predictive distribution at allowable selected values of a, b and c. Specifically,

$$\Pr\{X = x|a, b, c\} = k(a, b, c) \int_0^{\infty} \int_0^{\infty} \lambda^{a+x-1} e^{-\nu(b+\log(x!))} Z(\lambda, \nu)^{-(c+1)} d\lambda d\nu.$$

Kadane *et al.* (2006) pointed that it may be difficult for the expert to give meaningful values for the hyperparameters a, b and c, since the distribution is likely to be new to her. They assumed that the expert may have some knowledge about $\Pr\{X = x\}$. Thus, the program plots the predictive distribution as feedback to the expert. She can type in or modify the values of a, b and c using sliders and see the direct impact on the predictive histogram.

However, it does not seem that the expert will be able to adjust three values simultaneously to assess a histogram that represents her prior belief. Also, some combinations are not allowed because of mathematical incoherence, and some others need large numbers of iterations to produce the histogram. A lot of adjustment may be needed before the expert is happy with the histogram, since no specific combination of the hyperparameter values is known in advance for any intended appearance of the histogram.

2.7 Concluding comments

In this chapter, we have reviewed some of the relevant research work on eliciting prior distributions for the Bayesian analysis of GLMs and multinomial models. We have also discussed and reviewed the main psychological aspects that are usually involved in making the assessments to elicit these prior distributions. In addition, we commented on some of the recent interactive graphical software that have been reported in the literature for implementing and facilitating the elicitation processes in some other statistical problems. However, this review has been restricted to work that is directly relevant to the elicitation methods proposed in this thesis. There is a huge body of research that handles elicitation problems and techniques in general. As noted earlier, psychological concerns and recommendations for efficient elicitation techniques will be taken into consideration while developing the elicitation methods proposed in this thesis. Available elicitation techniques and computer software will feed into the methods developed in the next chapters and will help in building the software to implement these proposed methods.

Chapter 3

The piecewise-linear model for prior elicitation in GLMs

3.1 Introduction

Generalized linear models (GLMs) constitute a natural generalization of classical linear models, where the linear predictor part is linked to the mean of the dependent variable through some link function. The distribution of the dependent variable is not necessarily assumed to be normal. The model is determined by a combination of the link function and the family of distributions to which the dependent variable belongs (see McCullagh and Nelder (1989) for an introduction to GLMs). Being very common in both frequentist and Bayesian data analysis, GLMs have attracted much research.

An important task in the Bayesian analysis of GLMs is to specify an informative prior distribution for model parameters. Suitable elicitation methods play a key role in this task of representing expert knowledge as a prior distribution (see, for example, Bedrick *et al.* (1996) and O'Leary *et al.* (2009)).

A method of quantifying opinion about a logistic regression model was developed by Garthwaite and Al-Awadhi (2006). They mentioned that the method is very flexible and can be generalized to GLMs with any link function, not just the logistic link. This generalization has been introduced by the same authors in an unpublished paper, Garthwaite and Al-Awadhi (2011). Their method has been used to quantify the opinions of ecologists (Al-Awadhi and Garthwaite (2006)) and medical doctors (Jenkinson (2007); Garthwaite *et al.* (2008)). However, the method makes simplifying assumptions regarding independence between the regression coefficients. One purpose of the current thesis is to extend the elicitation method so that these assumptions are unnecessary. Different methods for this extension are proposed in Chapter 4. This will significantly increase the range of situations where the method is useful.

The original method for logistic regression was developed and implemented in user-friendly interactive software. The software was re-written in Java by Jenkinson (2007) who also extended it to elicit expert opinion about some other GLMs. It has been modified and extended further by the author of the current thesis.

The software is interactive, requiring the expert to either type in assessments or plot points on graphs and bar-charts using interactive graphics. An executable stand-alone version of the current software is available as a java executable (jar) file and a Windows executable file (with *.exe* extension). The stand-alone versions together with the user manual and the source code are freely available as Prior Elicitation Graphical Software for Generalized Linear Models (PEGS-GLM) at http://statistics.open.ac.uk/elicitation. The software is aimed to be executable on any machine regardless of its operating systems and without need of any other software packages.

The current modified version of the software is more flexible in determining the options available for the user, especially for data input and results output. Some important modifications involve broadening the scope of available models and the range of the link functions, and giving the user many suggestions, help notes and video clips, questions, warning messages and directions aimed at making the software more interactive and easy to use for non-statistical experts. Useful feedback has also been added.

In this current chapter, the piecewise-linear model of Garthwaite and Al-Awadhi (2006) is reviewed, and we describe the elicitation method they propose together with the above modifications. The assessment tasks that the expert performs quantify her opinion about the regression coefficients as a multivariate normal prior distribution. The largest extension to the current version of the software is a new section for assessing expert knowledge about correlated covariates. This will be introduced in Chapter 4. Important options have been added to the method that quantify opinion about the extra parameter in GLMs that involve gamma and normal distributions. The theoretical derivation and implementation of these options are proposed in Chapter 5.

3.2 The elicitation method for piecewise-linear models (GA method)

For quantifying expert's opinion about GLMs, Garthwaite and Al-Awadhi (2011) proposed a method to elicit expert opinion about the prior distribution of regression coefficients and its hyperparameters. As mentioned before, the method, which will be referred to here as GA, is a generalization of the same authors' piecewise-linear model that they used for quantifying opinion for logistic regression (Garthwaite and Al-Awadhi (2006)).

In their work, the relationship between each continuous predictor variable and the link function (assuming all other variables are held fixed) was modeled as a piecewise-linear function. Figure 3.1 illustrates a piecewise-linear relationship between the quantity of interest Y, and a continuous covariate "Weight"; the relationship correspondence to a sequence of straight lines that form a continuous line. The endpoints of the straight lines are refereed to as knots.

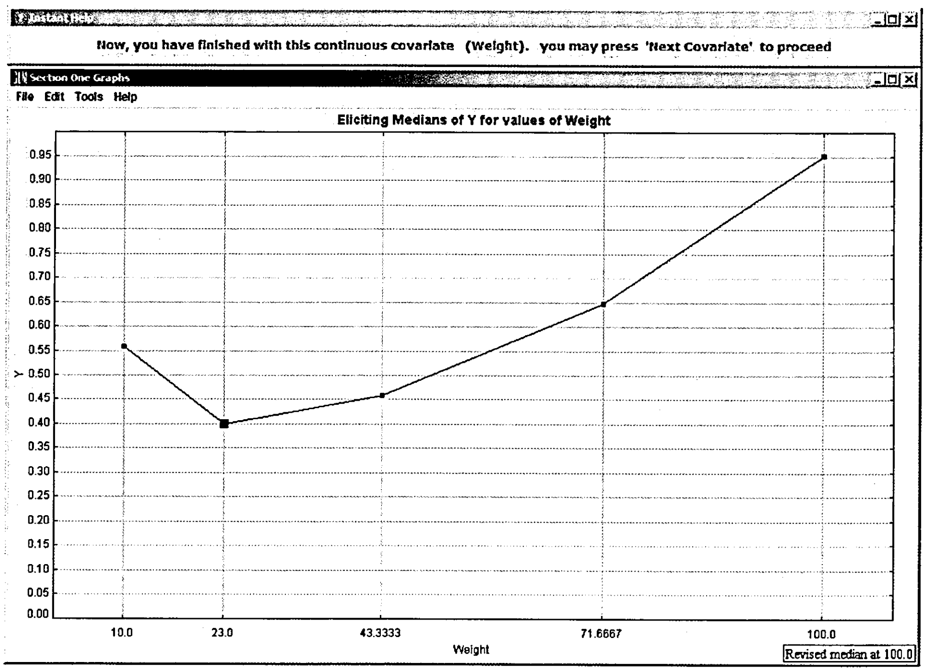


Figure 3.1: A piecewise-linear relationship given by median assessments

If the predictor variable is a categorical covariate, it is referred to as a factor. Its relationship with Y corresponds to a bar chart as in Figure 3.2, where the factor takes four levels:

Very large, Large, Normal and Small.

The aim of the elicitation process is to quantify opinion about the slopes of the straight lines (for continuous variables) and the heights of the bars (for factors). In the GA method, a multivariate normal distribution was used to represent prior knowledge about the regression coefficients. These coefficients were allowed to be dependant if associated with a single variable. A detailed discussion of their model is given next.

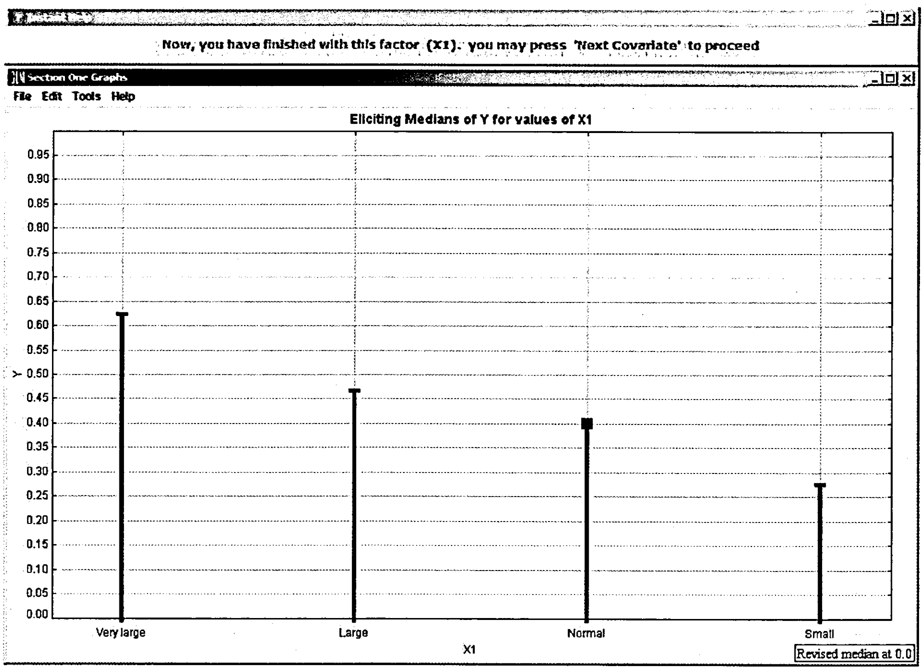


Figure 3.2: A bar chart relationship for a factor given by median assessments

3.2.1 The piecewise-linear model

Consider a response variable ζ, with m continuous covariates $R_1, R_2, \cdots, R_m$ and n categorical variables (factors) $R_{m+1}, R_{m+2}, \cdots, R_{m+n}$. Each variable R_i has $\delta(i) + 1$ knots, $r_{i,0}, r_{i,1}, \cdots, r_{i,\delta(i)}$, where $r_{i,j-1} < r_{i,j}$ for $j = 1, 2, \cdots, \delta(i)$ and $i = 1, 2, \cdots, m+n$. These knots represents the dividing points of the piecewise-linear relation for the continuous variables, or levels for factors, with $r_{i,0}$ taken as the reference point of each continuous covariate R_i, $i = 1, \cdots, m$, or the reference level of each factor R_i, $i = m+1, \cdots, m+n$.

Let $\underline{r}_0$ be the overall reference point, where all variables are at their reference values, i.e.

$$\underline{r}_0 = (r_{1,0}, \quad r_{2,0}, \quad \cdots, \quad r_{m+n,0})'. \tag{3.1}$$

For the response variable ζ, the expert is asked about its mean values given points on the space of the explanatory variables, i.e. about

$$\mu(\underline{r}) = E(\zeta|\underline{R} = \underline{r}), \tag{3.2}$$

where $\underline{R} = (R_1, \quad R_2, \quad \cdots, \quad R_{m+n})'$, and $\underline{r}$ is any specific value of $\underline{R}$.

Let

$$Y = g[\mu(\underline{r})] = \alpha + \underline{\beta}_1'\underline{X}_1 + \underline{\beta}_2'\underline{X}_2 + \cdots + \underline{\beta}_{m+n}'\underline{X}_{m+n}, \tag{3.3}$$

where $g(.)$ is any monotonic increasing link function. If $g(.)$ is monotonic decreasing we multiply it by -1, then change the sign of the resulting regression coefficients. We put

$$\underline{X}_i = (X_{i,1}, \quad X_{i,2}, \quad \cdots, \quad X_{i,\delta(i)})', \qquad i = 1, 2, \cdots, m+n, \tag{3.4}$$

$$\underline{\beta}_i = (\beta_{i,1}, \quad \beta_{i,2}, \quad \cdots, \quad \beta_{i,\delta(i)})', \qquad i = 1, 2, \cdots, m+n. \tag{3.5}$$

The relation between R_i and $\underline{X}_i$, for continuous covariates is that:

$$X_{i,j} = \begin{cases} 0 & \text{if } R_i \leq r_{i,j-1} \\ R_i - r_{i,j-1} & \text{if } r_{i,j-1} < R_i \leq r_{i,j} \\ d_{i,j} & \text{if } r_{i,j} < R_i, \end{cases} \tag{3.6}$$

for $i = 1, 2, \cdots, m$, and $j = 1, 2, \cdots, \delta(i)$, where

$$d_{i,j} = r_{i,j} - r_{i,j-1}. \tag{3.7}$$

For factors, $X_{i,j}$ is defined by:

$$X_{i,j} = \begin{cases} 1 & \text{if } R_i = r_{i,j} \\ 0 & \text{otherwise,} \end{cases} \tag{3.8}$$

for $i = m+1, m+2, \cdots, m+n$, and $j = 1, 2, \cdots, \delta(i)$.

Note that, if $R_i = r_{i,0}$, then $\underline{X}_i$ is a zero vector $(i = 1, \cdots, m+n)$.

The method concentrates on an expert's opinion about each covariate R_i separately, one at a time, assuming that all other covariates are kept at their reference values. Hence, for any specific value r, $Y_i(r)$ is defined as

$$Y_i(r) = g[\mu_i(r)], \tag{3.9}$$

where

$$\mu_i(r) = \mu((r_{1,0}, \quad \cdots, \quad r_{i-1,0}, \quad r, \quad r_{i+1,0}, \quad \cdots, \quad r_{m+n,0})'), \tag{3.10}$$

denotes the mean value of ζ when R_i has a value of r, and $R_j = r_{j,0}$, $j \neq i$.

Then

$$Y_i(r) = \alpha + \underline{\beta}_i' \underline{X}_i, \qquad i = 1, 2, \cdots, m+n. \tag{3.11}$$

Now, for $i = 1, \cdots, m+n$, $j = 1, \cdots, \delta(i)$, let

$$Y_{i,j} = Y_i(r_{i,j}). \tag{3.12}$$

For $\underline{\beta}_i$ as in (3.5), if R_i is a factor and $r = r_{i,j}$, then, in view of (3.8),

$$Y_{i,j} = \alpha + \beta_{i,j}, \tag{3.13}$$

hence, for factors, where $i = m+1, \cdots, m+n$, $j = 1, \cdots, \delta(i)$, we have

$$\beta_{i,j} = Y_{i,j} - Y_{i,0}. \tag{3.14}$$

For continuous covariates, from (3.6) and (3.7), for $i = 1, \cdots, m$, $j = 1, \cdots, \delta(i)$,

$$\beta_{i,j} = \frac{Y_{i,j} - Y_{i,j-1}}{d_{i,j}}. \tag{3.15}$$

The values of $\beta_{i,j}$ are the slopes of the piecewise-linear relation in Figure 3.1.

The prior distribution of α and $\underline{\beta} = (\underline{\beta}_1', \quad \underline{\beta}_2', \quad \cdots, \quad \underline{\beta}_{m+n}')'$ is assumed to be the following multivariate normal distribution,

$$\begin{pmatrix} \alpha \\ \underline{\beta} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} b_0 \\ \underline{b} \end{pmatrix}, \begin{pmatrix} \sigma_{0,0} & \underline{\sigma}_1' \\ \underline{\sigma}_1 & \Sigma \end{pmatrix} \right). \tag{3.16}$$

The elicitation of the hyperparameters b_0, $\underline{b}$, $\sigma_{0,0}$, $\underline{\sigma}_1$ and Σ is reviewed in the next section. The matrix Σ is assumed to have a block-diagonal structure, as the vectors $\underline{\beta}_1', \underline{\beta}_2', \cdots, \underline{\beta}_{m+n}'$ are assumed to be independent *a priori.* We propose three elicitation methods that relax this assumption in the next chapter.

3.2.2 Eliciting the hyperparameters of the multivariate normal prior

The assessments that are required for eliciting all the prior hyperparameters are only medians and quartiles of $\mu_i(r)$. The monotone increasing function $g(.)$ in (3.9) is then used to transform these assessments into medians and quartiles of $Y_i(r)$. Two main properties of the assumed normal distribution of Y are used extensively to elicit the hyperparameters from medians and quartiles. Namely, these properties are equating means to medians and getting variances from interquartile ranges.

It is well-known that, for normally distributed Y,

$$\mathrm{Var}(Y) = \left[\frac{Q_3 - Q_1}{1.349}\right]^2, \tag{3.17}$$

where Q_1 and Q_3 are the lower and upper quartiles of Y, respectively, as 1.349 is the interquartile range of a standard normal distribution.

Using the above approach, the elicitation of each hyperparameter is detailed below.

Eliciting b_0 and $\sigma_{0,0}$

Let $m_{0,0.5}$, $m_{0,0.25}$ and $m_{0,0.75}$ be the median, lower and upper quartiles, respectively, of $\mu(\underline{r}_0)$. Recall that $\underline{r}_0$ is defined in (3.1) as the reference point of all variables, in which case, Y is equal to the constant term α. The expert assesses $m_{0,0.5}$, $m_{0,0.25}$ and $m_{0,0.75}$, which are then transformed into the corresponding quartiles of Y, using the monotone increasing link function $g(.)$ in (3.3), as

$$y_{0,q} = g(m_{0,q}), \qquad \text{for } q = 0.25, 0.5, 0.75. \tag{3.18}$$

So, b_0 and $\sigma_{0,0}$ are elicited, in view of (3.17), as

$$b_0 = y_{0,0.5}, \tag{3.19}$$

$$\sigma_{0,0} = \left[\frac{y_{0,0.75} - y_{0,0.25}}{1.349}\right]^2. \tag{3.20}$$

Eliciting $\underline{b}$

The expert is asked to assume that her previously assessed value $m_{0,0.5}$ is the true value of the mean of ζ at the reference point $r_{i,0}$, i.e. assume that $\mu_i(r_{i,0}) = m_{0,0.5}$, for each covariate i in turn, $i = 1, 2, \cdots, m+n$. Given this information, she then assesses the conditional median of $\mu_i(r)$ at all other knots of R_i. These conditional medians are denoted by $m_{i,j,0.5}$, for $j = 1, 2, \cdots, \delta(i)$.

Hence

$$m_{i,j,0.5} = \text{ The Median of } [\mu_i(r_{i,j})|\mu_i(r_{i,0}) = m_{0,0.5}]. \tag{3.21}$$

The use of the software to assess these conditional medians is reviewed in detail in Section 3.3.3.

From (3.16),

$$\underline{b} = E(\underline{\beta}) = E(\underline{\beta}|\alpha = b_0), \tag{3.22}$$

but, from (3.1), (3.10), (3.18) and (3.19), we have

$$\underline{b} = E[\underline{\beta}|\mu_i(r_{i,0}) = m_{0,0.5}]. \tag{3.23}$$

From the conformaly partitioning in (3.16), each element of $\underline{b}$ in (3.23) is of the form

$$b_{i,j} = E[\beta_{i,j}|\mu_i(r_{i,0}) = m_{0,0.5}]. \tag{3.24}$$

Applying $g(.)$ on both sides of (3.21), in view of (3.9) and (3.12), we get

$$E[Y_{i,j}|\mu_i(r_{i,0}) = m_{0,0.5}] = y_{i,j,0.5}, \tag{3.25}$$

where

$$y_{i,j,0.5} = g(m_{i,j,0.5}). \tag{3.26}$$

Now, from (3.24) and (3.25), $b_{i,j}$ can be elicited for factors, in view of (3.14), as

$$b_{i,j} = y_{i,j,0.5} - y_{0,0.5}, \tag{3.27}$$

for $i = m+1, \cdots, m+n$, $j = 1, \cdots, \delta(i)$, and for continuous covariates, in view of (3.15), as

$$b_{i,j} = \frac{y_{i,j,0.5} - y_{i,j-1,0.5}}{d_{i,j}}, \tag{3.28}$$

for $i = 1, \cdots, m$, $j = 1, \cdots, \delta(i)$.

Eliciting $\underline{\sigma}_1$

For any value α^* satisfying $\alpha^* \neq b_0$, it can be seen, from (3.16) and the theory of multivariate normal distribution, that

$$E(\underline{\beta}|\alpha = \alpha^*) = \underline{b} + \underline{\sigma}_1 \sigma_{0,0}^{-1}(\alpha^* - b_0), \tag{3.29}$$

from which

$$\underline{\sigma}_1 = \frac{[E(\underline{\beta}|\alpha = \alpha^*) - \underline{b}]\sigma_{0,0}}{\alpha^* - b_0}. \tag{3.30}$$

So, $\underline{\sigma}_1$ can be elicited using assessments of $\underline{\beta}_{|\alpha^*} = E(\underline{\beta}|\alpha = \alpha^*)$, or equivalently, the expert is asked to assess

$$m_{i,j,0.5|\alpha^*} = \text{ The Median of } [\mu_i(r_{i,j})|\mu_i(r_{i,0}) = g^{-1}(\alpha^*)], \tag{3.31}$$

Following the same approach as in (3.27) and (3.28), equation (3.31) implies, for factors, that

$$b_{i,j|\alpha^*} = y_{i,j,0.5|\alpha^*} - \alpha^*, \tag{3.32}$$

for $i = m+1, \cdots, m+n$, $j = 1, \cdots, \delta(i)$, while for continuous covariates it implies that

$$b_{i,j|\alpha^*} = \frac{y_{i,j,0.5|\alpha^*} - y_{i,j-1,0.5|\alpha^*}}{d_{i,j}}, \tag{3.33}$$

for $i = 1, \cdots, m$, $j = 1, \cdots, \delta(i)$, where

$$y_{i,j,0.5|\alpha^*} = g(m_{i,j,0.5|\alpha^*}). \tag{3.34}$$

Using the interactive software, α^* is taken as $y_{0,0.75}$, and the task of assessing $m_{i,j,0.5|y_{0,0.75}}$ is detailed in Section 3.3.5.

Eliciting Σ

For eliciting the variance-covariance matrix Σ of the multivariate normal prior distribution of $\underline{\beta}$, the method of GA adopts a structured approach that recursively elicits conditional lower and upper quartiles given incremented sets of the previously assessed median values. The aim of using this structural elicitation is to be able to ensure that assessments yield a matrix Σ that is positive-definite, as required for mathematical coherence.

The idea is that assessed conditional quartiles are transformed, under the normality assumption, into sets of conditional variances that determine all elements of Σ. The positive-definiteness of Σ is guaranteed under a very logical condition that is quite simple to recognize and which the expert can fulfill during the elicitation process. Specifically, the expert is asked to keep reducing her uncertainty as a set of conditional values is increased. Conditioning on more information should increase her confidence in her assessed values, especially as the conditions say that her previous median assessments were accurate.

In what follows, we review the method of GA for eliciting Σ, using the same notations and equations of Garthwaite and Al-Awadhi (2006). In the next chapter, we propose a generalization of the method for the case of correlated vectors of regression coefficients.

Let the conditions that $\mu_i(r_{i,0}) = m_{0,0.5}$ and $\mu_i(r_{i,j}) = m_{i,j,0.5}$ be denoted by $m_{i,0}^0$ and $m_{i,j}^0$, respectively, for $i = 1, 2, \cdots, m+n$, $j = 1, 2, \cdots, \delta(i)$.

For each covariate R_i, $i = 1, 2, \cdots, m+n$, the assessment process consists of $\delta(i)$ steps. At step k, $k = 1, 2, \cdots, \delta(i)$, the expert is asked to assume that conditions $m_{i,0}^0$, $m_{i,1}^0$, $\cdots$, $m_{i,k-1}^0$ hold. Given this information, she assesses the conditional lower and upper quartiles of $\mu_i(r_{i,j})$, denoted by $m_{i,j,0.25}|m_{i,0}^0, \cdots, m_{i,k-1}^0$ and $m_{i,j,0.75}|m_{i,0}^0, \cdots, m_{i,k-1}^0$, respectively, for $j = k, k+1, \cdots, \delta(i)$.

The use of the interactive software to obtain the assessments of these conditional quartiles is discussed in Section 3.3.6.

For $i = 1, 2, \cdots, m+n$, $k = 1, 2, \cdots, \delta(i)$, $j = k, k+1, \cdots, \delta(i)$, using (3.17), the assessed conditional quartiles are used to elicit the conditional variance:

$$\operatorname{Var}(Y_{i,j}|y_{i,0}^0,\cdots,y_{i,k-1}^0) = \left[\frac{g(m_{i,j,0.75}|m_{i,0}^0,\cdots,m_{i,k-1}^0) - g(m_{i,j,0.25}|m_{i,0}^0,\cdots,m_{i,k-1}^0)}{1.349}\right]^2, \tag{3.35}$$

where $y_{i,l}^0$ denotes the condition that $Y_{i,l} = y_{i,l,0.5}$, which is equivalent to $m_{i,l}^0$ from (3.10), (3.12) and(3.26).

For mathematical coherence, conditioning on more values at each further step must reduce the value of the conditional variance in (3.35). Consequently, the expert must steadily reduce her uncertainty when she moves from one step to another. In view of (3.35), this means that the assessment of the interquartile range in step k must be less than that in step $k-1$, which guarantees that

$$\operatorname{Var}(Y_{i,j}|y_{i,0}^0, y_{i,1}^0,\cdots,y_{i,k-1}^0) \geq \operatorname{Var}(Y_{i,j}|y_{i,0}^0, y_{i,1}^0,\cdots,y_{i,k}^0). \tag{3.36}$$

For $i = 1, 2, \cdots, m+n$, $k = 0, 1, \cdots, \delta(i) - 1$, let the conditional variance-covariance matrix $\Lambda_{i,k}$ be defined as

$$\Lambda_{i,k} = \operatorname{Var}(Y_{i,k+1},\cdots,Y_{i,\delta(i)}|y_{i,0}^0, y_{i,1}^0,\cdots,y_{i,k}^0). \tag{3.37}$$

To elicit the full matrix $\Lambda_{i,0}$ in the last step and investigate its positive definiteness, mathematical induction is used to obtain a positive-definite matrix $\Lambda_{i,k-1}$ from $\Lambda_{i,k}$ that has the same property.

To achieve this, let

$$\Lambda_{i,k-1} = \begin{pmatrix} \phi_{i,k,k} & \underline{\phi}_{i,k}' \\ \underline{\phi}_{i,k} & \Phi_{i,k} \end{pmatrix}, \tag{3.38}$$

for $k = 1, 2, \cdots, \delta(i)$, where $\phi_{i,k,k}$ is a scalar, $\underline{\phi}_{i,k}$ is a vector and $\Phi_{i,k}$ is a square matrix. In particular, the scalar $\phi_{i,k,k}$ in (3.38) is given by

$$\phi_{i,k,k} = \operatorname{Var}(Y_{i,k}|y_{i,0}^0, y_{i,1}^0,\cdots,y_{i,k-1}^0), \qquad \text{for } k = 1, 2, \cdots, \delta(i). \tag{3.39}$$

The scalar $\phi_{i,k,k}$ can thus be directly elicited using (3.35).

The vector $\underline{\phi}_{i,k}$ takes the form:

$$\underline{\phi}_{i,k} = \left(\phi_{i,k,k+1}, \quad \phi_{i,k,k+2}, \quad \cdots, \quad \phi_{i,k,\delta(i)}\right)'. \tag{3.40}$$

From the theory of conditional multivariate normal distributions, and for $j = k+1, \cdots, \delta(i)$, we have

$$\mathrm{Var}(Y_{i,j}|y_{i,0}^0, \cdots, y_{i,k}^0) = \mathrm{Var}(Y_{i,j}|y_{i,0}^0, \cdots, y_{i,k-1}^0) - \phi_{i,k,k}^{-1}\phi_{i,k,j}^2. \tag{3.41}$$

Hence, from (3.36) and (3.41), $\phi_{i,k,j}$, $j = k+1, \cdots, \delta(i)$, in (3.40) is given by

$$\phi_{i,k,j} = \{\phi_{i,k,k}[\mathrm{Var}(Y_{i,j}|y_{i,0}^0, y_{i,1}^0, \cdots, y_{i,k-1}^0) - \mathrm{Var}(Y_{i,j}|y_{i,0}^0, y_{i,1}^0, \cdots, y_{i,k}^0)]\}^{\frac{1}{2}}. \tag{3.42}$$

What is left to be elicited in (3.38) is the matrix $\Phi_{i,k}$, which can be computed, using the conditional multivariate normal theory, as

$$\Phi_{i,k} = \Lambda_{i,k} + \underline{\phi}_{i,k}\phi_{i,k,k}^{-1}\underline{\phi}_{i,k}'. \tag{3.43}$$

Hence, the matrix $\Lambda_{i,k-1}$ in (3.38) can be obtained from $\Lambda_{i,k}$, for $k = 1, 2, \cdots, \delta(i) - 1$.

Finally, $\Lambda_{i,0}$ is the result of applying the same routine recursively, starting with $\Lambda_{i,\delta(i)-1}$ as

$$\Lambda_{i,\delta(i)-1} = \mathrm{Var}(Y_{i,\delta(i)}|y_{i,0}^0, y_{i,1}^0, \cdots, y_{i,\delta(i)-1}). \tag{3.44}$$

It can be seen, from (3.35) and (3.44), that

$$\Lambda_{i,\delta(i)-1} > 0. \tag{3.45}$$

From (3.38) and (3.43), we can write the determinant of $\Lambda_{i,k-1}$ as

$$\begin{aligned} |\Lambda_{i,k-1}| &= \phi_{i,k,k}|\Phi_{i,k} - \underline{\phi}_{i,k}\phi_{i,k,k}^{-1}\underline{\phi}_{i,k}'| \\ &= \phi_{i,k,k}|\Lambda_{i,k}|. \end{aligned} \tag{3.46}$$

Hence, from (3.45) and (3.46), $\Lambda_{i,0}$ is positive-definite.

Under the independence assumption between the elements of different vectors of regression coefficients, the matrix Λ can be defined as

$$\Lambda = \begin{pmatrix} D_1^{-1}\Lambda_{1,0}(D_1^{-1})' & O & \cdots & \cdots & \cdots & O \\ O & \ddots & O & \vdots & \vdots & \vdots \\ \vdots & O & D_m^{-1}\Lambda_{m,0}(D_m^{-1})' & O & \vdots & \vdots \\ \vdots & \vdots & O & \Lambda_{m+1,0} & O & \vdots \\ \vdots & \vdots & \vdots & O & \ddots & O \\ O & \cdots & \cdots & \cdots & O & \Lambda_{m+n,0} \end{pmatrix}, \tag{3.47}$$

where, for $i = 1, 2, \cdots, m$, each D_i is a lower triangular matrix given by

$$D_i = \begin{pmatrix} d_{i,1} & 0 & 0 & \cdots & 0 \\ d_{i,1} & d_{i,2} & 0 & \cdots & 0 \\ d_{i,1} & d_{i,2} & d_{i,3} & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ d_{i,1} & d_{i,2} & d_{i,3} & \cdots & d_{i,\delta(i)} \end{pmatrix}. \tag{3.48}$$

With $d_{i,j}$ as defined in (3.7), $d_{i,j} \neq 0$, and hence D_i^{-1} exists. Since, for continuous covariates, from (3.15), we have

$$(Y_{i,1}, \quad Y_{i,2}, \quad \cdots, \quad Y_{i,\delta(i)})' = (\alpha, \quad \cdots, \quad \alpha)' + D_i \underline{\beta}_i, \tag{3.49}$$

then

$$\mathrm{Var}(D_i \underline{\beta}_i | \alpha) = \mathrm{Var}((Y_{i,1}, \quad Y_{i,2}, \quad \cdots, \quad Y_{i,\delta(i)})' | \alpha) = \Lambda_{i,0}, \quad \text{for } i = 1, 2, \cdots, m. \tag{3.50}$$

Hence,

$$V(\underline{\beta}_i | \alpha) = \begin{cases} D_i^{-1}\Lambda_{i,0}(D_i^{-1})', & \text{for } i = 1, 2, \cdots, m, \\ \Lambda_{i,0}, & \text{for } i = m+1, m+2, \cdots, m+n. \end{cases} \tag{3.51}$$

In view of (3.16), the matrix Σ, as the unconditional variance of $\underline{\beta}$, can be given by

$$\Sigma = \Lambda + \underline{\sigma}_1 \sigma_{0,0}^{-1} \underline{\sigma}_1'. \tag{3.52}$$

The full variance-covariance matrix of $(\alpha, \quad \underline{\beta}')'$ is thus positive-definite, from (3.16), (3.47) and (3.52), since

$$\begin{vmatrix} \sigma_{0,0} & \underline{\sigma}_1' \\ \underline{\sigma}_1 & \Sigma \end{vmatrix} = \sigma_{0,0}|\Sigma - \underline{\sigma}_1\sigma_{0,0}^{-1}\underline{\sigma}_1'| = \sigma_{0,0}|\Lambda|. \tag{3.53}$$

The needed assessment tasks in order to elicit all the hyperparameters b_0, $\underline{b}$, $\sigma_{0,0}$, $\underline{\sigma}_1$ and Σ, are given in detail with the software description in Section 3.3.

3.2.3 Computing values for the suggested assessments

For larger elicitation problems, with many covariates or large numbers of knots per covariate, the number of required assessments increases and may represent an overload on the expert. To reduce this number of assessed quantities and help the expert to go through the elicitation process more easily, the method of GA suggests some values of assessments that can be presented by the software to the expert, as a guide for her possible assessed conditional medians and quartiles.

The expert may accept these suggestions if she finds them a reasonable representation of her opinion. Or, instead, she may change or modify them to the best of her knowledge and experience. The method of GA chooses values to suggest by extrapolation from the previously assessed medians and quartiles, assuming some patterns of dependence or independence at different knots of each covariate. The derivations of these suggestions are reviewed below.

Suggesting conditional medians

Assuming independence between α and $\underline{\beta}$, the conditional medians $m_{i,j,0.5}|\alpha^*$ in (3.31) that are required for eliciting $\underline{\sigma}_1$, can be suggested as follows.

Conditioning on $\alpha = \alpha^*$, and under the independence assumption, we have

$$b_{i,j}|\alpha^* = b_{i,j}, \qquad \forall i, j. \tag{3.54}$$

Taking $\alpha^* = y_{0,0.75}$, and equating the right hand sides of (3.27), (3.28) to those of (3.32), (3.33), respectively, equation (3.54) implies that

$$(y_{i,j,0.5}|y_{0,0.75}) - y_{0,0.75} = y_{i,j,0.5} - y_{0,0.5}, \tag{3.55}$$

for $i = m+1, \cdots, m+n$, $j = 1, 2, \cdots, \delta(i)$, and

$$(y_{i,j,0.5}|y_{0,0.75}) - (y_{i,j-1,0.5}|y_{0,0.75}) = y_{i,j,0.5} - y_{i,j-1,0.5}, \tag{3.56}$$

for $i = 1, 2, \cdots, m$, $j = 1, 2, \cdots, \delta(i)$.

Now, from both (3.55) and (3.56), we have

$$(y_{i,j,0.5}|y_{0,0.75}) - y_{i,j,0.5} = y_{0,0.75} - y_{0,0.5}, \tag{3.57}$$

for $i = 1, 2, \cdots, m+n$, $j = 1, 2, \cdots, \delta(i)$.

Hence, from (3.34), (3.57) and the independence assumption, a reasonable suggestion denoted by $\tilde{m}_{i,j,0.5}|y_{0,0.75}$ for $m_{i,j,0.5}|y_{0,0.75}$ is given by

$$\tilde{m}_{i,j,0.5}|y_{0,0.75} = g^{-1}(y_{0,0.75} - y_{0,0.5} + y_{i,j,0.5}), \tag{3.58}$$

for $i = 1, 2, \cdots, m+n$, $j = 1, 2, \cdots, \delta(i)$.

All the components in the right hand side of (3.58) can be computed from the previous assessments as in (3.18) and (3.26). Of course, accepting these suggested medians by the expert will lead to a zero vector as a value of $\underline{\sigma}_1$.

Suggesting conditional quartiles for factors

The simple idea here is to assume that the expert's opinion at one factor level is independent of her opinion at other levels. These lead to conditional quartiles that are unchanged as the number of conditions increases.

In particular, let $\tilde{m}_{i,j,0.25}|m^0_{i,0}, \cdots, m^0_{i,k}$ and $\tilde{m}_{i,j,0.75}|m^0_{i,0}, \cdots, m^0_{i,k}$ be the suggested values of the conditional lower and upper quartiles, $m_{i,j,0.25}|m^0_{i,0}, \cdots, m^0_{i,k}$ and $m_{i,j,0.75}|m^0_{i,0}, \cdots, m^0_{i,k}$, respectively, as required in (3.35), for $i = m+1, \cdots, m+n$, $k = 1, 2, \cdots, \delta(i) - 1$ and $j = k+1, k+2, \cdots, \delta(i)$.

Under the independence assumption, the suggested values are

$$\tilde{m}_{i,j,0.25}|m^0_{i,0}, \cdots m^0_{i,k} = m_{i,j,0.25}|m^0_{i,0}, \cdots m^0_{i,k-1}, \tag{3.59}$$

and

$$\tilde{m}_{i,j,0.75}|m_{i,0}^{0},\cdots m_{i,k}^{0}=m_{i,j,0.75}|m_{i,0}^{0},\cdots m_{i,k-1}^{0}, \tag{3.60}$$

for $i=m+1,\cdots,m+n$, $k=1,2,\cdots,\delta(i)-1$ and $j=k+1,k+2,\cdots,\delta(i)$.

Again, the expert can change any of these suggestions should she wish.

Suggesting conditional quartiles for continuous covariates

In offering suggestions for the conditional quartiles, $m_{i,j,0.25}|m_{i,0}^{0},\cdots,m_{i,k}^{0}$ and $m_{i,j,0.75}|m_{i,0}^{0},\cdots,m_{i,k}^{0}$, as required in (3.35), the method of GA distinguishes between two cases, the case where $k=0$, and the case where $k>0$.

In the case of $k=0$, the assumption is that the relation between Y and R_i is approximately linear, instead of being piecewise-linear. Hence, we may imagine three lines emerging from $y_{0,0.5}$ at the reference knot $r_{i,0}$. The middle line connects all the medians $y_{i,j,0.5}$, while the lower (upper) line connects all the lower (upper) quartiles $g(m_{i,j,0.25}|m_{i,0}^{0})$ $(g(m_{i,j,0.75}|m_{i,0}^{0}))$, at all other knots, $r_{i,j}$, for $j=1,2,\cdots,\delta(i)$.

The linearity assumption ensures that the slopes of each of these three lines are equal at all knots $r_{i,j}$, $j=1,2,\cdots,\delta(i)$. This implies that, for any value $l=1,2\cdots,\delta(i)$, $l\neq j$,

$$\frac{y_{i,j,0.5}-g(m_{i,j,0.25}|m_{i,0}^{0})}{|r_{i,j}-r_{i,0}|}=\frac{y_{i,l,0.5}-g(m_{i,l,0.25}|m_{i,0}^{0})}{|r_{i,l}-r_{i,0}|}, \tag{3.61}$$

and

$$\frac{g(m_{i,j,0.75}|m_{i,0}^{0})-y_{i,j,0.5}}{|r_{i,j}-r_{i,0}|}=\frac{g(m_{i,l,0.75}|m_{i,0}^{0})-y_{i,l,0.5}}{|r_{i,l}-r_{i,0}|}. \tag{3.62}$$

Once the expert has assessed one conditional quartile, $m_{i,l,0.25}|m_{i,0}^{0}$ or $m_{i,l,0.75}|m_{i,0}^{0}$, equation (3.61) or (3.62) can be used to suggest conditional quartiles as

$$\tilde{m}_{i,j,0.25}|m_{i,0}^{0}=g^{-1}\left\{y_{i,j,0.5}-[y_{i,l,0.5}-g(m_{i,l,0.25}|m_{i,0}^{0})]\frac{|r_{i,j}-r_{i,0}|}{|r_{i,l}-r_{i,0}|}\right\}, \tag{3.63}$$

or

$$\tilde{m}_{i,j,0.75}|m_{i,0}^{0}=g^{-1}\left\{y_{i,j,0.5}-[y_{i,l,0.5}-g(m_{i,l,0.75}|m_{i,0}^{0})]\frac{|r_{i,j}-r_{i,0}|}{|r_{i,l}-r_{i,0}|}\right\}, \tag{3.64}$$

respectively, for $j=1,2,\cdots,\delta(i)$, $j\neq l$.

Suggestions for all conditional lower (upper) quartiles are extrapolated from only one assessed value of the conditional lower (upper) quartile. This helps a lot in saving the expert's time and effort during the elicitation process.

For the remaining assessment tasks, where $k = 1, 2, \cdots, \delta(i) - 1$, a new assumption is imposed to obtain the suggested quartiles $\tilde{m}_{i,j,0.25}|m^0_{i,0}, \cdots m^0_{i,k}$ and $\tilde{m}_{i,j,0.75}|m^0_{i,0}, \cdots m^0_{i,k}$. The conditional correlation coefficient between $Y_{i,j}$ and $Y_{i,k}$, for $j = k+1, k+2, \cdots, \delta(i)$, is assumed to be of the form

$$\mathrm{Corr}(Y_{i,j}, Y_{i,k}|y^0_{i,0}, \cdots, y^0_{i,k-1}) = \rho_{i,k-1}^{|r_{i,j}-r_{i,k}|}. \tag{3.65}$$

From which, using theory of bivariate normal distributions, the conditional variance is given by

$$\mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k-1}, y^0_{i,k}) = (1 - \rho_{i,k-1}^{2|r_{i,j}-r_{i,k}|})\mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k-1}), \tag{3.66}$$

for $j = k+1, k+2, \cdots, \delta(i)$.

Once the expert has assessed both a lower and an upper conditional quartiles at any one knot, say $r_{i,k+1}$, the value of $\mathrm{Var}(Y_{i,k+1}|y^0_{i,0}, \cdots, y^0_{i,k-1}, y^0_{i,k})$ can be elicited from equation (3.35). Since $\mathrm{Var}(Y_{i,k+1}|y^0_{i,0}, \cdots, y^0_{i,k-1})$ has already been elicited in step $k-1$, then the value of $\rho_{i,k-1}$ can be computed from (3.66) for $j = k+1$.

Substituting with $\rho_{i,k-1}$ again in (3.66), and using the already elicited values of $\mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k-1})$, for $j = k+2, \cdots, \delta(i)$, the value of $\mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k-1}, y^0_{i,k})$ can be obtained for all $j = k+2, \cdots, \delta(i)$.

After the value of $\mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k})$ has been elicited, we can solve the following two equations for $\tilde{y}_{i,j,0.25}|y^0_{i,0}, \cdots, y^0_{i,k}$ and $\tilde{y}_{i,j,0.75}|y^0_{i,0}, \cdots, y^0_{i,k}$,

$$\mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k}) = \left[\frac{(\tilde{y}_{i,j,0.75}|y^0_{i,0}, \cdots, y^0_{i,k}) - (\tilde{y}_{i,j,0.25}|y^0_{i,0}, \cdots, y^0_{i,k})}{1.349}\right]^2, \tag{3.67}$$

and

$$\frac{(\tilde{y}_{i,j,0.75}|y^0_{i,0}, \cdots, y^0_{i,k}) - y_{i,j,0.5}}{y_{i,j,0.5} - (\tilde{y}_{i,j,0.25}|y^0_{i,0}, \cdots, y^0_{i,k})} = \frac{g(m_{i,j,0.75}|m^0_{i,0}, \cdots, m^0_{i,k-1}) - y_{i,j,0.5}}{y_{i,j,0.5} - g(m_{i,j,0.25}|m^0_{i,0}, \cdots, m^0_{i,k-1})}. \tag{3.68}$$

The use of equation (3.68) aims to ensures that asymmetry of the suggested quartiles around the median at step k is the same as any asymmetry of the assessed quartiles at step $k-1$.

Finally, in view of (3.26), the suggested quartiles are given by

$$\tilde{m}_{i,j,0.25} = g^{-1}(\tilde{y}_{i,j,0.25}|y^0_{i,0}, \cdots, y^0_{i,k}), \tag{3.69}$$

and

$$\tilde{m}_{i,j,0.75} = g^{-1}(\tilde{y}_{i,j,0.75}|y^0_{i,0}, \cdots, y^0_{i,k}), \tag{3.70}$$

for $i = 1, 2, \cdots, m$, $k = 1, 2, \cdots, \delta(i) - 1$ and $j = k+1, k+2, \cdots, \delta(i)$.

3.3 Assessment tasks and software description

The assessment procedure divides naturally into five stages, which are described in turn. A description of the method and theory for using the assessments to estimate the hyperparameters of the prior distribution was reviewed in Section 3.2.2.

3.3.1 Defining the model

The modified version of the software, PEGS-GLM, offers the expert different options for the model to be fitted. The choices available are ordinary linear regression, logistic regression, Poisson regression and any other user defined model. Ordinary linear regression assumes a normal distribution for the response variable with the identity link function. For the logistic regression the assumed distribution is Bernoulli with the logit link function. Poisson regression assumes a Poisson distribution with the logarithm link function.

The expert can choose to define any other model, in which case she will be asked to give a distribution and a link function. Available distributions are the normal, Poisson, binomial, gamma, inverse normal (inverse Gaussian), negative binomial, Bernoulli, geometric and exponential. The user is also asked for some parameters of the selected distribution where appropriate. However, the expert has the option to elicit the extra parameters of the normal and gamma distributions. Novel methods for eliciting these parameters are proposed in Chapter 5.

Available link functions are the canonical, identity, logarithm, logit, reciprocal, square root, probit, log-log, complementary log-log, power, log ratio and user defined link function. For a detailed definition of these link functions see McCullagh and Nelder (1989). For the power link function the software expects the exponent of the power function to be entered by the expert, a value of (-2) is suggested as a default. On choosing the distribution the software suggests the suitable canonical link function so as to help the expert (see Figure 3.3).

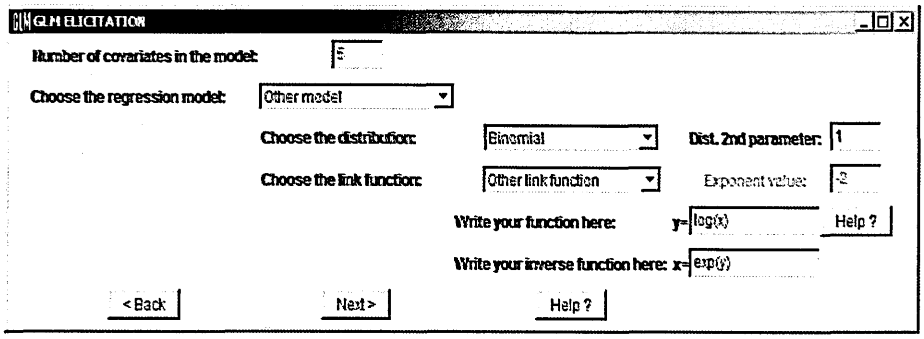


Figure 3.3: The dialogue box for defining the model

An important modification to the software (made by the author) is that it offers a large range of GLM's. It also lets the expert write her own link function and its inverse. The programm can parse both formulas and check their validity as mathematical expressions. Moreover, the program can help by checking whether the functions are valid inverses of each other.

3.3.2 Defining the response variable and covariates

The expert determines the dependant variable with its minimum and maximum values in a dialogue box. The modified version of the software suggests the maximum and minimum values of the response variable whenever possible. The expert may still change them, but, in the light of the chosen model with the specified link function, invalid values are not accepted, and the expert is shown a warning message (For example, the range for a binomial proportion must not extend outside the interval (0,1)).

A set of explanatory variables (covariates) are chosen by the expert. Each covariate is treated as either a continuous random variable or a factor. Continuous covariates are specified with their minimum and maximum, factors are specified with their levels. For each continuous covariate, knots are chosen by the expert or suggested by the software. A reference point is chosen for each covariate, while the origin is the setting for which every covariate is at its reference point. After determining the number, names and types (continuous covariate or categorical factor) of the variables, the expert has only to give the maximum and minimum for each of her continuous covariates together with the value of its reference knot, and the modified software then suggests a suitable number of knots and the position of the reference knot relative to the other knots. The software can then divide the range and gives the value of each knot. This process is done automatically to reduce the burden of data entry, but, again, the expert can change any of these.

The fractional part of each single numeric value is always being rounded to four decimal places, so as to avoid large decimal numbers which are not easily readable nor suitable for graph axis. If higher precision is to be used, measurement units can be modified to use data values of no more than four decimal places. For categorical factors, the expert gives the value of each level. In some cases, when the factor levels are ordinal data, for example, the expert may wish to keep the order of the factor levels, while still being able to select any level as the reference level. The author's modification of the software gives an option to select the reference level of each factor without restricting it to be the first knot (see Figure 3.2).

Using a dialogue box, the median, lower and upper quartiles of $\mu_i(\underline{r}_0)$ at the origin are assessed, namely, $m_{0,0.25}$, $m_{0,0.5}$ and $m_{0,0.75}$, as denoted in Section 3.2.2. These values must be inside the previously specified range of the response variable; if not, the software warns the expert and asks her to resolve this conflict. In the expert's opinion, the true value of $\mu_i(\underline{r}_0)$ is equally likely to be bigger or smaller than the assessed median. Together with the median, these quartiles should divide the range into four equally likely intervals. The expert is encouraged to modify her median and quartile assessments until they divide the range into four intervals that each seem equally likely to her. These assessed values are used as in equations (3.18), (3.19) and (3.20), in Section 3.2.2, to estimate b_0 and $\sigma_{0,0}$.

3.3.3 Initial medians assessments

In the remainder of the elicitation procedure, the expert is separately questioned about each covariate in turn. She is asked to assume the other covariates are at their reference values/levels and forms a piecewise-linear graph or bar chart to represent her opinion about each separate covariate.

The previous stage elicited the expert's median estimate, $m_{0,0.5}$, of $\mu_i(\underline{r}_0)$ at the origin $\underline{r} = \underline{r}_0$. The software plots this value on the reference vertical line and the expert is told to treat it as being correct. The expert then plots her median estimates, $m_{i,j,0.5}$, of $\mu_i(r_{i,j})$, as given in equation (3.21), to form the remainder of the graph. She does this by using the computer mouse to 'click' points on the vertical lines. Straight lines are drawn by the computer between the 'clicked' points, which the expert can change until she feels the graph corresponds to her opinions.

As an illustration, Figure 3.1 shows a software graph for the variable "Weight". The horizontal axis gives values for the variable and the vertical axis gives values of Y. Thus the graph plots the effect on Y as the value of "Weight" varies. The experts is told that, if the graph is fairly flat, then the variable has less influence on Y than if the graph is more curved. The axes and vertical lines are drawn by the software.

For factors, bar charts are formed to represent the expert's opinion. The value of Y has been elicited earlier for the reference level and this gives the height of the reference bar. The expert is told to assume that this bar is correct and to judge the appropriate heights for other bars relative to it. These heights give the value of Y for each level when the other covariates are at their reference values/levels. The software draws thin vertical lines for each level and the expert specifies the height of a bar by clicking on the line with the mouse. This is illustrated in Figure 3.2 where all bars have been specified.

The expert could change an assessment by re-clicking on a line. These median assessments,

$m_{i,j,0.5}$, for the continuous covariates and factors yield estimates of the hyperparameter $\underline{b}$, the mean of the regression coefficient vector $\underline{\beta}$. Theoretical derivation of this estimation is given in detail in Section 3.2.2, equations (3.26), (3.27) and (3.28).

3.3.4 The feedback stage

It is important to help the expert check that her assessments have resulted in a prior distribution that is a reasonable representation of her opinion. This is done through a feedback stage, in which the expert is informed of some other measurements that are inferred from her assessments. She can review and revise her original assessments, in the light of this feedback, if necessary. The current elicitation method has quantified the relationship between the response variable and each covariate in turn, while assuming that all other covariates are at their reference points. Hence, it is important that the expert has feedback that shows her implied predictions of the response variable when all covariates are simultaneously changed from their reference points.

The software computes the values of the response variable at some suggested design points and presents these values to the expert to check that they are reasonable representation of her opinion about the response variable at each suggested design point. Figure 3.4 illustrates a feedback screen, in which the software suggests 6 design points, each of which is a combination of the values and levels of all covariates. Combinations 1 and 4 are the covariate values that gives the minimum and maximum response values, respectively. Combinations 2 and 3 consist of the values that divide each covariate range into one-third and two-thirds, respectively. Minimum and maximum values of each covariate are suggested in combinations 5 and 6, respectively. The expert is asked to specify other design points of interest and to revise any design points offered by the computer that are unrealistic combinations of covariates.

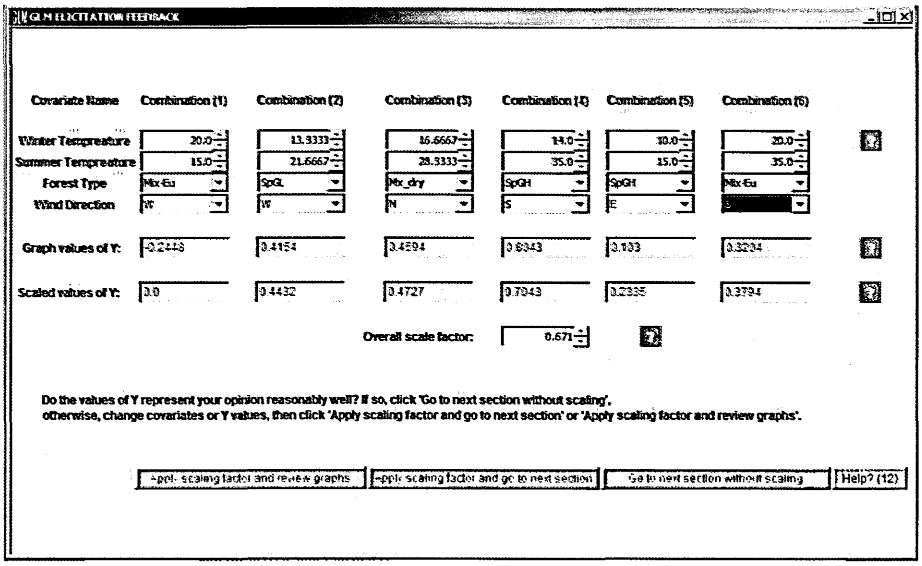


Figure 3.4: The feedback screen

The expert is asked to check that the row of "Graph values of Y", as given in Figure 3.4, is an acceptable representation of her opinion at each design point. These values are predicted from the graphs of medians that were assessed by the expert in Section 3.3.3. The values that are outside the range of the response variable, which was specified at the start of the elicitation process, are flagged in red. The expert can change the unacceptable values by varying the "Overall scale factor" until the row of the "Scaled values of Y", in Figure 3.4, represents her opinion reasonably well in terms of the predicted values at each design point. The scaled values of Y are computed by multiplying all regression coefficients, except the constant term, by the selected value of the overall scale factor.

The expert may choose to review and revise the scaled median assessments again as in Section 3.3.3. Then she will be shown an updated feedback screen. The process will continue until the expert is happy with the graph values of Y as presented in the feedback.

3.3.5 Conditional medians assessments

During this stage the expert is asked to assess her conditional medians, $m_{i,j,0.5}|m_{0,0.75}$, for each covariate in turn, $i = 1, 2, \cdots, m + n$. This is done by changing the conditioning value at the reference point from the median, $m_{0,0.5}$, to the upper quartile, $m_{0,0.75}$. See Figure 3.5

in which median assessments made in the previous stage are given together with the upper quartile at the reference point. The expert assumes that the true value of Y at the reference point is the given upper quartile and she is asked to change the median values at other points to assess $m_{i,j,0.5}|m_{0,0.75}$ in the light of this new conditioning value. Conditional medians for all values have been assessed by the expert in Figure 3.5.

These assessments are needed to elicit a part of the covariance matrix Λ, namely, $\underline{\sigma}_1$, the covariances between α and each of the components of $\underline{\beta}$, see equations (3.32), (3.33) and (3.34), in Section 3.2.2. Suggested values of these conditional medians, $\tilde{m}_{i,j,0.5}|m_{0,0.75}$, are given by the software, assuming that α and $\underline{\beta}$ components are independent, see equation (3.58) in Section 3.2.3. The expert can change these suggested values if she wishes.

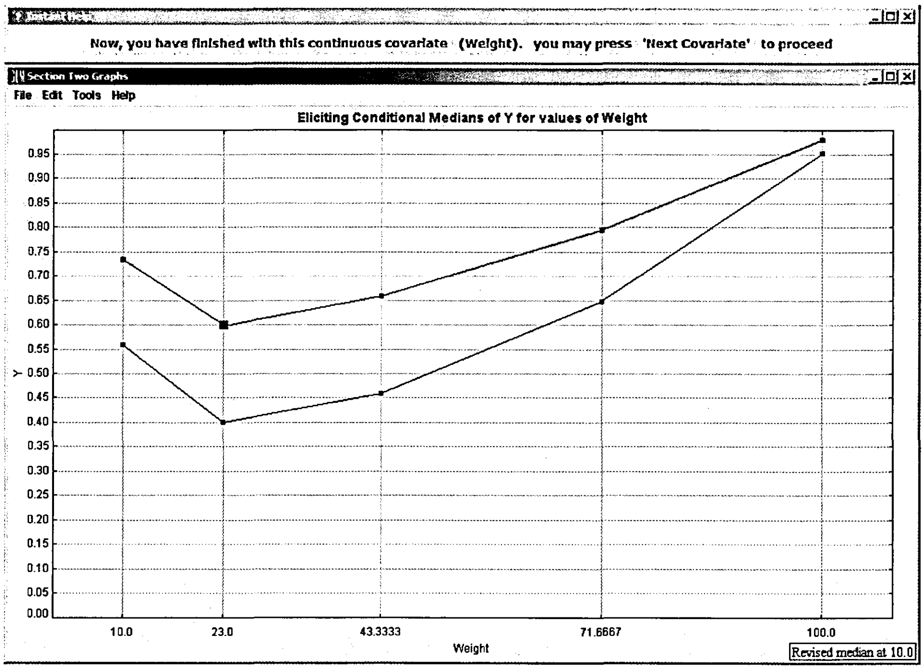


Figure 3.5: Conditional median assessments for the continuous covariate "Weight"

3.3.6 Conditional quartiles assessments

The median assessments provide point estimates of the relationship between different covariates and the variable Y. The remaining task is to quantify the expert's confidence in these estimates and their interrelationship, i.e. how accurate she believes the estimates to be and the correlations between them for each covariate individually. Correlations between coefficients of different covariates are estimated in three different methods proposed in Chapter 4.

In this stage, assessments of conditional lower and upper quartiles, $m_{i,j,0.25}|m_{i,0}^0$ and $m_{i,j,0.75}|m_{i,0}^0$, respectively, are elicited. Assessing quartiles is a harder task for an expert than assessing medians, and quite a large number of quartile assessments are required. To assist the expert, the software suggests some quartile values by extrapolating from other quartile assessments of the expert. The theoretical procedure for getting these suggested values, $\tilde{m}_{i,j,0.25}|m_{i,0}^0$ and $\tilde{m}_{i,j,0.75}|m_{i,0}^0$, as reviewed in Section 3.2.3, was programmed into the software to effectively help the expert during the current stage. The expert can change these assessments and commonly does so but, even then, a starting value to consider seems to make the task easier.

For each continuous covariate in turn, the software displays the graph of the medians that had been assessed earlier, $m_{i,j,0.5}$, and then sets of *conditional* quartile assessments, $m_{i,j,0.25}|m_{i,0}^0$ and $m_{i,j,0.75}|m_{i,0}^0$, are elicited. For this first set of assessments, the condition is that the value of Y at the reference value/level equals the median assessment, i.e. $\mu_i(r_{i,0}) = m_{0,0.5}$.

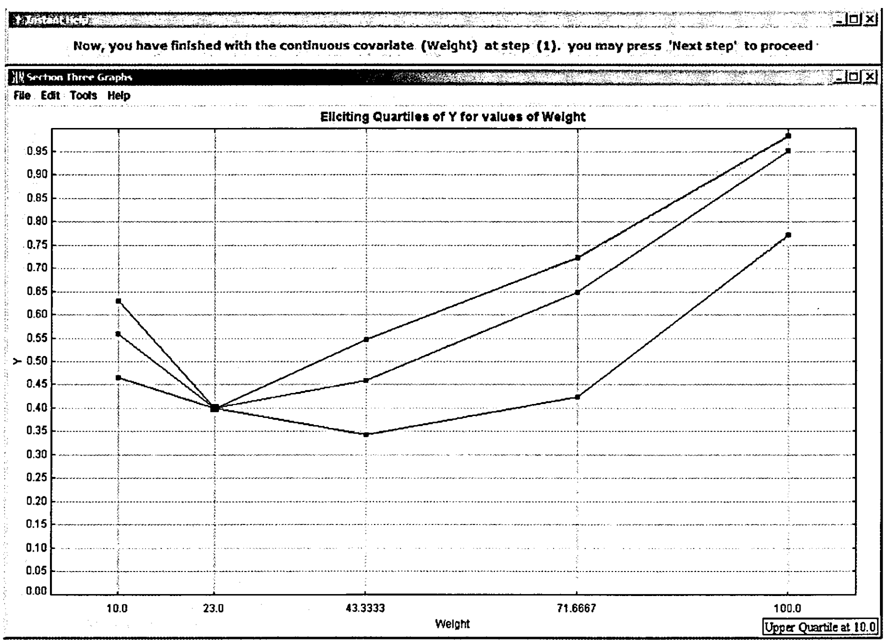


Figure 3.6: Quartile assessments for a continuous covariate

In an interactive graph like Figures 3.6 and 3.7, the expert is asked to give her lower and upper quartiles for Y at one point on each side of the medians for each value/level of the covariate except for the reference value/level. The lines joining quartiles look similar to confidence intervals and it is emphasized to the expert that there should only be a 50% chance that the value of Y is between the lines at any point. The expert uses the computer mouse to make assessments or change values suggested by the software.

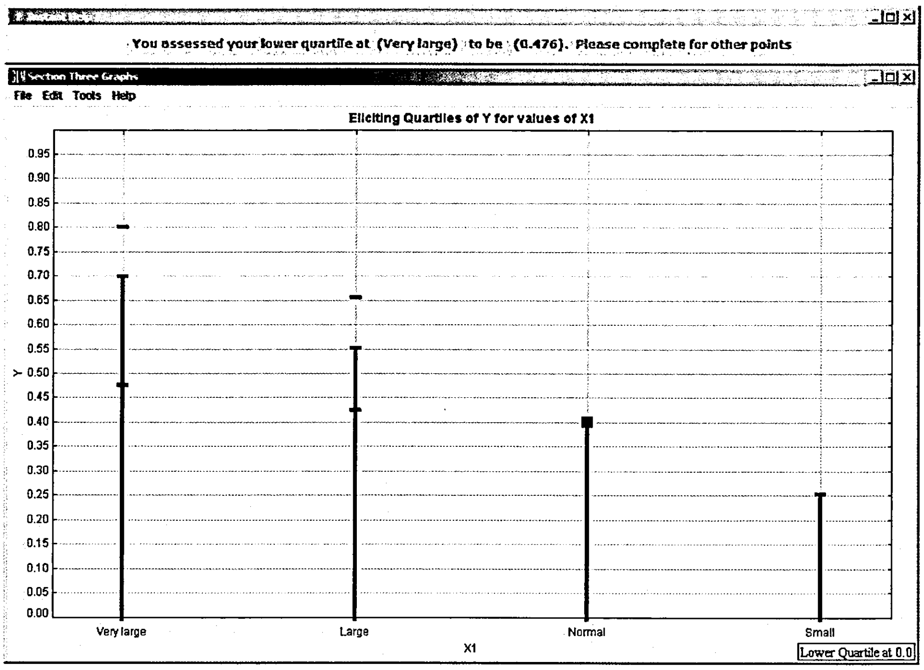


Figure 3.7: Quartile assessments for a factor

For the second set of conditional assessments, the expert is asked to assume that the median estimates of Y are correct at both the reference value/level and the nearest points on each side of it, i.e. conditions $m_{i,0}^0$, $m_{i,1}^0$, $\cdots$, $m_{i,k}^0$, in Section 3.2.2. The expert gives lower and upper quartiles at another point, $r_{i,k+1}$, and the software suggests quartiles, $\tilde{m}_{i,j,0.25}|m_{i,0}^0, m_{i,1}^0, \cdots, m_{i,k}^0$ and $\tilde{m}_{i,j,0.75}|m_{i,0}^0, m_{i,1}^0, \cdots, m_{i,k}^0$, for the remaining points, $j = k+2, \cdots, \delta(i)$. In Figure 3.8 lower quartiles have been assessed while upper quartiles are to be assessed. The expert modifies quartile values so as to represent her opinion, subject to the restriction that the current values must be within the previous set of quartile assessments, $m_{i,j,0.25}|m_{i,0}^0, m_{i,1}^0, \cdots, m_{i,k-1}^0$ and $m_{i,j,0.75}|m_{i,0}^0, m_{i,1}^0, \cdots, m_{i,k-1}^0$. The idea is that as conditions increase, uncertainty should reduce. As detailed in Section 3.2.2, this condition guarantees that the covariance matrix of correlation coefficients is positive definite.

Figure 3.8 illustrates the graph formed at that stage. The two red lines (the outer lines) represent the previous set of quartile assessments, the second highest (black) line gives the median assessments, and the second lowest (blue) line joins the new lower quartile assessments. The black line joining the median at the right two bold points represents the condition that these medians should be treated as being correct. In assessing quartiles, the expert is told to consider the points to which she thinks the blue line may reasonably extend.

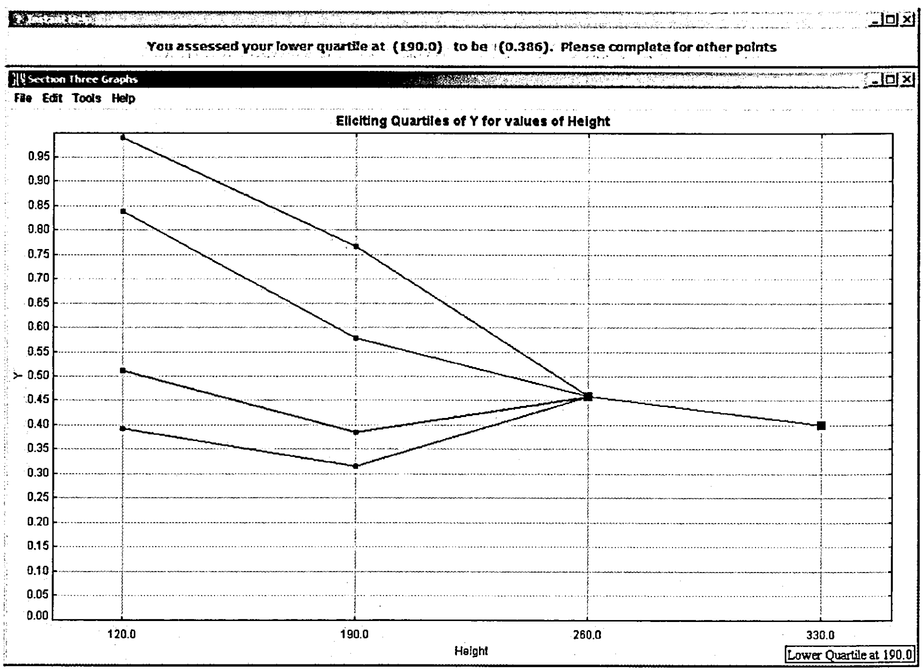


Figure 3.8: Assessing quartiles conditioning on two fixed points

Conditional assessments are also needed for factors. The software displays the bar chart that was formed during the assessment of medians. Conditional on the value of the bar at the reference level being correct, i.e. on $m_{i,0}^0$, the expert assesses a lower and an upper quartile, $m_{i,j,0.25}|m_{i,0}^0$ and $m_{i,j,0.75}|m_{i,0}^0$, respectively, for other factor levels.

For each further set of conditional assessments, for both continuous covariates and factors, the expert is asked to assume that a further median given by another value/level was correct and to give her opinion about quartiles for the remaining values/levels. This is continued until the condition includes all but one of the values/levels at one side or one at both sides, when the expert gives her opinion about just the last one or two values/levels (see Figure 3.9).

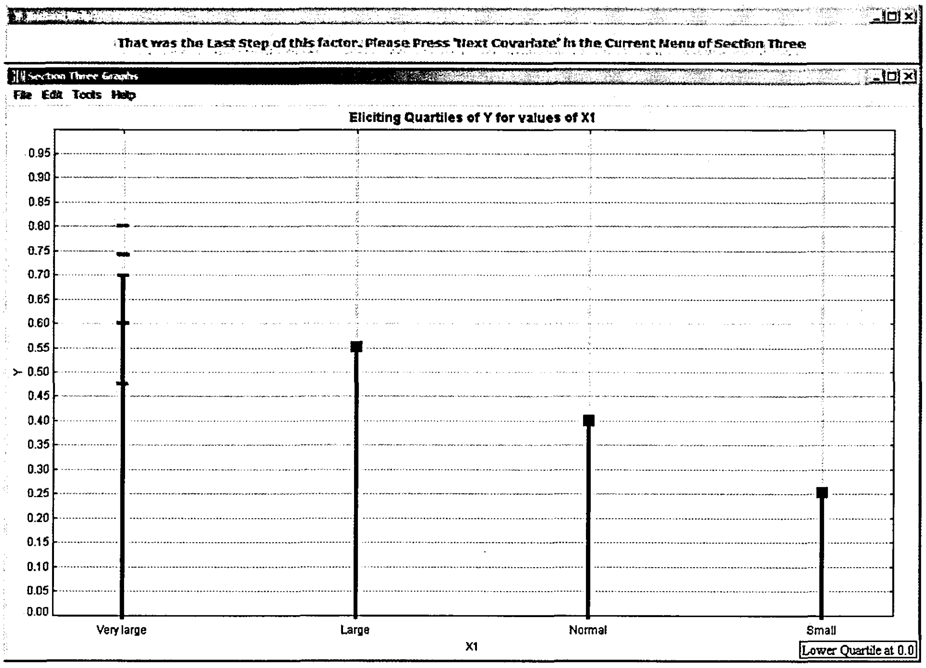


Figure 3.9: Assessing conditional quartiles for the last level of a factor

As in other parts of the elicitation procedure, the expert uses the mouse to make assessments. Figure 3.9 illustrates the bar chart when conditioning values are specified (indicated by the solid squares); quartiles for the last level are marked with short horizontal blue lines (the inner two lines), while the highest and lowest (red) lines represent the previous quartiles conditioning on fewer medians. Again, current conditional quartiles are not allowed to lay outside these red lines. The conditional quartile assessments, $m_{i,j,0.25}|m_{i,0}^0, m_{i,1}^0, \cdots, m_{i,k}^0$ and $m_{i,j,0.75}|m_{i,0}^0, m_{i,1}^0, \cdots, m_{i,k}^0$, yield estimates of the variance, Σ, of the hyperparameter $\underline{\beta}$, see Section 3.2.2.

The conditional assessments complete the elicitation procedure for the case of independent coefficients as required in Section 3.2.2.

3.4 Concluding comments

The piecewise-linear elicitation method for logistic regression introduced by Garthwaite and Al-Awadhi (2006), as reviewed in this chapter, is widely applicable for GLMs with any monotonic increasing link function. The method only requires conditional and unconditional medians and quartiles to be assessed from the expert, these assessment tasks are easy to perform using the bisection method. The number of assessed quantities is sufficient to elicit a mean vector and a positive-definite variance-covariance matrix for a multivariate normal prior distribution of the regression parameters of any GLM. The available modified software has increased the applicability of the method and made its implementation easier for the expert. However, the independence assumption between different regression coefficients that is imposed by the method is sometimes unrealistic and need to be relaxed. Extended methods that relax this assumption are proposed in the next chapter.

Chapter 4

Eliciting a covariance matrix for dependant coefficients in GLMs

4.1 Introduction

For quantifying expert's opinion about generalized linear models (GLM), Garthwaite and Al-Awadhi (2011) proposed a method of eliciting opinion about the prior distribution of the regression coefficients. This method, which will be referred to here as GA, is a generalization of the same authors' piecewise-linear model that they used for quantifying opinion for logistic regression (Garthwaite and Al-Awadhi (2006)). A detailed description of their method has been given in the previous chapter.

In their work, the relationship between each continuous predictor variable and the dependant variable (assuming all other variables are held fixed) was modeled as a piecewise-linear function. They used a multivariate normal distribution to represent prior knowledge about the regression coefficients. These coefficients were allowed to be dependant if they were associated with a single variable. However, they assumed that there was no interaction between any variables, in the sense that regression coefficients were *a priori* independent if associated with different variables.

Our aim in this chapter is to relax the independence assumption between coefficients of different variables. In fact, in many practical situations, it may be thought that regression coefficients of different variables should be related in the prior distribution, if the prior distribution is to give a reasonable representation of the expert's opinion. The expert may be asked to state which variables this applies to. We propose three different elicitation methods that are implemented in interactive graphical software. The software is freely available as PEGS-GLM (Correlated Coefficients) at http://statistics.open.ac.uk/elicitation.

In the first method, after assessing additional conditional quartiles, GA's method of estimating the variance-covariance matrix is generalized and used to estimate the variance-covariance matrix in generalized linear models where pairs of correlated vectors of coefficients are not necessarily independent in the prior distribution. The second method is designed to require a smaller number of assessments. Its generalization to the case of various vectors of correlated coefficients is straightforward, where the required conditions for positive-definiteness can be easily investigated. A third flexible method is proposed in which the expert assesses the relative correlation structure for all pairs of vectors, then chooses one of the other two methods to specify the coefficient for the highest correlated vectors. This method automatically fulfil the requirement that the whole variance-covariance matrix must be positive-definite. The three proposed methods are detailed below.

4.2 A proposed method for eliciting the variance-covariance matrix of a pair of correlated vectors of coefficients

In this section, we propose an elicitation method that generalizes the method of GA to handle correlated coefficients in GLMs. We start by generalizing the equations given in the previous chapter to make them applicable to the case of correlated coefficients. The underlying mathematical framework is given in Section 4.2.1. The equations given there show how the required conditional assessments are mathematically treated to elicit the variance-covariance matrix. Our approach to assess these conditional quartiles from the expert using interactive software is detailed in Section 4.2.2.

4.2.1 Notations and theoretical framework

Consider the piecewise-linear GLM of GA, with m continuous covariates $R_1, R_2, \cdots, R_m$ and n categorical variables (factors) $R_{m+1}, R_{m+2}, \cdots, R_{m+n}$. The model has been defined in Chapter 3, equations (3.1) to (3.15).

Recall that the prior distribution of α and $\underline{\beta} = (\underline{\beta}'_1, \ \underline{\beta}'_2, \ \ldots, \ \underline{\beta}'_{m+n})'$ is assumed to be a multivariate normal distribution

$$\begin{pmatrix} \alpha \\ \underline{\beta} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} b_0 \\ \underline{b} \end{pmatrix}, \begin{pmatrix} \sigma_{0,0} & \underline{\sigma}'_1 \\ \underline{\sigma}_1 & \Sigma \end{pmatrix} \right). \tag{4.1}$$

The elicitation of the hyperparameters b_0, $\underline{b}$, $\sigma_{0,0}$, $\underline{\sigma}_1$ and Σ has been reviewed in Section 3.2.2.

Equation (3.52) states that $\Sigma = \Lambda + \underline{\sigma}_1 \sigma_{0,0}^{-1} \underline{\sigma}'_1$, where Λ has been assumed to have the block-diagonal structure

$$\Lambda = \begin{pmatrix} D_1^{-1}\Lambda_{1,0}(D_1^{-1})' & O & \cdots & \cdots & \cdots & O \\ O & \ddots & O & \vdots & \vdots & \vdots \\ \vdots & O & D_m^{-1}\Lambda_{m,0}(D_m^{-1})' & O & \vdots & \vdots \\ \vdots & \vdots & O & \Lambda_{m+1,0} & O & \vdots \\ \vdots & \vdots & \vdots & O & \ddots & O \\ O & \cdots & \cdots & \cdots & O & \Lambda_{m+n,0} \end{pmatrix}, \tag{4.2}$$

where, for $i = 1, 2, \cdots, m$, each D_i is a lower triangular matrix given by

$$D_i = \begin{pmatrix} d_{i1} & 0 & 0 & \cdots & 0 \\ d_{i1} & d_{i2} & 0 & \cdots & 0 \\ d_{i1} & d_{i2} & d_{i3} & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ d_{i1} & d_{i2} & d_{i3} & \cdots & d_{i\delta(i)} \end{pmatrix}. \tag{4.3}$$

Hence, for continuous covariates

$$(Y_{i,1}, \quad Y_{i,2}, \quad \cdots, \quad Y_{i,\delta(i)})' = (\alpha, \quad \cdots, \quad \alpha)' + D_i\underline{\beta}_i,$$

and

$$\mathrm{Var}(D_i\underline{\beta}_i|\alpha) = \mathrm{Var}((Y_{i,1}, \quad Y_{i,2}, \quad \cdots, \quad Y_{i,\delta(i)})'|\alpha) = \Lambda_{i,0}, \quad \text{for } i = 1, 2, \cdots, m,$$

where

$$Y_{i,j} = g[\mu((r_{1,0}, \quad \cdots, \quad r_{i-1,0}, \quad r_{i,j}, \quad r_{i+1,0}, \quad \cdots, \quad r_{m+n,0})')].$$

As required, Y is a continuous piecewise-linear function of the variable R_i, if all other variables are kept at their reference values. Hence,

$$\mathrm{Var}(\underline{\beta}_i|\alpha) = \begin{cases} D_i^{-1}\Lambda_{i,0}(D_i^{-1})', & \text{for } i = 1, 2, \cdots, m, \\ \Lambda_{i,0}, & \text{for } i = m+1, m+2, \cdots, m+n. \end{cases} \tag{4.4}$$

Formulae for $\Lambda_{i,0}$ are given in GA as reviewed in the previous chapter, see equations (3.37) to (3.44).

Instead of assuming the block-diagonal structure given by (4.2), we will conformally partition Λ as

$$\Lambda = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,m+n} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,m+n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m+n,1} & \Sigma_{m+n,2} & \cdots & \Sigma_{m+n,m+n} \end{pmatrix}, \tag{4.5}$$

where

$$\Sigma_{i,i} = \mathrm{Var}(\underline{\beta}_i|\alpha), \quad \text{for } i = 1, 2, \cdots, m+n, \tag{4.6}$$

and the submatrices $\Sigma_{s,t}$ are not necessarily zero matrices ($s = 1, 2, \cdots, m+n$, $t = 1, 2, \cdots, m+n$ and $s \neq t$). We will estimate the $\Sigma_{s,t}$ matrices in (4.5) by generalizing the method of GA.

Assume that the expert believes that $\underline{\beta}_s$ and $\underline{\beta}_t$ are correlated. For $s < t$, we must estimate the upper diagonal covariance submatrix $V_{s,t}$ of V,where,

$$V = \mathrm{Var}[(\underline{\beta}_s' \quad \underline{\beta}_t')'|\alpha] \equiv \begin{pmatrix} V_{s,s} & V_{s,t} \\ V_{t,s} & V_{t,t} \end{pmatrix}. \tag{4.7}$$

As a variance-covariance matrix is symmetric, $V_{t,s} = V_{s,t}'$.

The correlation relationships are handled one pair at a time. Suppose we are currently interested only in the pair $\underline{\beta}_s$, $\underline{\beta}_t$, and that these are correlated in the prior distribution. (The same procedure can be followed for each pair that is correlated.)

For $s = 1, 2, \cdots, m+n$, $t = 1, 2, \cdots, m+n$, and $s < t$, let $\delta_{st} = \delta(s) + \delta(t)$, and for $k = 0, 1, \cdots, \delta_{st} - 1$, put

$$\Lambda_{st,k} = \begin{cases} \mathrm{Var}(Y_{s,k+1}, \cdots, Y_{s,\delta(s)}, Y_{t,1}, \cdots, Y_{t,\delta(t)}|y_{s,0}^0, \cdots, y_{s,k}^0), \\ \qquad\qquad \text{for} \quad 0 \le k \le \delta(s) - 1. \\ \mathrm{Var}(Y_{t,k-\delta(s)+1}, \cdots, Y_{t,\delta(t)}|y_{s,0}^0, \cdots, y_{s,\delta(s)}^0, y_{t,0}^0, \cdots, y_{t,k-\delta(s)}^0), \\ \qquad\qquad \text{for} \quad \delta(s) \le k \le \delta_{st} - 1. \end{cases}$$

Specifying conditional values $y^0_{i,j}$, is equivalent to conditioning on the corresponding assessed medians $m^0_{i,j}$, as detailed in the previous chapter.

We start with

$$\Lambda_{st,\delta_{st}-1} = \mathrm{Var}(Y_{t,\delta(t)}|y^0_{s,0}, \cdots, y^0_{s,\delta(s)}, y^0_{t,0}, \cdots, y_{t,\delta(t)-1}), \tag{4.8}$$

which can be computed from the conditional quartile assessments of the covariate R_t at $\delta(t)$. The conditioning specifies the values of Y at all previous knots of R_t and all knots of R_s as well. Given these conditions, the expert assesses conditional quartiles $m_{t,\delta(t),0.25}$ and $m_{t,\delta(t),0.75}$. The method of assessing these quartiles is detailed in Section 4.2.2. The formula for computing the variance ensures that $\Lambda_{st,\delta_{st}-1} > 0$, since

$$\begin{aligned}\Lambda_{st,\delta_{st}-1} = {} & [g(m_{t,\delta(t),0.75}|m^0_{s,0}, \cdots, m^0_{s,\delta(s)}, m^0_{t,0}, \cdots, m^0_{t,\delta(t)-1}) \\ & - g(m_{t,\delta(t),0.25}|m^0_{s,0}, \cdots, m^0_{s,\delta(s)}, m^0_{t,0}, \cdots, m^0_{t,\delta(t)-1})/1.349]^2. \end{aligned} \tag{4.9}$$

We put

$$\Lambda_{st,k-1} = \begin{pmatrix} \phi_{st,k,k} & \underline{\phi}'_{st,k} \\ \underline{\phi}_{st,k} & \Phi_{st,k} \end{pmatrix}, \tag{4.10}$$

for $k = 1, 2, \cdots, \delta_{st}$, where $\phi_{st,k,k}$ is a scalar, $\underline{\phi}_{st,k}$ is a vector and $\Phi_{st,k}$ is a square matrix. In particular, the scalar $\phi_{st,k,k}$ in (4.10) is given by:

$$\phi_{st,k,k} = \begin{cases} \mathrm{Var}(Y_{s,k}|y^0_{s,0}, \cdots, y^0_{s,k-1}), & \text{for} \quad 1 \le k \le \delta(s), \\ \mathrm{Var}(Y_{t,k-\delta(s)}|y^0_{s,0}, \cdots, y^0_{s,\delta(s)}, y^0_{t,0}, \cdots, y^0_{t,k-\delta(s)-1}), & \\ & \text{for} \quad \delta(s)+1 \le k \le \delta_{st}. \end{cases} \tag{4.11}$$

Recall from the previous chapter that, for $j = k+1, \cdots, \delta(i)$,

$$\mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k}) = \mathrm{Var}(Y_{i,j}|y^0_{i,0}, \cdots, y^0_{i,k-1}) - \phi^{-1}_{i,k,k}\phi^2_{i,k,j}, \tag{4.12}$$

as a result of the theory about conditional multivariate normal distributions. Equation (4.12) can be generalized for the case where there are two correlated vectors of coefficients. Then, the vector $\underline{\phi}_{st,k}$ in (4.10) takes the form:

$$\underline{\phi}_{st,k} = (\phi_{st,k,k+1}, \quad \cdots, \quad \phi_{st,k,\delta_{st}})',$$

where

$$\phi_{st,k,j} = \begin{cases} [\phi_{st,k,k}\{\text{Var}(Y_{s,j}|y^0_{s,0},\cdots,y^0_{s,k-1}) - \text{Var}(Y_{s,j}|y^0_{s,0},\cdots,y^0_{s,k})\}]^{\frac{1}{2}}, \\ \qquad\qquad \text{for} \quad 1 \le k < \delta(s), \\ \qquad\qquad\quad j = k+1,\cdots,\delta(s). \\ [\phi_{st,k,k}\{\text{Var}(Y_{t,j-\delta(s)}|y^0_{s,0},\cdots,y^0_{s,k-1}) \\ \qquad\qquad -\text{Var}(Y_{t,j-\delta(s)}|y^0_{s,0},\cdots,y^0_{s,k})\}]^{\frac{1}{2}}, \\ \qquad\qquad \text{for} \quad 1 \le k \le \delta(s), \\ \qquad\qquad\quad j = \delta(s)+1,\cdots,\delta_{st}. \\ [\phi_{st,k,k}\{\text{Var}(Y_{t,j-\delta(s)}|y^0_{s,0},\cdots,y^0_{s,\delta(s)},y^0_{t,0},\cdots,y^0_{t,k-\delta(s)-1}) \\ \quad -\text{Var}(Y_{t,j-\delta(s)}|y^0_{s,0},\cdots,y^0_{s,\delta(s)},y^0_{t,0},\cdots,y^0_{t,k-\delta(s)})\}]^{\frac{1}{2}}, \\ \qquad\qquad \text{for} \quad \delta(s)+1 \le k < \delta_{st}, \\ \qquad\qquad\quad j = k+1,\cdots,\delta_{st}. \end{cases} \tag{4.13}$$

The main constraint needed here is that conditioning on more values at each further step must reduce the value of a conditional variance. The expert must therefore reduce her uncertainty as the elicitation process progresses. It means that her assessments of each interquartile range must steadily decrease. This will ensure that, for $i = 1,2,\cdots,m+n$, $j > k$, $1 \le k < \delta_{st}$,

$$\text{Var}(Y_{i,j}|y^0_{i,0},\cdots,y^0_{i,k-1}) > \text{Var}(Y_{i,j}|y^0_{i,0},\cdots,y^0_{i,k}). \tag{4.14}$$

Conditional variances in (4.11) and (4.13) can be written in terms of the assessed conditional quartiles as

$$\text{Var}(Y_{s,j}|y^0_{s,0},\cdots,y^0_{s,k}) = \left[\frac{g(m_{s,j,0.75}|m^0_{s,0},\cdots,m^0_{s,k}) - g(m_{s,j,0.25}|m^0_{s,0},\cdots,m^0_{s,k})}{1.349}\right]^2,$$

$$\text{for } 0 \le k < \delta(s), \quad j = k+1,\cdots,\delta(s), \tag{4.15}$$

$$\text{Var}(Y_{t,j}|y^0_{s,0},\cdots,y^0_{s,k}) = \left[\frac{g(m_{t,j,0.75}|m^0_{s,0},\cdots,m^0_{s,k}) - g(m_{t,j,0.25}|m^0_{s,0},\cdots,m^0_{s,k})}{1.349}\right]^2,$$

$$\text{for } 0 \le k \le \delta(s), \quad j = 1,\cdots,\delta(t), \tag{4.16}$$

$$\mathrm{Var}(Y_{t,j}|y^0_{s,0},\cdots,y^0_{s,\delta(s)},y^0_{t,0},\cdots,y^0_{t,k}) = [g(m_{t,j,0.75}|m^0_{s,0},\cdots,m^0_{s,\delta(s)},m^0_{t,0},\cdots,m^0_{t,k})$$

$$- g(m_{t,j,0.25}|m^0_{s,0},\cdots,m^0_{s,\delta(s)},m^0_{t,0},\cdots,m^0_{t,k})/1.349]^2,$$

$$\text{for } 1 \leq k < \delta(t), \quad j = 1,\cdots,\delta(t). \tag{4.17}$$

What is left to be estimated in (4.10) is the matrix $\Phi_{st,k}$, which can be computed, using the conditional multivariate normal theory, as

$$\Phi_{st,k} = \Lambda_{st,k} + \underline{\phi}_{st,k}\phi^{-1}_{st,k,k}\underline{\phi}'_{st,k}. \tag{4.18}$$

Hence, the matrix $\Lambda_{st,k-1}$ in (4.10) can be obtained from $\Lambda_{st,k}$, for $k = 1,2,\cdots,\delta_{st}-1$. Finally, $\Lambda_{st,0}$ is the result of applying the same routine recursively, starting with $\Lambda_{st,\delta_{st}-1}$ as in (4.8).

If $\Lambda_{st,0}$ is conformally partitioned as

$$\Lambda_{st,0} = \begin{pmatrix} \Lambda_{s,s} & \Lambda_{s,t} \\ \Lambda_{t,s} & \Lambda_{t,t} \end{pmatrix}, \tag{4.19}$$

then its submatrices can be used to obtain the required conformally partitioned matrix in (4.7), as follows. Take

$$V = \begin{pmatrix} V_{s,s} & V_{s,t} \\ V'_{s,t} & V_{t,t} \end{pmatrix},$$

where $V_{s,s}$ is the variance of $\underline{\beta}_s$ given α. Clearly, $V_{s,s} = \Sigma_{s,s}$ of equation (4.5), also $V_{s,s} = \Lambda_{s,s}$ of equation (4.19). Hence, from (4.2),

$$V_{s,s} = \begin{cases} D_s^{-1}\Lambda_{s,0}(D_s^{-1})', & \text{for } s = 1,2,\cdots,m, \\ \Lambda_{s,0}, & \text{for } s = m+1, m+2,\cdots,m+n. \end{cases}$$

The submatrix $V_{s,t}$ is the covariance of $\underline{\beta}_s$ and $\underline{\beta}_t$ given α, of the form

$$V_{s,t} \equiv \begin{cases} D_s^{-1}\Lambda_{s,t}(D_t^{-1})', & \text{for } s = 1,2,\cdots,m, \\ & \quad t = 1,2,\cdots,m, \\ D_s^{-1}\Lambda_{s,t}, & \text{for } s = 1,2,\cdots,m, \\ & \quad t = m+1, m+2,\cdots,m+n, \\ \Lambda_{s,t}, & \text{for } s = m+1, m+2,\cdots,m+n, \\ & \quad t = m+1, m+2,\cdots,m+n. \end{cases}$$

Noting that $\Lambda_{t,t}$ in (4.19) is the conditional variance of $\underline{\beta}_t$ given $\underline{\beta}_s$ and α, another version conditional only on α can be taken as

$$V_{t,t} = \begin{cases} D_t^{-1}\Lambda_{t,t}(D_t^{-1})' + V_{s,t}'V_{s,s}^{-1}V_{s,t}, & \text{for } t = 1,2,\cdots,m, \\ \Lambda_{t,t} + V_{s,t}'V_{s,s}^{-1}V_{s,t}, & \text{for } t = m+1, m+2,\cdots,m+n. \end{cases}$$

With this construction, in Section 4.2.3 below, the matrix V is shown to be positive-definite.

4.2.2 Assessment tasks and software description

The modified elicitation software PEGS-GLM (Correlated Coefficients), that is freely available at http://statistics.open.ac.uk/elicitation, elicits the expert's conditional quartiles that are needed to estimate the covariance matrix of correlated pair of covariates. The mathematical details have been given in Section 4.2.1. The expert is asked whether the regression coefficients of any pair of covariates are dependent in her prior distribution. If so, she will be asked to name the two variables that have such dependence. Then she will be shown a panel that simultaneously displays two graphs (see Figure 4.1 or Figure 4.2).

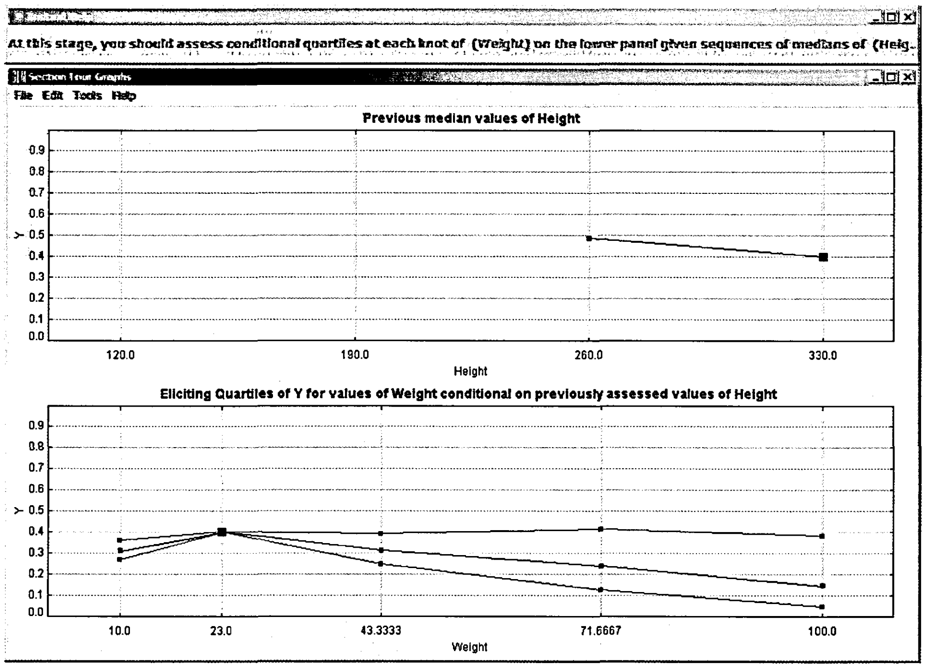


Figure 4.1: Assessments needed in the first phase for correlated covariates

The upper graph of the panel is for one variable of the correlated pair. It shows the previously assessed median values for that variable, denoted by $m^0_{i,j}$, $i = 1, 2, \cdots, m+n$, $j = 1, 2, \cdots, \delta(i)$, as in equations (4.9) and (4.15)-(4.17). The expert is asked to assume that these median values are the correct values of Y at the given knots. That is, they are accurate estimates of the mean response for the specified covariate values. Conditional on this information, the expert clicks on the lower interactive graph to assess new conditional quartile values, denoted by $m_{i,j,0.25}$ and $m_{i,j,0.75}$, $i = 1, 2, \cdots, m+n$, $j = 1, 2, \cdots, \delta(i)$, in equations (4.9) and (4.15)-(4.17).

The procedure consists of two phases; in the first phase the expert assesses quartile values for the variable in the lower graph given sets of medians for the variable in the upper graph. Specifically, these medians are denoted by $m^0_{s,0}, \cdots, m^0_{s,k}$ in equation (4.16). The set of conditioning values of the first variable in the upper graph are incremented by one extra value at each new step. The expert is asked to take account of the additional information and re-assess conditional quartiles. This gives the assessments denoted by $m_{t,j,0.25}$ and $m_{t,j,0.75}$ in equation (4.16).

Step 1 of the first phase is shown in Figure 4.1, where the expert is asked to assess conditional quartiles for different knots of the "Weight" variable in the lower graph conditioning on the previously assessed medians $m^0_{s,0}$, $m^0_{s,1}$ of the "Height" variable at its reference knot and one other knot. These two medians are connected by the rightmost (black) line in the upper graph. The conditioning set includes also the median of the "Weight" variable at its reference knot (23.0).

The upper and lower (red) curves in Figures 4.1 and 4.2, represent the previous quartile assessments conditioning on fewer medians. Current conditional quartiles are not allowed to lay outside these red lines. This fulfils condition (4.14), which guarantees the positive-definiteness of the variance-covariance matrix, as discussed before. Specifying these conditions by drawing boundary lines on the graph makes it easier for the expert to absorb what the conditional values are and what they imply. This helps her apply the idea of reducing uncertainty as conditions increase.

The second phase starts after conditioning on the median values at all knots in the top graph, denoted by $m^0_{s,0}, \cdots, m^0_{s,\delta(s)}$ in equation (4.17). Each further step in this second phase adds an extra median value from the lower graph to the conditioning set. These additional values are $m^0_{t,0}, \cdots, m^0_{t,k}$ in equation (4.17). Further conditional quartiles $m_{t,j,0.25}$ and $m_{t,j,0.75}$ are assessed in the lower graph and used in equation (4.17).

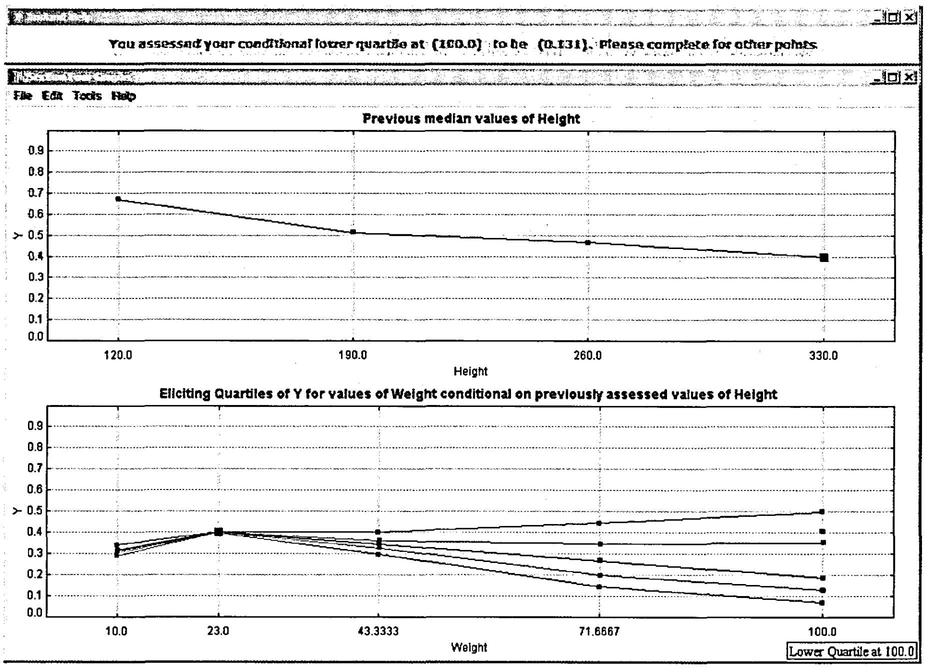


Figure 4.2: Assessments needed in the second phase for correlated covariates

This phase is very similar to the assessment of conditional quartiles in the GA method, as reviewed in the previous chapter, where incremented sets of medians of the same variable are used as conditioning sets for assessing conditional quartiles. However, in this phase previously assessed median values at knots for a different variable (R_s) are also taken into consideration when assessing conditional quartiles of R_t, where $s < t$.

One of the steps of the second phase is shown in Figure 4.2. In this step, the expert is asked to assess conditional quartiles $m_{t,j,0.25}$ and $m_{t,j,0.75}$, for $j = 1, \cdots, 4$, for different knots of the "Weight" variable in the lower graph. Some of the conditioning values are the previously assessed medians, $m^0_{s,0}, \cdots, m^0_{s,3}$, of the "Height" variable at all of its four knots. These are connected by the black line in the upper graph. The other conditioning values are the median, $m^0_{t,0}$ of the "Weight" variable at its reference knot (23.0).

Suggested conditional quartiles are computed by extrapolating from other quartile assessments in the same manner as in GA method; see the previous chapter. The middle (green) lines in the lower graph in Figure 4.2 represent these suggested values.

On finishing all phases of the assessment for this pair of explanatory variables, the user is asked about other correlated pairs, and the process starts again for the new pair, if there is one. The modified software outputs data in three different files, one containing the basic setup data, the second containing all assessments made by the expert, and the third containing the resulting mean vector and covariance matrix of the hyperparameter vector, which are in a form suitable for further Bayesian analysis.

4.2.3 On the positive-definiteness of the elicited covariance matrix

After generalizing GA's method, as shown in Section 4.2.1 above, to estimate the variance-covariance matrix of $\underline{\beta}_s$ and $\underline{\beta}_t$, we ended up with

$$V = \mathrm{Var}((\underline{\beta}'_s \quad \underline{\beta}'_t)'|\alpha) \equiv \begin{pmatrix} \Sigma_{s,s} & V_{s,t} \\ V'_{s,t} & V_{t,t} \end{pmatrix}, \tag{4.20}$$

where $\Sigma_{s,s}$ is estimated using the method of GA. Now $V_{t,t} \neq \Sigma_{t,t}$. Instead,

$$V_{t,t} = \Sigma^*_{t,t} + V'_{s,t}\Sigma^{-1}_{s,s}V_{s,t},$$

with

$$\Sigma^*_{t,t} \equiv \mathrm{Var}(\underline{\beta}_t|\underline{\beta}_s, \alpha) = \begin{cases} D_t^{-1}\Lambda_{t,t}(D_t^{-1})', & \text{for } t = 1, 2, \cdots, m, \\ \Lambda_{t,t}, & \text{for } t = m+1, m+2, \cdots, m+n. \end{cases}$$

To check the positive-definiteness of the variance-covariance matrix $\mathrm{Var}((\underline{\beta}'_s \quad \underline{\beta}'_t)'|\alpha)$, we proceed as follows. First, we will show that V in (4.20) is positive-definite. Then we will find a transformation to replace the sub-matrix $V_{t,t}$ of V by the directly elicited unconditional variance matrix $\Sigma_{t,t}$. This transformation replaces V with a new matrix, say A, which will be shown to be positive-definite.

Now, in the matrix V, we have:

- From (4.4) and (4.6) $\Sigma_{s,s}$ is positive-definite, since $\Lambda_{s,0}$ is positive-definite as shown in the previous chapter, and from (4.3), D_s are lower triangular for $s = 1, 2, \cdots, m$.
- $\Sigma^*_{t,t}$ is positive-definite since it was computed in the manner of $\Sigma_{s,s}$, above.

- Since $\Sigma_{s,s}$ is positive-definite, so is $\Sigma_{s,s}^{-1}$, and $V_{s,t}'\Sigma_{s,s}^{-1}V_{s,t}$ is sure to be positive semi-definite. In fact, $\forall \underline{x} \neq \underline{0}$,

$$\underline{x}'\left(V_{s,t}'\Sigma_{s,s}^{-1}V_{s,t}\right)\underline{x} = \left(V_{s,t}\underline{x}\right)'\Sigma_{s,s}^{-1}\left(V_{s,t}\underline{x}\right) \geq 0,$$

  from the positive-definiteness of $\Sigma_{s,s}^{-1}$.

- $V_{t,t}$ is thus the sum of a positive-definite and a positive semi-definite matrix, hence $V_{t,t}$ is positive-definite.

For V to be positive-definite, we use the Schurr complement (Abadir and Magnus, 2005, p.228) to show that

$$\left(\Sigma_{t,t}^{*} + V_{s,t}'\Sigma_{s,s}^{-1}V_{s,t}\right) - V_{s,t}'\Sigma_{s,s}^{-1}V_{s,t} = \Sigma_{t,t}^{*}$$

is positive-definite, which is the case.

We believe that the submatrix $\Sigma_{t,t}$ is better than $V_{t,t}$ as an estimate of $\text{Var}(\underline{\beta}_t|\alpha)$. Note that $V_{t,t}$ was computed by conditioning on both α and $\underline{\beta}_s$.

Our aim now is to introduce a new matrix, A, conformally partitioned as,

$$A = \begin{pmatrix} \Sigma_{s,s} & A_{s,t} \\ A_{s,t}' & \Sigma_{t,t} \end{pmatrix},$$

to replace V, where we believe A will generally be a better estimate of the variance-covariance matrix of $(\underline{\beta}_s' \quad \underline{\beta}_t')'|\alpha$.

To this end, put

$$B = \begin{pmatrix} I & O \\ O & \Sigma_{t,t}^{\frac{1}{2}}V_{t,t}^{-\frac{1}{2}} \end{pmatrix},$$

and take $A = BVB'$. Then

$$A = \begin{pmatrix} I & O \\ O & \Sigma_{t,t}^{\frac{1}{2}} V_{t,t}^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \Sigma_{s,s} & V_{s,t} \\ V'_{s,t} & V_{t,t} \end{pmatrix} \begin{pmatrix} I & O \\ O & V_{t,t}^{-\frac{1}{2}} \Sigma_{t,t}^{\frac{1}{2}} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{s,s} & V_{s,t} V_{t,t}^{-\frac{1}{2}} \Sigma_{t,t}^{\frac{1}{2}} \\ \Sigma_{t,t}^{\frac{1}{2}} V_{t,t}^{-\frac{1}{2}} V'_{s,t} & \Sigma_{t,t}^{\frac{1}{2}} \left(V_{t,t}^{-\frac{1}{2}} V_{t,t} V_{t,t}^{-\frac{1}{2}} \right) \Sigma_{t,t}^{\frac{1}{2}} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{s,s} & A_{s,t} \\ A'_{s,t} & \Sigma_{t,t} \end{pmatrix},$$

with $A_{s,t} = V_{s,t} V_{t,t}^{-\frac{1}{2}} \Sigma_{t,t}^{\frac{1}{2}}$. We next investigate whether A is necessarily positive-definite.

Since $\Sigma_{s,s}$ and $\Sigma_{t,t}$ are positive-definite, A is positive-definite, using the Schurr complement again, if and only if

$$\Sigma_{s,s} - A_{s,t} \Sigma_{t,t}^{-1} A'_{s,t},$$

is positive-definite. But

$$\begin{aligned} \Sigma_{s,s} - A_{s,t} \Sigma_{t,t}^{-1} A'_{s,t} &= \Sigma_{s,s} - \left(V_{s,t} V_{t,t}^{-\frac{1}{2}} \Sigma_{t,t}^{\frac{1}{2}} \right) \Sigma_{t,t}^{-1} \left(\Sigma_{t,t}^{\frac{1}{2}} V_{t,t}^{-\frac{1}{2}} V'_{s,t} \right) \\ &= \Sigma_{s,s} - V_{s,t} V_{t,t}^{-\frac{1}{2}} \left(\Sigma_{t,t}^{\frac{1}{2}} \Sigma_{t,t}^{-1} \Sigma_{t,t}^{\frac{1}{2}} \right) V_{t,t}^{-\frac{1}{2}} V'_{s,t} \\ &= \Sigma_{s,s} - V_{s,t} V_{t,t}^{-1} V'_{s,t}. \end{aligned}$$

Thus $\Sigma_{s,s} - A_{s,t} \Sigma_{t,t}^{-1} A'_{s,t}$ is positive-definite from the positive-definiteness of the matrix V. It can be simply seen also from the matrix equation $A = BVB'$ that A is positive-definite since V is positive-definite, and B is non singular (Abadir and Magnus, 2005, p.221).

Now, although each variance-covariance matrix A for any pair of correlated vectors of coefficients, has been shown to be positive-definite, some extra conditions must be imposed for the whole variance-covariance matrix Λ in (4.5) to be positive-definite. For that, a structural elicitation method should be applied to the whole matrix. In which case, a huge number of conditional assessments will be needed to inter-relate all pairs, even though many of them may be slightly correlated. This puts an extra assessment burden on the expert and there may be no real gain.

However, the power of this method is apparent when only one pair of vectors is highly correlated. Another good situation for its application is when there are only a few correlated pairs and the whole variance-covariance matrix can be re-arranged so that these are 2×2 partitioned matrices on the main diagonal and off-diagonal covariance matrices are zeros. The whole matrix is sure to be positive-definite in this case. The expert should, of course, be willing to use the proposed method to elicit each main diagonal 2×2 partitioned matrix by assessing all the required conditional quartiles.

Although the variance-covariance matrix cannot be guaranteed to be positive-definite when there are many correlated pairs of vectors, it can still be checked for positive-definiteness. The expert may be asked to review her assessments, if needed, to fulfil the property. However, we propose another elicitation method in the next section that not only fulfils the positive-definiteness of Λ in (4.5), but which also requires a smaller number of assessments.

We also combine the two methods to give a flexible approach in which the expert assesses the variance-covariance matrix for the highest correlated pair of vectors using the current method. She then assesses the relative correlation of other pairs of vectors in comparison with the most highly correlated pair of vectors. These relative correlations are scaled to give the whole matrix. The details of this approach are presented in the next two sections.

4.3 Another elicitation method for the variance-covariance matrix of correlated coefficients

One possible drawback of the elicitation method proposed in Section 4.2 is that the number of conditional quartiles that the expert must assess will become uncomfortably large, if many pairs of covariates are thought to be correlated. For such situations, another method is proposed here to elicit the off-diagonal covariance matrices. It uses a small number of coefficients to reflect the pattern of correlation between pairs of vectors and this reduces the number of assessments that are required. At the same time, the assessments can be used to induce all the elements of the covariance matrix and, under suitable conditions, the resulting variance-covariance matrix is positive-definite. These conditions can be translated into allowable ranges shown to the expert on an interactive graph; the expert will be asked to restrict her assessments so that conditional medians lie inside these ranges. The mathematical details of the proposed method are given in Sections 4.3.1 and 4.3.2 below. The required assessments for the equations in these two sections are discussed in detail in Section 4.3.3, where the use of the interactive software to obtain the conditional medians is also discussed.

4.3.1 The case of two vectors of correlated coefficients

To reduce the number of required assessments for estimating the covariance matrix of any correlated vectors $\underline{\beta}_s$ and $\underline{\beta}_t$, we assume a fixed pattern of correlation between the elements of these two vectors. We must make some simplifying assumptions about the correlation between these vectors. If the variance-covariance matrix of $\underline{\beta}_s$ were the identity matrix and the same were true for $\underline{\beta}_t$, then it might be reasonable to assume that any component of $\underline{\beta}_s$ had the same correlation with each component of $\underline{\beta}_t$, and vice-versa. Of course, the variances of $\underline{\beta}_s$ and $\underline{\beta}_t$ are not identity matrices. Instead, we transform $\underline{\beta}_s$ and $\underline{\beta}_t$ into $\underline{\xi}_s$ and $\underline{\xi}_t$, respectively, for which, $\text{Var}(\underline{\xi}_s) = I_{\delta(s)}$, $\text{Var}(\underline{\xi}_t) = I_{\delta(t)}$. Then we assume that the correlation coefficient between any element $\xi_{s,i}$ $(i = 1, 2, \cdots, \delta(s))$ of $\underline{\xi}_s$ and any element $\xi_{t,j}$ $(j = 1, 2, \cdots, \delta(t))$ of $\underline{\xi}_t$ is a fixed number, $c_{s,t}$. We elicit the value of $c_{s,t}$ using a small number of conditional assessments.

The matrices $\text{Var}(\underline{\beta}_s) = \Sigma_{s,s}$ and $\text{Var}(\underline{\beta}_t) = \Sigma_{t,t}$ may be estimated using the method of GA that was reviewed in Chapter 3. These matrices are positive-definite, so there exist non-singular matrices A and B such that

$$A\Sigma_{s,s}A' = I_{\delta(s)},$$

$$B\Sigma_{t,t}B' = I_{\delta(t)}.$$

In fact, we take A and B as the inverse of the two unique symmetric positive-definite square roots that can be obtained from the eigenvalue decomposition of $\Sigma_{s,s}$ and $\Sigma_{t,t}$, respectively, i.e.

$$A = \Sigma_{s,s}^{-\frac{1}{2}},$$

$$B = \Sigma_{t,t}^{-\frac{1}{2}}.$$

Let $\underline{\xi}_s = A\underline{\beta}_s$ and $\underline{\xi}_t = B\underline{\beta}_t$, then

$$\underline{\xi}_s \sim \text{MVN}(A\underline{b}_s, I_{\delta(s)}),$$

$$\underline{\xi}_t \sim \text{MVN}(B\underline{b}_t, I_{\delta(t)}).$$

We assume that

$$\text{Cov}(\underline{\xi}_s, \underline{\xi}_t) = C_{s,t} = \begin{pmatrix} c_{s,t} & \cdots & c_{s,t} \\ \vdots & \cdots & \vdots \\ c_{s,t} & \cdots & c_{s,t} \end{pmatrix}_{\delta(s)\times\delta(t)} = c_{s,t}\ \underline{1}_{\delta(s)}\ \underline{1}'_{\delta(t)}. \tag{4.21}$$

So that

$$\begin{pmatrix} \underline{\xi}_s \\ \underline{\xi}_t \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} A\underline{b}_s \\ B\underline{b}_t \end{pmatrix}, \begin{pmatrix} I_{\delta(s)} & C_{s,t} \\ C'_{s,t} & I_{\delta(t)} \end{pmatrix}\right). \tag{4.22}$$

Assume further that

$$E(\underline{\xi}_t | \underline{\xi}_s = A\underline{b}_s + \underline{\eta}_s) = B\underline{b}_t + \underline{\theta}_t, \tag{4.23}$$

where

$\underline{\eta}_s = (\eta_s \quad \eta_s \quad \cdots \quad \eta_s)' = \eta_s\ \underline{1}$, for an arbitrary chosen value $\eta_s > 0$,

$\underline{\theta}_t = (\theta_t \quad \theta_t \quad \cdots \quad \theta_t)' = \theta_t\ \underline{1}$.

But it is known, from the conditional multivariate normal theory, that

$$E(\underline{\xi}_t | \underline{\xi}_s = A\underline{b}_s + \underline{\eta}_s) = B\underline{b}_t - C'_{s,t} I^{-1}_{\delta(s)} \left[A\underline{b}_s - (A\underline{b}_s + \underline{\eta}_s)\right] = B\underline{b}_t + C'_{s,t}\underline{\eta}_s. \tag{4.24}$$

Thus, from (4.23) and (4.24), we get

$$\underline{\theta}_t = C'_{s,t}\underline{\eta}_s. \tag{4.25}$$

The expert will be asked to determine the conditional mean of $\underline{\xi}_t$ given a specific value of $\underline{\xi}_s$, hence the value of θ_t will be computed from the expert's assessment of $E(\underline{\xi}_t | \underline{\xi}_s)$. In fact, the expert assesses only conditional medians of Y, which are then transformed, under normality assumption, into conditional means of the slopes of the piecewise-linear relation, or bar heights for factors, as will be detailed in Section 4.3.3.

From (4.25), the value of $c_{s,t}$ is simply estimated as

$$c_{s,t} = \frac{\theta_t}{\delta(s) \times \eta_s}. \tag{4.26}$$

It will be shown that $\mathrm{Var}(\underline{\beta}'_s \;\; \underline{\beta}'_t)'$ is a positive-definite matrix if, and only if,

$$|c_{s,t}| < \frac{1}{\sqrt{\delta(s) \times \delta(t)}}. \tag{4.27}$$

Using (4.26), this condition can be written in terms of θ_t, as

$$|\theta_t| < \eta_s \sqrt{\frac{\delta(s)}{\delta(t)}}. \tag{4.28}$$

To prove (4.27), note that

$$V \equiv \mathrm{Var}\begin{pmatrix} \underline{\beta}_s \\ \\ \underline{\beta}_t \end{pmatrix} = \mathrm{Var}\begin{pmatrix} \Sigma_{s,s}^{\frac{1}{2}} \; \underline{\xi}_s \\ \\ \Sigma_{t,t}^{\frac{1}{2}} \; \underline{\xi}_t \end{pmatrix} = \begin{pmatrix} \Sigma_{s,s} & \Sigma_{s,s}^{\frac{1}{2}} \; C_{s,t} \; \Sigma_{t,t}^{\frac{1}{2}} \\ \\ \Sigma_{t,t}^{\frac{1}{2}} \; C'_{s,t} \; \Sigma_{s,s}^{\frac{1}{2}} & \Sigma_{t,t} \end{pmatrix}.$$

Since $\Sigma_{s,s}$ and $\Sigma_{t,t}$ are both positive-definite matrices, V is positive-definite, using the Schurr complement, if and only if

$$\Sigma_{s,s} - \Sigma_{s,s}^{\frac{1}{2}} \; C_{s,t}C'_{s,t} \; \Sigma_{s,s}^{\frac{1}{2}} = \Sigma_{s,s}^{\frac{1}{2}} \; (I_{\delta(s)} - C_{s,t}C'_{s,t}) \; \Sigma_{s,s}^{\frac{1}{2}} \tag{4.29}$$

is positive-definite, or equivalently

$$\Sigma_{t,t} - \Sigma_{t,t}^{\frac{1}{2}} \; C'_{s,t}C_{s,t} \; \Sigma_{t,t}^{\frac{1}{2}} = \Sigma_{t,t}^{\frac{1}{2}} \; (I_{\delta(t)} - C'_{s,t}C_{s,t}) \; \Sigma_{t,t}^{\frac{1}{2}} \tag{4.30}$$

is positive-definite.

In other words, from (4.29) or (4.30), V is positive-definite if and only if

$$\mathrm{Var}(\underline{\xi}_s, \underline{\xi}_t) = \begin{pmatrix} I_{\delta(s)} & C_{s,t} \\ \\ C'_{s,t} & I_{\delta(t)} \end{pmatrix}$$

is positive-definite.

Let,

$$F \equiv I_{\delta(s)} - C_{s,t}C'_{s,t},$$

$$G \equiv \frac{1}{\delta(s)} \; \underline{1}_{\delta(s)} \; \underline{1}'_{\delta(s)},$$

and note that G is a symmetric idempotent matrix with $\text{rank}(G) = \text{trace}(G) = 1$. Then F can be written as

$$
\begin{aligned}
F &= I_{\delta(s)} - C_{s,t}\ C'_{s,t} \\
&= I_{\delta(s)} - \delta(s)\delta(t)c_{s,t}^2 G \\
&= I_{\delta(s)} - G + G - \delta(s)\delta(t)c_{s,t}^2 G \\
&= (I_{\delta(s)} - G) + (1 - \delta(s)\delta(t)c_{s,t}^2)G \\
&\equiv \alpha_1(I_{\delta(s)} - G) + \alpha_2 G,
\end{aligned}
$$

with

$$
\begin{aligned}
\alpha_1 &= 1, \\
\alpha_2 &= (1 - \delta(s)\delta(t)c^2).
\end{aligned}
$$

As both G and $(I_{\delta(s)} - G)$ are idempotent matrices summing up to $I_{\delta(s)}$, the eigenvalues of F are $\alpha_1 = 1$, with multiplicity $\text{rank}(I_{\delta(s)} - G) = \text{trace}(I_{\delta(s)} - G) = \delta(s) - 1$ and $\alpha_2 = (1 - \delta(s)\delta(t)c_{s,t}^2)$ with multiplicity one. Hence, the necessary and sufficient condition for the matrix F, and consequently for V, to be positive-definite is that both α_1 and α_2 must be positive. Since $\alpha_1 = 1$, the matrix V is positive-definite if and only if $(1 - \delta(s)\delta(t)c_{s,t}^2) > 0$, which gives the condition (4.27).

The same condition can also be deduced from the quadratic form of the matrix F. First, recall, from Cauchy's inequality, that

$$
\left(\sum_{i=1}^{n} x_i\right)^2 \leq n \sum_{i=1}^{n} x_i^2.
$$

Then $\forall \underline{x} \neq \underline{0}$,

$$
\begin{aligned}
\underline{x}' F \underline{x} &= \sum_{i=1}^{\delta(s)} (1 - \delta(t) c_{s,t}^2) x_i^2 + \sum_{i \neq j} (-\delta(t) c_{s,t}^2) x_i x_j \\
&= \sum_{i=1}^{\delta(s)} x_i^2 - \delta(t) c_{s,t}^2 \left[\sum_{i=1}^{\delta(s)} x_i^2 + \sum_{i \neq j} x_i x_j \right] \\
&= \sum_{i=1}^{\delta(s)} x_i^2 - \delta(t) c_{s,t}^2 \left[\sum_{i=1}^{\delta(s)} x_i \right]^2 \\
&\geq \sum_{i=1}^{\delta(s)} x_i^2 - \delta(t) \delta(s) c_{s,t}^2 \sum_{i=1}^{\delta(s)} x_i^2 \\
&= (1 - \delta(s) \delta(t) c_{s,t}^2) \sum_{i=1}^{\delta(s)} x_i^2 .
\end{aligned}
$$

Since $\sum_{i=1}^{\delta(s)} x_i^2 > 0$, F is positive-definite if and only if $(1 - \delta(s)\delta(t)c_{s,t}^2) > 0$.

4.3.2 The case of various vectors of correlated coefficients

When there are more than two correlated explanatory variables, the method given in Section 4.3.1 is still valid. We next obtain a set of $n(n-1)/2$ conditions that are necessary and sufficient for the full variance-covariance matrix to be positive-definite, for any number $n > 2$ of correlated explanatory variables. The number of assessments required for eliciting a variance-covariance matrix using this proposed method when $n > 2$ is only $n(n-1)/2$.

The case of $n = 2$ has been considered already. For $n = k > 2$ explanatory variables, let

$$
V_i \equiv \text{Var} \begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_i \end{pmatrix}, \qquad \underline{e}_i \equiv E \begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_i \end{pmatrix}, \qquad \text{for } i = 1, 2, ..., k.
$$

Assume that V_i, $i = 1, 2, ..., k-1$, have been obtained and that they are known to be positive-definite matrices.

Let

$$\underline{\xi}_i \equiv V_i^{-\frac{1}{2}} \begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_i \end{pmatrix}, \quad \text{for } i = 1, 2, ..., k-1,$$

$$\underline{\xi}_k \equiv \Sigma_{k,k}^{-\frac{1}{2}} \ \underline{\beta}_k,$$

with

$$\Sigma_{k,k} = \text{Var}(\underline{\beta}_k).$$

We assume that

$$\text{Cov}(\underline{\xi}_{k-1}, \underline{\xi}_k) \equiv C_k \equiv \begin{pmatrix} C_{k,1} \\ C_{k,2} \\ \vdots \\ C_{k,k-1} \end{pmatrix}, \tag{4.31}$$

where C_k is a matrix of order $(\sum_{i=1}^{k-1} \delta(i)) \times \delta(k)$, and that each $C_{k,i}$ is a submatrix of order $\delta(i) \times \delta(k)$, taking the form

$$C_{k,i} \equiv \begin{pmatrix} c_{k,i} & \cdots & c_{k,i} \\ \vdots & \cdots & \vdots \\ c_{k,i} & \cdots & c_{k,i} \end{pmatrix}, \quad \text{for } i = 1, 2, ..., k-1. \tag{4.32}$$

Then

$$\begin{pmatrix} \underline{\xi}_{k-1} \\ \underline{\xi}_k \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} V_{k-1}^{-\frac{1}{2}} \ \underline{e}_{k-1} \\ \Sigma_{k,k}^{-\frac{1}{2}} \ \underline{b}_k \end{pmatrix}, \begin{pmatrix} I_{\sum_{i=1}^{k-1} \delta(i)} & C_k \\ C_k' & I_{\delta(k)} \end{pmatrix} \right). \tag{4.33}$$

Now suppose that

$$E(\underline{\xi}_k | \underline{\xi}_i = V_i^{-\frac{1}{2}} \ \underline{e}_i + \underline{\tau}_{k,i}) = \Sigma_{k,k}^{-\frac{1}{2}} \ \underline{b}_k + \underline{\theta}_{k,i}, \quad \text{for } i = 1, 2, ..., k-1, \tag{4.34}$$

where

$\underline{\tau}_{k,i} = (\underline{\eta}_{k,1}' \quad \underline{\eta}_{k,2}' \quad \cdots \quad \underline{\eta}_{k,i}')'$,

$\underline{\eta}_{k,j} = (\eta_{k,j} \quad \eta_{k,j} \quad \cdots \quad \eta_{k,j})' = \eta_{k,j} \ \underline{1}$, $j = 1, 2, ..., i$, for arbitrary chosen $\eta_{k,j} > 0$,

$$\underline{\theta}_{k,i} = (\theta_{k,i} \quad \theta_{k,i} \quad \cdots \quad \theta_{k,i})' = \theta_{k,i} \ \underline{1}.$$

The process will consist of $k-1$ steps, at the i^{th} step, an elicited value of $E(\underline{\xi}_k | \underline{\xi}_i = V_i^{-\frac{1}{2}} \ \underline{e}_i + \underline{\tau}_{k,i})$ will be obtained. This can be done by asking the expert to assess conditional median values of Y that can be transformed, under the normality assumption, to conditional means of the slopes of the piecewise-linear relation, or bar heights for factors, as will be discussed in Section 4.3.3. We can then obtain the conditional medians $E(\underline{\xi}_k | \underline{\xi}_i = V_i^{-\frac{1}{2}} \ \underline{e}_i + \underline{\tau}_{k,i})$ in (4.34) from $E(\underline{\beta}_k | \underline{\beta}_1, \underline{\beta}_2, \cdots, \underline{\beta}_i)$. The conditional values of $\underline{\beta}_1, \underline{\beta}_2, \cdots, \underline{\beta}_i$, will be displayed through a set of i graphs, each of which gives a value for a different $\underline{\beta}_j$, $j = 1, 2, ..., i$.

Moreover, from the conditional multivariate normal theory, equation (4.33) gives

$$E(\underline{\xi}_k | \underline{\xi}_i = V_i^{-\frac{1}{2}} \ \underline{e}_i + \underline{\tau}_{k,i}) = \Sigma_{k,k}^{-\frac{1}{2}} \ \underline{b}_k + (C'_{k,1} \quad C'_{k,2} \quad \cdots \quad C'_{k,i}) I_{\sum_{j=1}^{i} \delta(j)} \underline{\tau}_{k,i}. \tag{4.35}$$

Then, from (4.34) and (4.35), we get

$$\underline{\theta}_{k,i} = (C'_{k,1} \quad C'_{k,2} \quad \cdots \quad C'_{k,i}) \underline{\tau}_{k,i}. \tag{4.36}$$

Hence, after finishing the $k-1$ steps, the following system of equations can be formed

$$\begin{aligned}
\theta_{k,1} &= \delta(1) c_{k,1} \eta_{k,1}, \\
\theta_{k,2} &= \delta(1) c_{k,1} \eta_{k,1} + \delta(2) c_{k,2} \eta_{k,2}, \\
&\vdots \\
\theta_{k,k-1} &= \delta(1) c_{k,1} \eta_{k,1} + \delta(2) c_{k,2} \eta_{k,2} + \cdots + \delta(k-1) c_{k,k-1} \eta_{k,k-1}.
\end{aligned}$$

To solve for $c_{k,i}$, $i = 1, 2, ..., k-1$, the system can be written as

$$\Omega \begin{pmatrix} c_{k,1} \\ c_{k,2} \\ \vdots \\ c_{k,k-1} \end{pmatrix} = \begin{pmatrix} \theta_{k,1} \\ \theta_{k,2} \\ \vdots \\ \theta_{k,k-1} \end{pmatrix}, \tag{4.37}$$

where

$$\Omega = \begin{pmatrix} \delta(1)\eta_{k,1} & 0 & \cdots & 0 \\ \delta(1)\eta_{k,1} & \delta(2)\eta_{k,2} & 0 & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \delta(1)\eta_{k,1} & \delta(2)\eta_{k,2} & \cdots & \delta(k-1)\eta_{k,k-1} \end{pmatrix}. \tag{4.38}$$

Provided that $\eta_{k,i} \neq 0$, $\forall\ i = 1, 2, ..., k-1$, the matrix Ω is non-singular and hence

$$\begin{pmatrix} c_{k,1} \\ c_{k,2} \\ \vdots \\ c_{k,k-1} \end{pmatrix} = \Omega^{-1} \begin{pmatrix} \theta_{k,1} \\ \theta_{k,2} \\ \vdots \\ \theta_{k,k-1} \end{pmatrix}. \tag{4.39}$$

Now, the variance-covariance matrix V_k can be estimated as follows:

$$V_k \equiv \mathrm{Var} \begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_k \end{pmatrix} = \mathrm{Var} \begin{pmatrix} V_{k-1}^{\frac{1}{2}}\ \underline{\xi}_{k-1} \\ \Sigma_{k,k}^{\frac{1}{2}}\ \underline{\xi}_k \end{pmatrix} = \begin{pmatrix} V_{k-1} & V_{k-1}^{\frac{1}{2}}\ C_k\ \Sigma_{k,k}^{\frac{1}{2}} \\ \Sigma_{k,k}^{\frac{1}{2}}\ C_k'\ V_{k-1}^{\frac{1}{2}} & \Sigma_{k,k} \end{pmatrix}. \tag{4.40}$$

We define the matrices $\Sigma_{i,k}$, for $i = 1, 2, \cdots, k-1$, that conformally partition $V_{k-1}^{\frac{1}{2}}\ C_k\ \Sigma_{k,k}^{\frac{1}{2}}$ as

$$\begin{pmatrix} \Sigma_{1,k} \\ \Sigma_{2,k} \\ \vdots \\ \Sigma_{k-1,k} \end{pmatrix} \equiv V_{k-1}^{\frac{1}{2}} \begin{pmatrix} C_{k,1} \\ C_{k,2} \\ \vdots \\ C_{k,k-1} \end{pmatrix} \Sigma_{k,k}^{\frac{1}{2}} = V_{k-1}^{\frac{1}{2}}\ C_k\ \Sigma_{k,k}^{\frac{1}{2}}. \tag{4.41}$$

Following the same steps as in the case $n = 2$, and since V_{k-1} and $\Sigma_{k,k}$ are positive-definite matrices, equations similar to (4.29) and (4.30) show that V_k is positive-definite if and only if the matrix

$$\begin{pmatrix} I_{\sum_{i=1}^{k-1}\delta(i)} & C_k \\ C_k' & I_{\delta(k)} \end{pmatrix}$$

is positive-definite. Putting

$$F_k \equiv I_{\delta(k)} - C_k' C_k$$
$$= I_{\delta(k)} - \sum_{i=1}^{k-1} C_{k,i}' C_{k,i},$$
$$G_k \equiv \frac{1}{\delta(k)} \, \underline{1}_{\delta(k)} \, \underline{1}_{\delta(k)}',$$

where G_k is idempotent of rank 1, it can be shown that

$$F_k = (I_{\delta(k)} - G_k) + \left(1 - \sum_{i=1}^{k-1} \delta(k)\delta(i) c_{k,i}^2\right) G_k$$

is positive-definite if and only if

$$\left(1 - \sum_{i=1}^{k-1} \delta(k)\delta(i) c_{k,i}^2\right) > 0. \tag{4.42}$$

This condition implies $k-1$ conditions for $c_{k,i}$, $i = 1, 2, ..., k-1$, of the form

$$|c_{k,i}| < \sqrt{\frac{1 - \sum_{j=1}^{i-1} \delta(k)\delta(j) c_{k,j}^2}{\delta(i) \times \delta(k)}}. \tag{4.43}$$

These $k-1$ conditions guarantee that the elicited matrix V_k in (4.40) is positive-definite, provided that V_{k-1} is positive-definite. Since V_2 is known to be positive-definite from Section 4.3.1, we can use mathematical induction to prove that the full variance-covariance matrix V_n is positive-definite, as follows. For any number ($n \geq 2$) of correlated vectors $\underline{\beta}_1$, $\underline{\beta}_2, \cdots, \underline{\beta}_n$, the whole matrix

$$V_n \equiv \mathrm{Var}\begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_n \end{pmatrix} \equiv \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,n} \\ \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{n,1} & \Sigma_{n,2} & \cdots & \Sigma_{n,n} \end{pmatrix}$$

is certain to be positive-definite if (a) the $n-1$ conditions in (4.43) hold and (b) V_{n-1} is positive-definite. This imposes an extra $i-1$ conditions on each matrix V_i, ($i = 2, \cdots, n-1$), so that each V_i is positive-definite. Then V_n is positive-definite under a number of $\sum_{k=2}^{n} k -$

$1 = \sum_{k=1}^{n-1} k = n(n-1)/2$ conditions of the form:

$$|c_{k,i}| < \sqrt{\frac{1 - \sum_{j=1}^{i-1} \delta(k)\delta(j)c_{k,j}^2}{\delta(i) \times \delta(k)}}, \quad \text{for } i = 1, 2, \cdots, k-1, \; k = 2, 3, \cdots, n. \tag{4.44}$$

Using these conditions, the range of each $\theta_{k,i}$, for $i = 1, 2, ..., k-1$, $k = 2, 3, ..., n$, can be computed and shown to the expert who can ensure that her assessed values fall within these ranges. This will guarantee that the estimated variance-covariance matrix is positive-definite. For $i = 1, 2, ..., k-1$, from (4.36) and (4.43), the range of $\theta_{k,i}$ is given by

$$\left[\left(\sum_{j=1}^{i-1} \delta(j) c_{k,j} \eta_{k,j} \right) \pm \eta_{k,i} \sqrt{\frac{\delta(i)}{\delta(k)} \left(1 - \sum_{j=1}^{i-1} \delta(k)\delta(j)c_{k,j}^2 \right)} \right]. \tag{4.45}$$

This formula for the allowable range of $\theta_{k,i}$ has a drawback: we cannot calculate these ranges until quite late in the assessment procedure, so the expert may sometimes be asked to revise assessments that she made some time earlier. Hence, we decided to find a different approach that gives a more direct range for each $\theta_{k,i}$, and which only asks the expert to modify recent assessments that she has made. At step i, when conditioning on the value of $\underline{\xi}_i$, the expert may be asked to modify the assessment she has made in step $i-1$, but she will not be asked to modify assessments she gave at stages before that. This can be formulated as follows. Instead of equation (4.34), let

$$E(\underline{\xi}_k | \underline{\xi}_i = V_i^{-\frac{1}{2}} \; \underline{e}_i + \underline{\tau}_{k,i}) = E(\underline{\xi}_k | \underline{\xi}_{i-1} = V_{i-1}^{-\frac{1}{2}} \; \underline{e}_{i-1} + \underline{\tau}_{k,i-1}) + \underline{\theta}_{k,i} \tag{4.46}$$

$$= \Sigma_{k,k}^{-\frac{1}{2}} \; \underline{b}_k + \sum_{j=1}^{i} \underline{\theta}_{k,j}, \quad \text{for } i = 1, 2, ..., k-1, \tag{4.47}$$

where

$\underline{\tau}_{k,i} = (\underline{\eta}'_{k,1} \quad \underline{\eta}'_{k,2} \quad \cdots \quad \underline{\eta}'_{k,i})'$,

$\underline{\eta}_{k,j} = (\eta_{k,j} \quad \eta_{k,j} \quad \cdots \quad \eta_{k,j})' = \eta_{k,j} \; \underline{1}$, $j = 1, 2, ..., i$, for arbitrary chosen $\eta_{k,j} > 0$,

$\underline{\theta}_{k,i} = (\theta_{k,i} \quad \theta_{k,i} \quad \cdots \quad \theta_{k,i})' = \theta_{k,i} \; \underline{1}$.

In this case, using standard results of conditional expectations, we get

$$\Sigma_{k,k}^{-\frac{1}{2}}\ \underline{b}_k + \sum_{j=1}^{i} \underline{\theta}_{k,j} = \Sigma_{k,k}^{-\frac{1}{2}}\ \underline{b}_k + C_k' \underline{\tau}_{k,i}, \quad \text{for } i = 1, 2, ..., k-1. \tag{4.48}$$

Then equation (4.36) becomes

$$\underline{\theta}_{k,i} = C_{k,i}' \underline{\eta}_{k,i}, \tag{4.49}$$

which gives

$$\theta_{k,i} = \delta(i) c_{k,i} \eta_{k,i}, \quad i = 1, 2, ..., k-1.$$

Hence

$$c_{k,i} = \frac{\theta_{k,i}}{\delta(i) \times \eta_{k,i}}, \quad i = 1, 2, ..., k-1. \tag{4.50}$$

The positive-definiteness of the whole variance-covariance matrix V_n is still guaranteed under the same conditions in (4.44). But the allowable range for each $\theta_{k,i}$ ($i = 1, 2, ..., k-1,$; $k = 2, 3, ..., n$) has the simplified form,

$$|\theta_{k,i}| < \eta_{k,i} \sqrt{\frac{\delta(i)}{\delta(k)} \left(1 - \sum_{j=1}^{i-1} \delta(k)\delta(j) c_{k,j}^2 \right)}. \tag{4.51}$$

This represents a simple range for $\theta_{k,i}$, in comparison with (4.45). The range in (4.51) depends only on the change $\eta_{k,i}$ in the ith variable, $\underline{\xi}_i$, not on the changes $\eta_{k,j}$ in all variables $\underline{\xi}_j$, $j = 1, \cdots, i-1$, as in (4.45).

4.3.3 Assessment tasks

- The current assessment tasks start after eliciting all variance matrices $\Sigma_{i,i}$ ($i = 1, 2, ..., k$).
- For any pair of correlated vectors $(\underline{\beta}_s, \underline{\beta}_t)$, we assume that

$$\text{Cov}(\underline{\beta}_s, \underline{\beta}_t) = \Sigma_{s,s}^{\frac{1}{2}} C_{s,t} \Sigma_{t,t}^{\frac{1}{2}}, \quad s \neq t, \tag{4.52}$$

  where $C_{s,t}$ is given in (4.21) and $\Sigma_{s,s}$ and $\Sigma_{t,t}$ are the variances of $\underline{\beta}_s$ and $\underline{\beta}_t$, respectively.
- The expert will be shown a panel that simultaneously displays two graphs and a slider (see Figure 4.3). For continuous covariates, the upper graph of the panel shows the piecewise-linear relation between Y and X_s. The slopes of the black (lower) curve

represent $\underline{b}_s = E(\underline{\beta}_s)$, while the slopes of the blue (upper) curve represent the change of $E(\underline{\beta}_s)$ by $\Sigma^{\frac{1}{2}}_{s,s}\underline{\eta}_s$, i.e. the slopes of the blue (upper) curve are $\underline{b}_s + \Sigma^{\frac{1}{2}}_{s,s}\underline{\eta}_s$. The black (lower) lines represent the expert's original median assessments but she is asked to suppose that the correct values are actually the blue (upper) lines. Given this information, the expert is asked to use the slider to change the position of the black (middle) curve in the lower panel so that it gives her new opinion about the median value that Y will take as X_t varies. The magnitude and direction of the change reflects the correlation between $\underline{\beta}_s$ and $\underline{\beta}_t$.

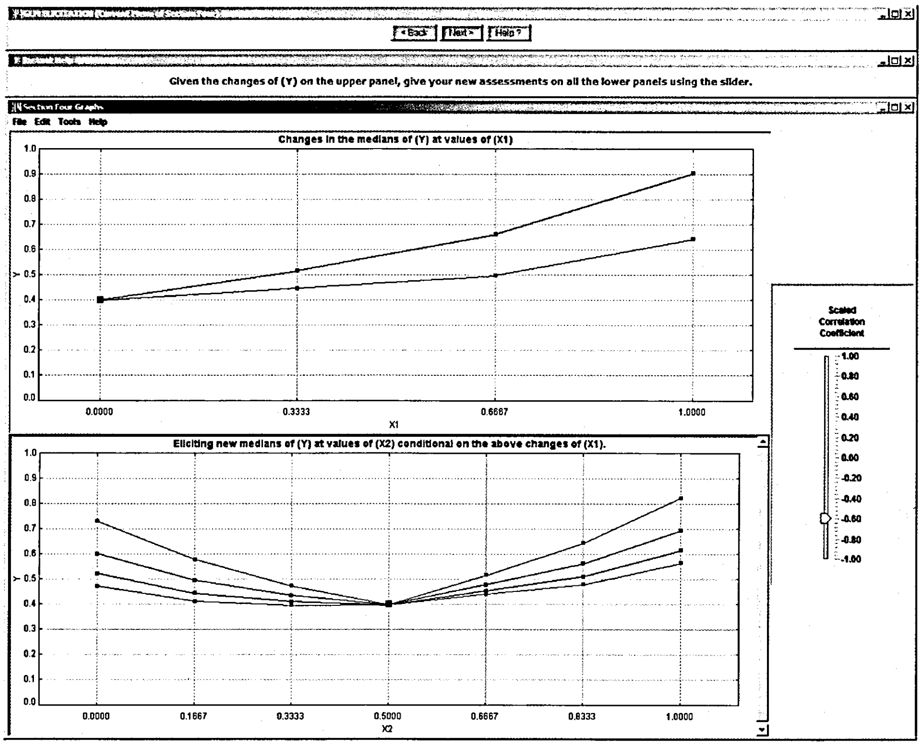


Figure 4.3: Assessments needed for two correlated variables

- The two red (outer) piecewise-linear curves in the lower panel of Figure 4.3 represent the allowable boundaries for the change of $\underline{\beta}_t$; these boundaries ensure that the resulting variance-covariance matrix is positive-definite. The boundaries are calculated from the condition given in equation (4.28). Moving the slider simultaneously changes the position of all the medians of Y in the lower panel. When the expert is happy with the new position of the curve on the lower panel, the corresponding value of the slider is used to compute $c_{s,t}$, as will be shown later.

- The expert is asked to assume that the slopes of X_s, in the upper panel of Figure 4.3, have changed from $\underline{b}_s$ to $\underline{b}_s + \Sigma_{s,s}^{\frac{1}{2}}\underline{\eta}_s$. Conditional on this information, she revises the slopes of X_t, in the lower panel, changing them from $\underline{b}_t$ to $\underline{b}_t + \Sigma_{t,t}^{\frac{1}{2}}\underline{\theta}_t$. The expert changes all the slopes simultaneously using the slider.

- The size of the change, η_s, in the conditioning variable, X_s, in the upper panel, is chosen such that the vertical distances between the two piecewise-linear curves in the upper graph do not exceed the upper quartile at any of the knots of X_s. This ensures that the new conditioning values $\underline{b}_s + \Sigma_{s,s}^{\frac{1}{2}}\underline{\eta}_s$ are not too far from $\underline{b}_s$, as they have to be values that the expert finds plausible. This choice is also not too close to $\underline{b}_s$, so it should prompt a measurable change in $\underline{b}_t$ in the lower panel of Figure 4.3.

- The software calculates medians to draw a piecewise-linear curve with slopes $\underline{b}_s + \Sigma_{s,s}^{\frac{1}{2}}\underline{\eta}_s$. For $i = 1, 2, \cdots, \delta(s)$, the median value of Y at each knot i, $m_{s,i,0.5}$, is changed to $m^*_{s,i,0.5}$, as follows.

  First, let

$$m^*_{s,0,0.5} = m_{s,0,0.5},$$

  and

$$d_{i,i-1} = r_{s,i} - r_{s,i-1}.$$

  Then, for $i = 1, 2, \cdots, \delta(s)$, we put

$$\begin{aligned}\frac{m^*_{s,i,0.5} - m^*_{s,i-1,0.5}}{d_{i,i-1}} &= b_{s,i} + \eta_s(\Sigma_{s,s}^{\frac{1}{2}})_i \\ &= \frac{m_{s,i,0.5} - m_{s,i-1,0.5}}{d_{i,i-1}} + \eta_s(\Sigma_{s,s}^{\frac{1}{2}})_i,\end{aligned}$$

  where $(\Sigma_{s,s}^{\frac{1}{2}})_i$ is the sum of the elements of the ith row of $\Sigma_{s,s}^{\frac{1}{2}}$.

  Hence,

$$m^*_{s,i,0.5} = m^*_{s,i-1,0.5} + m_{s,i,0.5} - m_{s,i-1,0.5} + \eta_s d_{i,i-1}(\Sigma_{s,s}^{\frac{1}{2}})_i. \tag{4.53}$$

If X_s is a factor, then

$$m^*_{s,i,0.5} = m_{s,i,0.5} + \eta_s(\Sigma^{\frac{1}{2}}_{s,s})_i. \tag{4.54}$$

In view of (4.53) and (4.54), η_s can be chosen as

$$\eta_s = \min_i \left(\frac{m_{s,i,0.75} - m_{s,i,0.5}}{\sum_{j=1}^{i} d_{j,j-1}(\Sigma^{\frac{1}{2}}_{s,s})_j} \right), \tag{4.55}$$

for continuous covariates. For factors, it can be chosen as

$$\eta_s = \min_i \left(\frac{m_{s,i,0.75} - m_{s,i,0.5}}{(\Sigma^{\frac{1}{2}}_{s,s})_i} \right). \tag{4.56}$$

- In order to draw the red (outer) boundaries in Figure 4.3, we require upper and lower bounds, $m^U_{t,i,0.5}$ and $m^L_{t,i,0.5}$. From (4.28), if X_t is a continuous covariate, we put

$$m^U_{t,i,0.5} = m^U_{t,i-1,0.5} + m_{t,i,0.5} - m_{t,i-1,0.5} + \eta_s\sqrt{\frac{\delta(s)}{\delta(t)}}d_{i,i-1}(\Sigma^{\frac{1}{2}}_{t,t})_i, \tag{4.57}$$

and

$$m^L_{t,i,0.5} = m^L_{t,i-1,0.5} + m_{t,i,0.5} - m_{t,i-1,0.5} - \eta_s\sqrt{\frac{\delta(s)}{\delta(t)}}d_{i,i-1}(\Sigma^{\frac{1}{2}}_{t,t})_i. \tag{4.58}$$

If X_s is a factor, we put

$$m^U_{t,i,0.5} = m_{t,i,0.5} + \eta_s\sqrt{\frac{\delta(s)}{\delta(t)}}(\Sigma^{\frac{1}{2}}_{t,t})_i, \tag{4.59}$$

and

$$m^L_{t,i,0.5} = m_{t,i,0.5} - \eta_s\sqrt{\frac{\delta(s)}{\delta(t)}}(\Sigma^{\frac{1}{2}}_{t,t})_i. \tag{4.60}$$

- Using the slider, in view of (4.27), the expert changes the value of $c_{s,t}$ between its two boundaries, $\pm 1/\sqrt{\delta(s) \times \delta(t)}$. To be interpretable by the expert, the slider presents a scaled range between -1 and 1 as a measure of correlation between $\underline{\beta}_s$ and $\underline{\beta}_t$. Hence

$$c_{s,t} = \text{The slider value } /\sqrt{\delta(s) \times \delta(t)}.$$

The corresponding new curve, say $m'_{t,i,0.5}$, is interactively changing with each movement of the slider. For continuous covariates, $m'_{t,i,0.5}$ is computed after $m'_{t,i-1,0.5}$ has been calculated:

$$m'_{t,i,0.5} = m'_{t,i-1,0.5} + m_{t,i,0.5} - m_{t,i-1,0.5} + c_{s,t}d_{i,i-1}(\Sigma^{\frac{1}{2}}_{t,t})_i. \tag{4.61}$$

For factors

$$m'_{t,i,0.5} = m_{t,i,0.5} + c_{s,t}(\Sigma_{t,t}^{\frac{1}{2}})_i. \tag{4.62}$$

When the expert is happy with the new position of the curve, the value of $c_{s,t}$ is used in (4.21) and (4.52) to calculate the covariances between $\underline{\beta}_s$ and $\underline{\beta}_t$.

- For $k > 2$ correlated vectors of coefficients, the process will consist of $k-1$ steps. At the i^{th} step, the expert will be asked to change the conditional medians of $(\underline{\beta}_k|\underline{\beta}_1, \underline{\beta}_2, \cdots, \underline{\beta}_i)$ by a value of $\theta_{k,i}$ given a set of i graphs, each of which shows a change with a different fixed value η_j for each $\underline{\beta}_j$, $j = 1, 2, ..., i$.

- However, we choose not to offer this general case as an option in the interactive software. Although it has been shown to have a consistent mathematical framework and adequate theoretical properties as proposed in Section 4.3.2, its practical implementation may raise some critical issues in the elicitation process. Conditioning on simultaneous changes in many graphs for different variables gives too much information for an expert to readily absorb. She may not be able to assess the direct conditional impact of these changes on the variable of concern.

- Another difficulty arises in choosing the different values η_j, $j = 1, 2, \cdots, i$, that control the change in the conditioning set used in step i. These values must be carefully specified so that the resulting simultaneous change represents a valid combination of values that is acceptable by the expert to condition on.

- A general problem in successive increment of variables in the conditioning set is that the allowable range of medians at the variable of concern gets tighter as we approach the last variable in the list. This problem is not only a practical one, but it has also been shown that variances, and hence covariances, of the last variables in the list are usually over estimated by the expert due to incremental conditioning (Garthwaite, 1994). These drawbacks constitute the motivation for the third elicitation method proposed in the next section.

4.4 A general flexible elicitation method for correlated coefficients

The aim here is to form an elicitation method suitable for GLMs that contain a large number of correlated vectors. We propose the following elicitation method as a promising new approach for eliciting the whole variance-covariance matrix. It uses only a small number of assessments that directly reflect the pattern of correlations between all pairs of vectors.

The method avoids the previously mentioned disadvantages of using incremented conditioning sets of variables. Instead, the method treats all variables symmetrically. As with the method proposed in Section 4.3.1, it assumes a fixed correlation structure for the elements of each pair of vectors. The current method differs from the generalization proposed in Section 4.3.2, in that it avoids incremented conditioning and assesses all covariances simultaneously.

The main idea is that the expert assesses the relative magnitudes of the average correlations between each pair of vectors. She is asked to ensure that these weights reflect the strength of the average correlation of each pair relative to each other pair. The expert need not be conscious of conditions that are required for mathematical coherence. Instead, the assessed relative weights will be scaled to ensure that the assessed variance-covariance matrix is positive-definite.

The current method can be used alone or together with one of the two methods proposed before in this chapter. In the latter case, the current method needs an assessment of the correlation of only one pair of vectors, then all other correlations are computed using the relative weights. This correlation assessment may be obtained using the method proposed in Section 4.2 or the method proposed in Section 4.3.1. With the latter method the expert might use a slider to adjust the slopes of one vector of a highly correlated pair.

In what follows, the method is introduced in detail and the scaling needed to obtain a positive-definite matrix will also be investigated.

Assuming that all the k covariates are correlated, let

$$\underline{\xi}_i = \Sigma_{i,i}^{-\frac{1}{2}} \underline{\beta}_i, \qquad i = 1, 2, \cdots, k, \tag{4.63}$$

then

$$\mathrm{Var}(\underline{\xi}_i) = I_{\delta(i)}, \qquad i = 1, 2, \cdots, k. \tag{4.64}$$

For all $i = 1, 2, \cdots, k$, $j = 1, 2, \cdots, k$, $i \neq j$, we assume that

$$\mathrm{Cov}(\underline{\xi}_i, \underline{\xi}_j) = C_{i,j} = \begin{pmatrix} c_{i,j} & \cdots & c_{i,j} \\ \vdots & \cdots & \vdots \\ c_{i,j} & \cdots & c_{i,j} \end{pmatrix}_{\delta(i)\times\delta(j)} = c_{i,j}\ \underline{1}_{\delta(i)}\ \underline{1}'_{\delta(j)}, \tag{4.65}$$

with $c_{j,i} = c_{i,j}$.

Then

$$\mathrm{Cov}(\underline{\beta}_i, \underline{\beta}_j) = \Sigma_{i,i}^{\frac{1}{2}} C_{i,j} \Sigma_{j,j}^{\frac{1}{2}}, \tag{4.66}$$

and hence

$$V \equiv \mathrm{Var}\begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_k \end{pmatrix} \equiv \Lambda_{\Sigma_{i,i}}^{\frac{1}{2}} C \Lambda_{\Sigma_{i,i}}^{\frac{1}{2}}, \tag{4.67}$$

where $\Lambda_{\Sigma_{i,i}}^{\frac{1}{2}}$ is a block-diagonal matrix with $\Sigma_{i,i}^{\frac{1}{2}}$ as the ith main diagonal block and

$$C \equiv \begin{pmatrix} I_{\delta(1)} & C_{1,2} & \cdots & C_{1,k} \\ C_{2,1} & I_{\delta(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & C_{k-1,k} \\ C_{k,1} & \cdots & C_{k,k-1} & I_{\delta(k)} \end{pmatrix}, \tag{4.68}$$

with

$$C_{j,i} = C'_{i,j}.$$

Since each $\Sigma_{i,i}^{\frac{1}{2}}$ is positive-definite, so is $\Lambda_{\Sigma_{i,i}}^{\frac{1}{2}}$. Hence, we can state that V in (4.67) is positive-definite if and only if C in (4.68) is positive-definite.

For $i = 1, 2, \cdots, k$, $j = 1, 2, \cdots, k$, $i \neq j$, let

$$c_{i,j} = cw_{i,j}, \tag{4.69}$$

where $w_{i,j}$ are the relative weights to be assessed from the expert and $c > 0$ is a fixed scaling constant that adjusts to ensure that C is positive-definite.

The main assessment task with this method consists of one dialogue box. An example is shown in Figure 4.4. The expert assesses the relative magnitudes (weights) and signs of different correlations between all pairs of vectors. Since the correlation matrix must be symmetric, we just require the elements below the main diagonal to be assessed. Hence, when there are n vectors of coefficients, we require $n(n-1)/2$ assessments for this stage. The main diagonal elements are necessarily set equal to ones, as C is a correlation matrix.

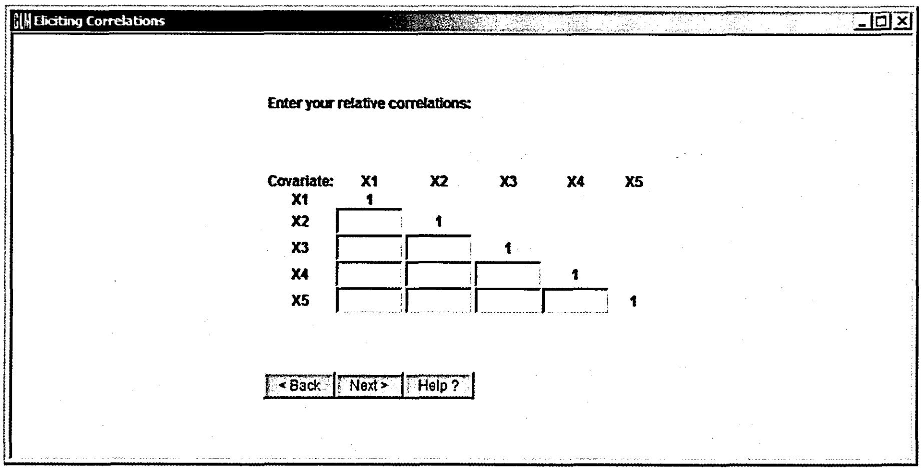


Figure 4.4: Assessments needed for five correlated variables

The relative weights that are assessed in this task need not be coherent correlation coefficients. For example, they are not necessarily restricted to be between -1 and 1. Instead, any assessed numbers are accepted; they must simply reflect the magnitude of the correlation between any pair of vectors relative to other pairs. Negative values are allowed and are appropriate when an expert believes a correlation is negative. The expert is asked to assess a single weight for each pair of vectors. The weight should reflect her opinion about the average correlation between all pairs of elements in that pair of vectors.

The relative weights assessed in Figure 4.4 will be denoted by $w^*_{i,j}$, where $w^*_{i,j}$ corresponds to the fixed average correlation between all elements of $\underline{\beta}_i$ and $\underline{\beta}_j$. The expert is asked to ensure that the relative magnitudes of $w^*_{i,j}$, $i = 1, 2, \cdots, k$, $j = 1, 2, \cdots, k$, $i > j$, model her opinion about the relative correlation of each pair compared to the others. As mentioned before, $w^*_{i,j}$ will be scaled later to attain mathematical coherent values of correlations.

For mathematical simplicity, we use the weights, $w_{i,j}$, of correlations between $\underline{\xi}_i$ and $\underline{\xi}_j$ when investigating the conditions required for the scaling constant c in (4.69). However, we assess the weights $w^*_{i,j}$ in terms of $\underline{\beta}_i$ and $\underline{\beta}_j$, as the expert cannot think about correlations between the transformed vectors $\underline{\xi}_i$ and $\underline{\xi}_j$. Hence, we need an explicit relationship between $w_{i,j}$ and $w^*_{i,j}$. We obtain one as follows.

For $i = 1, 2, \cdots, k$, $j = 1, 2, \cdots, k$, $i > j$, let

$$c^*_{i,j} = cw^*_{i,j} \tag{4.70}$$

be the scaled average correlation between $\underline{\beta}_i$ and $\underline{\beta}_j$. Then

$$c^*_{i,j} = \frac{\sum_{r=1}^{\delta(i)} \sum_{s=1}^{\delta(j)} [\text{Cov}(\beta_{i,r}, \beta_{j,s})/\sigma_r \sigma_s]}{\delta(i)\delta(j)}, \tag{4.71}$$

where, as in (4.66), $\text{Cov}(\beta_{i,r}, \beta_{j,s})$ is the (r, s) element of $\text{Cov}(\underline{\beta}_i, \underline{\beta}_j)$, and σ_r and σ_s are the square roots of the rth and sth main diagonal elements of $\Sigma_{i,i}$ and $\Sigma_{j,j}$, respectively.

Hence, from (4.65), (4.66) and (4.71),

$$c^*_{i,j} = c_{i,j} \frac{\sum_{r=1}^{\delta(i)} \sum_{s=1}^{\delta(j)} [\sigma_{r,s}/\sigma_r \sigma_s]}{\delta(i)\delta(j)}, \tag{4.72}$$

where $\sigma_{r,s}$ is the (r, s) element of $\Sigma^{\frac{1}{2}}_{i,i}(1)_{\delta(i)\delta(j)} \Sigma^{\frac{1}{2}}_{j,j}$, i.e.

$$c_{i,j} = c^*_{i,j} \frac{\delta(i)\delta(j)}{\sum_{r=1}^{\delta(i)} \sum_{s=1}^{\delta(j)} [\sigma_{r,s}/\sigma_r \sigma_s]}. \tag{4.73}$$

So, in view of (4.69) and (4.70), we have

$$w_{i,j} = w^*_{i,j} \frac{\delta(i)\delta(j)}{\sum_{r=1}^{\delta(i)} \sum_{s=1}^{\delta(j)} [\sigma_{r,s}/\sigma_r \sigma_s]}. \tag{4.74}$$

It remains now to investigate the allowable range for the positive scaling constant c, so that C in (4.68), and consequently V in (4.67), are positive-definite.

First, from (4.69), we write C in (4.68) as

$$C = I + cW, \tag{4.75}$$

where I is the identity matrix of order $\sum_{j=1}^{k} \delta(j)$, W is a conformally partitioned matrix with main diagonal zero block matrices, and all the elements of each (i, j) off-diagonal block are equal to $w_{i,j}$.

Let $\lambda_{W,i}$, $i = 1, 2, \cdots, \sum_{j=1}^{k} \delta(j)$, be the eigenvalues of W. We have that

$$\min_i(\lambda_{W,i}) < 0,$$

since if not, W with zero main diagonal elements will be a nonnegative-definite matrix, in which case

$$w_{i,j}^2 \leq w_{i,i}w_{j,j} = 0, \quad \forall i \neq j,$$

which is true if and only if W is a zero matrix.

Since I and W are symmetric, C in (4.75) is positive-definite if and only if all its eigenvalues, say $\lambda_{C,i}$, $i = 1, 2, \cdots, \sum_{j=1}^{k} \delta(j)$, are strictly greater than zero.

But

$$\lambda_{C,i} = 1 + c\lambda_{W,i}, \qquad i = 1, 2, \cdots, \sum_{j=1}^{k} \delta(j). \tag{4.76}$$

Consequently, C is positive-definite if and only if

$$\min_i(\lambda_{C,i}) > 0,$$

i.e. if and only if

$$c < \frac{-1}{\min_i(\lambda_{W,i})}. \tag{4.77}$$

The condition in (4.77) guarantees that C and V are positive-definite, and also that $c_{i,j} = cw_{i,j}$, $i \neq j$, are coherent correlation values, since, from the positive-definiteness of C,

$$c_{i,j}^2 < c_{i,i}c_{j,j} = 1.$$

The software obtains the value of $\min_i(\lambda_{W,i})$ using the eigenvalue decomposition of the matrix W. Then the boundary of c in (4.77) is computed.

With the software, different options are available to the expert for assessing a value of c that fulfils condition (4.77). The default option is to use a slider. The expert chooses the value of c that represents her opinion on the basis of interactive graphs. Specifically, the software displays a panel with k graphs, as illustrated in Figure 4.5.

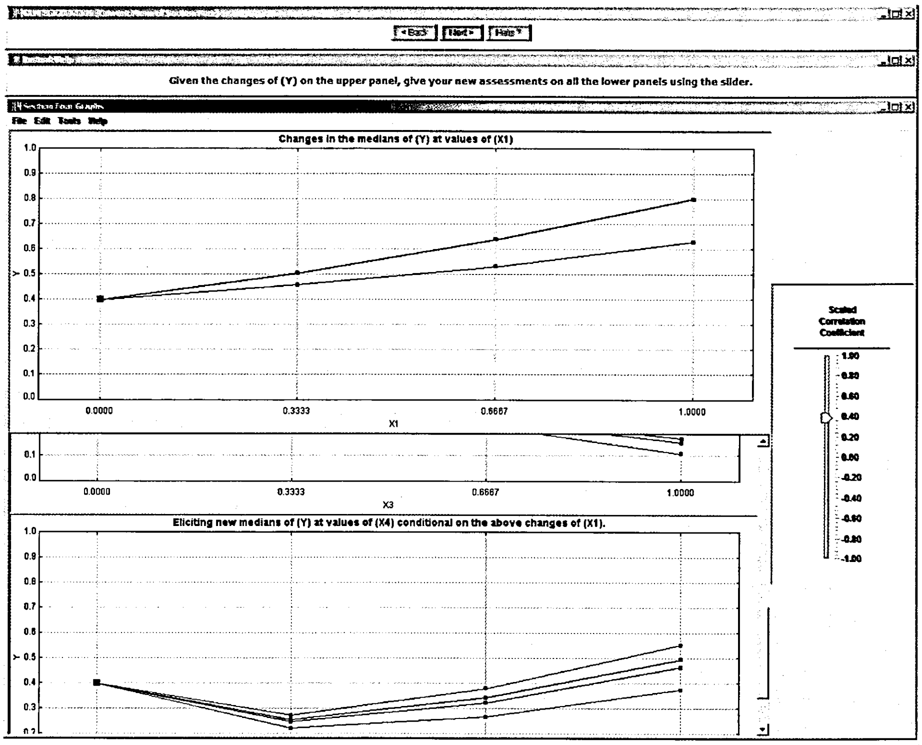


Figure 4.5: Assessments needed for various correlated variables

The upper graph shows the slopes for one continuous covariate after each of its slopes has been changed by a fixed amount, η. This covariate is one of the mostly highly correlated pair of vectors. In the same manner as in Section 4.3.3, the expert is asked to assess the new medians of all other $k-1$ covariates (factors) given the change in the above graph. Apart from the condition in (4.77), other equations needed for drawing the graphs are exactly as in Section 4.3.3.

Instead of using the slider and all graphs in Figure 4.5, another two options are also available to the expert after assessing the relative weights $w^*_{i,j}$. As the first option, the expert can choose to use the method proposed in Section 4.2, to elicit different covariances for the elements of the highest correlated pair, say $\underline{\beta}_s$ and $\underline{\beta}_t$. An averaging argument as in (4.71) is then used to get $c^*_{s,t}$. As the second option, the expert might use the method proposed in Section 4.3.1 to obtain $c^*_{s,t}$. In both cases, the value of c may be taken, for a small $\epsilon > 0$, as

$$c = \min \left\{ \frac{-1}{\min_i(\lambda_{W,i})} - \epsilon, \frac{c^*_{s,t}}{w^*_{s,t}} \right\}. \tag{4.78}$$

The expert may choose the option that suits her most. For example, the option that combines the current method with the one in Section 4.2.1 is flexible although it requires more conditional assessments. However, we favour the default option as it gives the expert a good chance to see how all the other covariates are affected by her choice of c.

The expert can, of course, go back in the software to change her assessed values of $w^*_{i,j}$, if she finds that the allowable range of c is not a reasonable representation of her opinion.

4.5 Concluding comments

Three different methods for eliciting expert opinion about the variance-covariance matrix of correlated coefficients in GLMs have been proposed.

The first method is the most flexible for modelling correlations between pairs of vectors - it is a good method if correlations are only substantial between a few pairs of variables, while the other correlations are near zero. However, it needs lots of assessments if there are lots of variables that are inter-related, and the number may become uncomfortably large. The positive-definiteness of the resulting matrix has only been investigated in the case of two vectors of correlated coefficients. No clear conditions have been investigated for the positive-definiteness of the whole matrix if many vectors of coefficients are thought to be correlated.

The second proposed method requires fewer assessments and has been shown to be a valid method for any number of vectors of correlated coefficients. Also, the required conditions for positive-definiteness of the covariance matrix in this method have been investigated. These were translated into boundaries for conditional assessments on the interactive graphs, which helps the expert fulfill the conditions. The disadvantage of the method is that it makes strong assumptions about the correlation structure between two vectors of coefficients, and sometimes the assumptions will be inappropriate.

The third proposed method requires a smaller number of assessments. For $n > 2$ correlated vectors of coefficients, the expert is required to make only $n(n-1)/2$ assessments of relative magnitudes of correlations between pairs of vectors. This leads to coherent estimates of correlations and a scaled variance-covariance matrix that is guaranteed to be positive-definite. The needed conditional medians can be easily assessed from the expert by the movement of one slider using the available user-friendly software. The method has been shown to give flexible options to the expert as an extension of the first or the second proposed methods. This third method is very promising. It also avoids incremented conditioning and treats all covariates symmetrically.

Chapter 5

Eliciting prior distributions for extra parameters of some GLMs

5.1 Introduction

So far, we have completed the process of eliciting the multivariate prior distribution for the vector of regression coefficients of any GLM. However, in some common GLMs, such as the normal and gamma regression models, the regression parameters are not the only parameters in the sampling model. The other parameters in these GLMs must be either assumed known or expert opinion about them must be quantified in a suitable way.

In normal GLMs, prior opinion about regression coefficients can be quantified using the methods discussed in the previous two chapters. However, prior opinion about the error variance in normal GLMs must also be quantified to complete the prior distribution of all the model parameters.

A limited number of elicitation methods for error variance in normal linear models has been proposed in the literature. See, for example, Kadane *et al.* (1980), Garthwaite and Dickey (1988) and Ibrahim and Laud (1994). However, these available methods have been criticized for using assessment tasks that the expert may not be very good at performing (Garthwaite *et al.*, 2005).

The method of Garthwaite and Dickey (1988) elicits a conjugate inverted chi-squared prior distribution for the error variance through conditional assessments that depend only upon the experimental error. The expert is required to assess her median of the absolute difference between two observed values of the response variable at the same design point. Then conditional medians of the same difference is assessed given a set of hypothetical data. These two assessments are sufficient to elicit the two hyperparameters of the inverted chi-squared prior of the normal error variance. However, it is better to specify several data sets and get a conditional median for each data set, then different assessments can be reconciled to elicit the two hyperparameters. In this chapter, we propose an elicitation method based on more than one data set of hypothetical future samples.

The second task addressed in this chapter is to assess prior distributions for the shape parameter of a gamma distribution and the scale parameter of gamma GLMs. Prior distributions for these parameters have been proposed in the literature [see for example Miller (1980), West (1985) or Chen and Ibrahim (2003)], but no prior elicitation method for these parameters has been suggested. To fill this gap, we propose a new method for eliciting lognormal prior distributions for such parameters. The proposed method is based on conditional quartile assessments given that the mean of the gamma distribution is known or has already been elicited.

In Section 5.2 of this chapter, we extend the method of Garthwaite and Dickey (1988) for eliciting the variance of random errors in normal GLMs. A novel method for eliciting a lognormal prior distribution for the scale parameter in gamma GLMs is proposed in Section 5.3. The two methods have been implemented as extra options in our elicitation software PEGS-GLM (Correlated Coefficients) that is freely available at http://statistics.open.ac.uk/elicitation.

5.2 Eliciting a prior distribution for the error variance in normal GLMs

The method of Garthwaite and Dickey (1988) is based on conditional assessments that depend only on the random error to elicit a conjugate inverted chi-squared prior distribution for the normal error variance. In their method, the expert is asked to assume that two observations are taken at the same design point. Then she assesses her median of their absolute difference - the two observations differ only because of random variation.

The method has been also used to quantify experts' opinion about multivariate normal distributions [Al-Awadhi and Garthwaite (1998, 2001), Garthwaite and Al-Awadhi (2001)]. However, it has been criticized for eliciting only the minimum number of assessments that are required to determine the hyperparameters. To overcome this, Garthwaite *et al.* (2005) suggested that it is a good idea to elicit more than one estimate of the hyperparameters and to then reconcile these estimates in some way.

The aim of this section is to extend the method of Garthwaite and Dickey (1988) by increasing the size and frequency of the hypothetical (virtual) sample data that are used as the conditioning set on which the expert is modifying her opinion. Our extended method is designed to elicit a conjugate prior for the error variance in normal GLMs. This will complete the prior distribution structure of these models when the prior distribution of their regression coefficients is elicited using the piecewise-linear model discussed in the previous chapters. However, the method developed here can be used to elicit the prior distribution of error variance in any normal model where the prior distribution of its regression coefficients is totally known or has been elicited using any other elicitation method.

The theoretical derivation of the proposed extension is detailed in Section 5.2.1. The implementation of the method has been programmed as a new option in the PEGS-GLM (Correlated Coefficients) software. The assessment tasks and the description of the procedure that implements our proposed method are discussed in Section 5.2.2.

5.2.1 The mathematical framework and notations

The normal GLM assumes that the link function $g(.)$ in (3.3) is the identity link function, which means, in view of (3.2), that

$$\zeta = \alpha + \beta_1' X_1 + \beta_2' X_2 + \cdots + \beta_{m+n}' X_{m+n} + \varepsilon, \tag{5.1}$$

where ε is assumed to be a normal random error with zero mean and an unknown variance σ_ε^2, i.e.

$$\varepsilon \sim \mathrm{N}(0, \sigma_\varepsilon^2). \tag{5.2}$$

A conjugate prior for σ_ε^2 is the inverted chi-squared distribution [see, for example, Pratt *et al.* (1995), Kadane *et al.* (1980) or Garthwaite and Dickey (1988)]. Equivalently, we assume that

$$\sigma_\varepsilon^2 \sim \text{Inverted Gamma}(\nu/2, \nu w/2), \tag{5.3}$$

with a pdf of the form

$$f(\sigma_\varepsilon^2; \nu, w) = \frac{1}{\Gamma(\nu/2)} \left(\frac{\nu w}{2}\right)^{\nu/2} \left(\frac{1}{\sigma_\varepsilon^2}\right)^{(\nu/2)+1} \exp\left(-\frac{\nu w}{2\sigma_\varepsilon^2}\right),$$

$$\sigma_\varepsilon^2 > 0,\ \nu > 0,\ w > 0. \qquad (5.4)$$

The aim now is to elicit the values of the hyperparameters ν and w of the pdf in (5.4). To attain this, the expert should preferably be asked to assess values that depend only on the random variation. For that, the method of Garthwaite and Dickey (1988) requires the expert to assess a median value, say q_0, of the absolute difference, $|\zeta_1 - \zeta_2|$, between two observed values of the response variable ζ at the same design point $(X_1, X_2, \cdots, X_{n+m})$.

The expert is then asked to assume that the true value of this absolute difference is a suggested value z. Given this piece of information, she gives her new median assessment , say q_1, of the absolute difference between two observations for any new hypothetical experiments at the same design point $(X_1, X_2, \cdots, X_{n+m})$. The difference between q_0 and the new median assessment, q_1, reflects the expert's confidence in her first median assessment q_0. Then both q_0 and q_1 were used in Garthwaite and Dickey (1988) to calculate the two hyperparameters ν and w.

To extend their method, instead of conditioning on only one hypothetical datum z, we repeat the assessment of the conditional median for a number of s steps. At each step, the condition is on a steadily increasing set of hypothetical data representing the response differences for pairs of experiments at the same design point.

At each step j, $j = 1, 2, \cdots, s$, the expert is asked to assume that a number $k(j) = 2^{j-1}$ of experiment pairs at the same design point has given a hypothetical data set of absolute differences, $z_1, z_2, \cdots, z_{k(j)}$. She is then asked to give her conditional median q_j of the absolute response difference of a new pair of experiments at the same design point. In what follows, we show how to use these assessments to estimate a number of elicited values that can be reconciled to give a better assessment of ν and w.

For $i = 1, \cdots, k$, where $k > 1$ is any integer number, let Z_i be the difference between the two observed values, $\zeta_{i,1}$ and $\zeta_{i,2}$, of the response variable ζ in any two experiments at the same design point $(\underline{X}_1, \cdots, \underline{X}_{m+n})$, i.e. $Z_i = \zeta_{i,1} - \zeta_{i,2}$.

Clearly, from (5.1) and (5.2), given σ_ε^2, the random variables $Z_1, \cdots, Z_k$ are independent and identically distributed normal variates, i.e. for $i = 1, 2, \cdots, k$,

$$Z_i|\sigma_\varepsilon^2 \sim \mathrm{N}(0, 2\sigma_\varepsilon^2), \tag{5.5}$$

with the joint distribution

$$f(z_1, \cdots, z_k|\sigma_\varepsilon^2) = \frac{1}{(4\sigma_\varepsilon^2\pi)^{k/2}} \exp\left(-\sum_{i=1}^{k} z_i^2/4\sigma_\varepsilon^2\right),$$

$$-\infty < z_i < \infty, \quad \sigma_\varepsilon^2 > 0. \tag{5.6}$$

From (5.4) and (5.6), the joint distribution of $Z_1, \cdots, Z_k$ and σ_ε^2 is given by

$$f(z_1, \cdots, z_k, \sigma_\varepsilon^2; \nu, w) = \frac{(\nu w/2)^{\nu/2}}{\Gamma(\nu/2)(4\pi)^{k/2}} \left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{\nu+k}{2}+1} \exp\left\{-\frac{1}{4\sigma_\varepsilon^2}\left(\sum_{i=1}^{k} z_i^2 + 2\nu w\right)\right\},$$

$$-\infty < z_i < \infty, \quad \sigma_\varepsilon^2, \nu, w > 0. \tag{5.7}$$

Integrating σ_ε^2 out from the RHS of (5.7), we get

$$f(z_1, \cdots, z_k; \nu, w) = \frac{\Gamma((\nu+k)/2)}{\Gamma(\nu/2)[\nu\pi(2w)]^{k/2}} \left[1 + \frac{\sum_{i=1}^{k} z_i^2}{\nu(2w)}\right]^{-\frac{\nu+k}{2}},$$

$$-\infty < z_i < \infty, \quad \nu, w > 0, \tag{5.8}$$

which is the k-variate version of the general three-parameter Student-t distribution with ν degrees of freedom, zero mean vector and a diagonal scale matrix $2wI_k$, where I_k is the identity matrix of order k, i.e.

$$Z_1, \cdots, Z_k \sim \text{MV-}t_\nu(\underline{0}_k, 2wI_k). \tag{5.9}$$

Now, the conditional distribution of σ_ε^2 given $Z_1 = z_1, \cdots, Z_k = z_k$, can be obtained by dividing the RHS of (5.7) by that of (5.8) to get

$$f(\sigma_\varepsilon^2|Z_1 = z_1, \cdots, Z_k = z_k; \nu, w) = \frac{1}{\Gamma((\nu+k)/2)} \left[\frac{1}{4}\left(2\nu w + \sum_{i=1}^{k} z_i^2\right)\right]^{(\nu+k)/2} \times$$

$$\left(\frac{1}{\sigma_\varepsilon^2}\right)^{\frac{\nu+k}{2}+1} \exp\left\{-\frac{1}{4\sigma_\varepsilon^2}\left[2\nu w + \sum_{i=1}^{k} z_i^2\right]\right\}, \qquad \sigma_\varepsilon^2, \nu, w > 0. \tag{5.10}$$

Since the inverted gamma distribution is a conjugate prior for σ_ε^2, comparing (5.10) with (5.4), we can write

$$(\sigma_\varepsilon^2|Z_1 = z_1, \cdots, Z_k = z_k) \sim \text{Inverted Gamma}\left(\frac{\nu+k}{2}, \frac{(\nu+k)w_k}{2}\right), \tag{5.11}$$

where

$$w_k = \frac{1}{\nu+k}\left[\nu w + \frac{\sum_{i=1}^k z_i^2}{2}\right]. \tag{5.12}$$

For $j = 0, 1, \cdots, s$, define a new set, $Z_{(j)} = \zeta_{(j),1} - \zeta_{(j),2}$, of the response variable differences for two further experiments at the same design point $(\underline{X}_1, \cdots, \underline{X}_{m+n})$. The variates in this new set are iid with the same normal distribution as in (5.5).

The conditional distribution of $(Z_{(j)}|Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)})$, with $k(j) = 2^{j-1}$, for $j = 1, \cdots, s$, is given by

$$f(Z_{(j)}|Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)}) =$$
$$\int_{\sigma_\varepsilon^2=0}^{\infty} f(Z_{(j)}|\sigma_\varepsilon^2) \times f(\sigma_\varepsilon^2|Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)})d\sigma_\varepsilon^2. \tag{5.13}$$

Using the normal distribution in (5.5), and putting $k = k(j)$ in (5.11), the integrand in (5.13) is similar to the RHS of (5.7) with k set equal to 1, ν replaced by $\nu + k(j)$ and w replaced by $w_{k(j)}$ with k set equal to $k(j)$ in (5.12).

As in (5.8) and (5.9), integrating σ_ε^2 out from (5.13) gives

$$(Z_{(j)}|Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)}) \sim t_{\nu+k(j)}(0, 2w_{k(j)}), \tag{5.14}$$

for $j = 1, \cdots, s$.

Similarly, for $j = 0$, the marginal unconditional distribution of $Z_{(0)}$ is obtained, from (5.4) and (5.5), as

$$Z_{(0)} \sim t_\nu(0, 2w). \tag{5.15}$$

As will be discussed in the next section, under reasonable choices of the conditioning values $z_1, \cdots, z_{k(j)}$, the expert assesses her median of the absolute value for each of the Student-t distributions in (5.14) and (5.15). These are exactly the upper quartiles of the t-variates, from symmetry about zero.

Let the assessed upper quartile of $Z_{(0)}$ and $(Z_{(j)}|Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)})$ be denoted by q_0 and q_j, for $j = 1, \cdots, s$, respectively. If we denote the upper quartile of a standard Student-t distribution with ν degrees of freedom by Q_ν, then we have

$$q_0 = (2w)^{1/2} Q_\nu, \tag{5.16}$$

and

$$q_j = (2w_{k(j)})^{1/2} Q_{\nu+k(j)}, \tag{5.17}$$

for $j = 1, \cdots, s$.

The aim now is to solve the above pairs of equations for ν and w. By division, for each pair, we get

$$\frac{q_0}{q_j} = \frac{Q_\nu}{Q_{\nu+k(j)}} \left[\frac{w}{w_{k(j)}}\right]^{1/2}. \tag{5.18}$$

Using (5.12), (5.16), we can eliminate w from (5.18), to get

$$\frac{q_0}{q_j} = \frac{Q_\nu}{Q_{\nu+k(j)}} \left[\frac{\nu + k(j)}{\nu + Q_\nu^2 \sum_{i=1}^{k(j)} (z_i/q_0)^2}\right]^{\frac{1}{2}}, \tag{5.19}$$

for $j = 1, \cdots, s$.

For each value of j, the assessed ratio of q_0/q_j is used by the software to search for the value of the degrees of freedom ν, say ν_j, that solves equation (5.19).

To guarantee the existence of a unique solution for ν using this approach, two conditions must be imposed on the function in (5.19). It must be strictly monotonic in ν on the interval of concern. For statistical coherence, the assessed quartile, q_j, must also be above a lower limit, say a_j, for $j = 1, 2, \cdots, s$.

To satisfy the latter condition, we assume that there is a reasonable minimum value of the elicited degrees of freedom, say $\min(\nu)$. Since q_0 has already been assessed, using the extreme value $\min(\nu)$ in the RHS of (5.19) gives the lower limit of q_j, as follows:

$$a_j = q_0 \frac{Q_{\min(\nu)+k(j)}}{Q_{\min(\nu)}} \left[\frac{\min(\nu) + Q_{\min(\nu)}^2 \sum_{i=1}^{k(j)} (z_i/q_0)^2}{\min(\nu) + k(j)}\right]^{\frac{1}{2}}, \tag{5.20}$$

for $j = 1, 2, \cdots, s$.

Setting this limit, we can now investigate the monotonicity condition. In fact, the monotonicity of (5.19) as a function of ν is required to ensure that there exists a unique value $\nu_j \geq \min(\nu)$ that satisfies (5.19) for $q_j \geq a_j$, $j = 1, 2, \cdots, s$.

In (5.19), if we put

$$C_j = \sum_{i=1}^{k(j)} (z_i/q_0)^2, \tag{5.21}$$

then the first derivative of q_0/q_j with respect to ν will take the form

$$\begin{aligned}\frac{\partial(q_0/q_j)}{\partial\nu} =& \frac{1/2(q_j/q_0)Q_\nu Q_{\nu+k(j)}}{Q^4_{\nu+k(j)}(C_jQ_\nu^2+\nu)^2}\{C_jQ_\nu^3[Q_{\nu+k(j)} - 2(\nu+k(j))Q'_{\nu+k(j)}]\\ &+ 2\nu(\nu+k(j))(Q'_\nu Q_{\nu+k(j)} - Q_\nu Q'_{\nu+k(j)}) - k(j)Q_\nu Q_{\nu+k(j)}\}.\end{aligned}$$

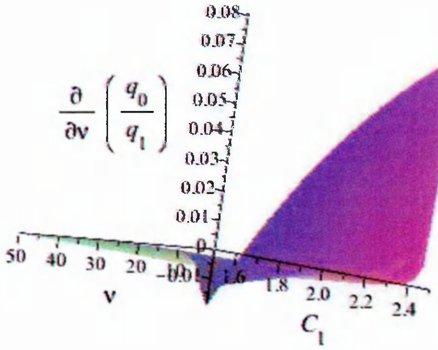
So, for all $\nu \geq \min(\nu)$,

$$\frac{\partial(q_0/q_j)}{\partial\nu} < 0$$

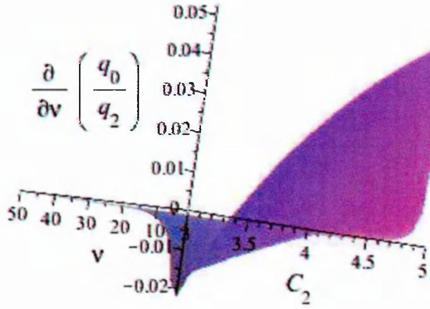
if and only if

$$C_j < \min_{\nu\geq\min(\nu)}\left\{\frac{k(j)Q_\nu Q_{\nu+k(j)} - 2\nu(\nu+k(j))(Q'_\nu Q_{\nu+k(j)} - Q_\nu Q'_{\nu+k(j)})}{Q_\nu^3[Q_{\nu+k(j)} - 2(\nu+k(j))Q'_{\nu+k(j)}]}\right\} = C_{j,0}. \tag{5.22}$$

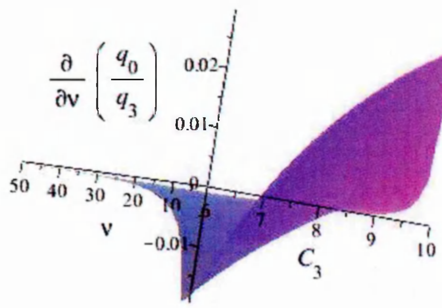
Since there does not exist a closed form for the derivative of a Student-t quantile with respect to its degrees of freedom, the values of $C_{j,0}$ cannot be found analytically. Instead, these values have been computed numerically using Maple 14 Software, for $s = 5$, $\nu \in [1, 50]$. Figure 5.1 lists these values of $C_{j,0}$, where the derivative $\partial(q_0/q_j)/\partial\nu$ is plotted against ν and C_j, for $j = 1, 2, \cdots, 5$.



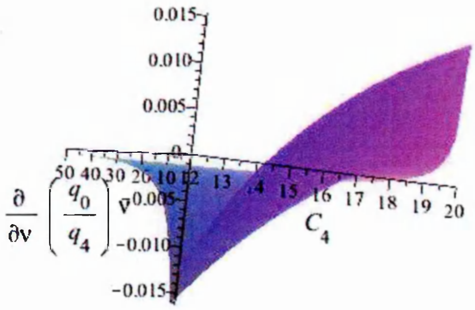
For $k(1) = 1$, $C_{1,0} = 1.626$.



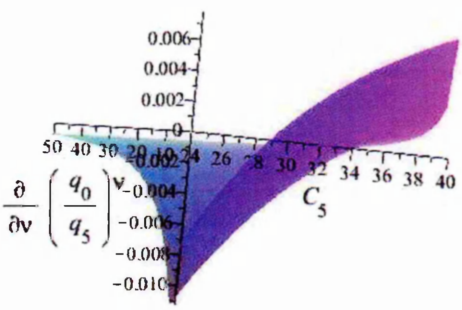
For $k(2) = 2$, $C_{2,0} = 3.367$.



For $k(3) = 4$, $C_{3,0} = 6.950$.



For $k(4) = 8$, $C_{4,0} = 14.222$.



For $k(5) = 16$, $C_{5,0} = 28.846$.

For $j = 1, 2, \cdots, 5$, $C_{j,0}$ is such that:

$$\frac{\partial}{\partial \nu}\left(\frac{q_0}{q_j}\right) < 0, \text{ for all } 1 \leq \nu \leq 50,$$

if and only if $C_j < C_{j,0}$.

Figure 5.1: Three dimension plots of $\partial(q_0/q_j)/\partial\nu$ against ν and C_j for various sample sizes $k(j)$.

It can be seen from Figure 5.1 that

$$C_{1,0} < \frac{C_{j,0}}{k(j)}, \qquad \text{for } j = 2, 3, 4, 5. \tag{5.23}$$

Now, from (5.21), (5.22), (5.23) and Figure 5.1, we can state that the function in (5.19) is strictly monotonic decreasing in ν, for all $1 \le \nu \le 50$ and $j = 1, 2, \cdots, 5$, if and only if

$$\frac{\sum_{i=1}^{k(j)} (z_i/q_0)^2}{k(j)} < 1.626. \tag{5.24}$$

Although we have not examined the case where $\nu > 50$, Figure 5.1 suggests that (5.24) holds for $\nu \ge 1$.

In the implementation of the method, the software generates the values of z_i that satisfy (5.24). Hence, for $j = 1, 2, \cdots, 5$, a unique solution ν_j can be obtained from (5.19), then the corresponding w_j can be obtained by substituting ν_j for ν in (5.16). We then reconcile the five different values of the degrees of freedom parameter ν by taking their geometric mean.

When averaging different assessments of a degrees of freedom parameter, taking their geometric mean is favored, by empirical evidence, rather than their arithmetic mean. See for example, Al-Awadhi (1997), Al-Awadhi and Garthwaite (1998) or Garthwaite *et al.* (2005). The elicited value of w can then be obtained from (5.16) by substituting for ν with the geometric mean of $\nu_1, \cdots, \nu_5$.

Finally, we assume that the regression coefficients vector of parameters $\beta = (\alpha, \beta_1, \cdots, \beta_{m+n})$ is independent from σ_ε^2 *a priori*, and give the full prior structure of the normal GLM as

$$f(\underline{\beta}, \sigma_\varepsilon^2) = f_1(\underline{\beta}) f_2(\sigma_\varepsilon^2), \tag{5.25}$$

where $f_1(\underline{\beta})$ can be taken as the multivariate normal prior distribution elicited in the previous chapters, and $f_2(\sigma_\varepsilon^2) = f(\sigma_\varepsilon^2; \nu, w)$ as given in (5.4) with the elicited hyperparameters ν and w.

5.2.2 Implementation and assessment tasks

The elicitation method proposed in the previous section has been programmed into the PEGS-GLM (Correlated Coefficients) software by the author of this thesis. The option of eliciting the prior distribution of the random error variance is given to the expert once she selects her model as an "ordinary linear regression" model. The same procedure has also been programmed in a separate piece of software that can be used as an add-on to any other elicitation software for normal models. This developed software is freely available as PEGS-Normal at http://statistics.open.ac.uk/elicitation.

In a dialogue box, the expert is asked to assume that two independent experiments have been conducted at the same design point, i.e. at the same values of the explanatory variables. She then assesses her median value, q_0, of the absolute difference, $|Z_{(0)}|$, between the observed values of the response variable after these two virtual experiments.

Since the distribution of $Z_{(0)}$ is symmetric about zero, see (5.15), the assessed median q_0 of $|Z_{(0)}|$ is exactly the upper quartile of $Z_{(0)}$. In fact, $\Pr\{|Z_{(0)}| < q_0\} = 0.5$ implies that $\Pr\{-q_0 < Z_{(0)} < q_0\} = 0.5$, which implies from symmetry that $\Pr\{Z_{(0)} < q_0\} = 0.75$. Similarly, from (5.14), each upper quartile q_j, for $j = 1, \cdots, s$ will be assessed as the median of the absolute difference $|Z_{(j)}|$ given that $Z_1 = z_1, \cdots, Z_{k(j)} = z_{k(j)}$.

In assessing the remaining conditional medians q_j, the choice of the conditioning values $z_1, z_2, \cdots, z_{k(j)}$, for $j = 1, 2, \cdots, 5$, is an important issue. As mentioned before, the method of Garthwaite and Dickey (1988) uses only one hypothetical data point z_1, for which they suggested a value of $z_1 = q_0/2$. They argued that, this choice will give a conditioning value that is not too close to q_0, so as to prompt a significant change in the expert's opinion in assessing q_1. This value of z_1 is, at the same time, not too far from q_0, so as to represent an acceptable value for the expert to condition on.

In our implementation of the extended method, the above two criteria will be considered in choosing values for z_i, $i > 1$. This means that the values should result in a considerable change in the expert's opinion, while the expert still find them plausible values. To attain this, we take $z_1 = q_0/2$, following Garthwaite and Dickey (1988). Then we generate four extra sets of hypothetical data, for $j = 2, \cdots, 5$, the jth set consists of $k(j) = 2^{j-1}$ data points. The first 2^{j-2} data points of each set, namely $z_1, \cdots, , z_{2^{j-2}}$, are taken as the same elements of the previous data set, while the new extra elements $z_{2^{j-2}+1}, \cdots, , z_{2^{j-1}}$, are generated as follows.

For $i = 2^{j-2} + 1, \cdots, 2^{j-1}$, we generate z_i as random variates from a population with a median of $q_0/2$. Hence, we choose each z_i as the absolute value of a normal variate with zero mean and a variance of $(q_0/1.349)^2$. Thus, the interquartile range of this normal distribution is q_0, and the upper quartile of the signed variates, which is also the median of the unsigned ones, is exactly $q_0/2$.

For any data set j, $j = 2, \cdots, 5$, if the generated values fail to satisfy the following condition

$$\frac{\sum_{i=1}^{k(j)} z_i^2}{k(j)} \leq \left(\frac{3}{4} q_0\right)^2, \tag{5.26}$$

we resample the new elements $z_{2^{j-2}+1}, \cdots, , z_{2^{j-1}}$, from the same normal distribution, until (5.26) is satisfied. This guarantees that the generated data should prompt the expert to revise her opinion by a substantial amount.

To implement the proposed procedure, The expert is asked to perform an assessment task that consists of $s = 5$ steps. In each step j, for $j = 1, \cdots, 5$, the software presents an interactive graph to the expert. The graph in Figure 5.2 is an example of the graph presented to the expert by the software at step $j = 3$.

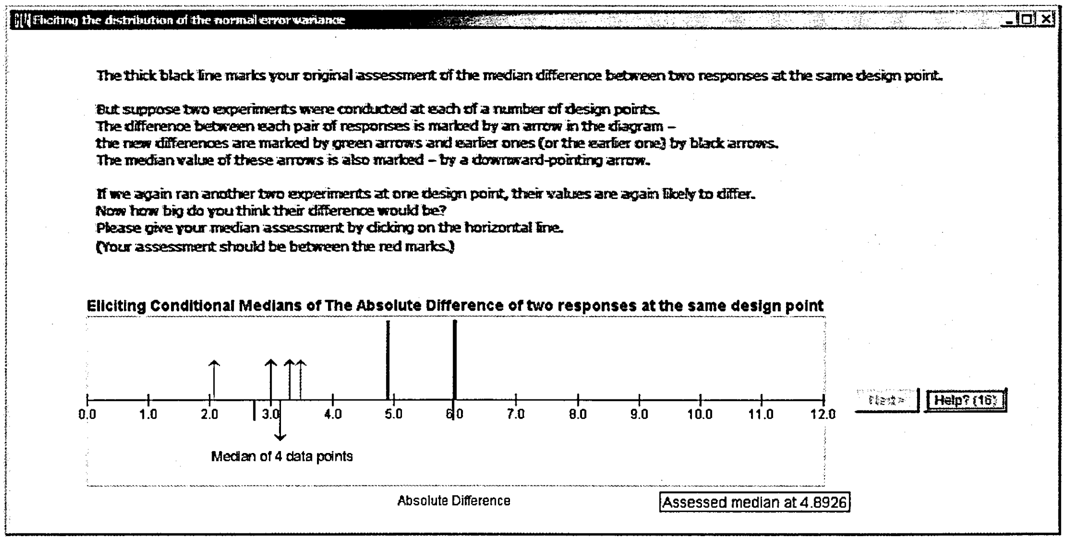


Figure 5.2: Assessing a median value conditioning on a set of data

This graph shows the expert's first unconditional median q_0 drawn as the thick black long line and the more recent assessed median in the second step $(j-1=2)$ as the other thick green long line. The graph also shows a number of $k(3) = 2^{3-1} = 4$ generated data points $z_1, \cdots, z_4$, represented by upward arrows, together with a downward arrow that shows the sample median of this virtual data set.

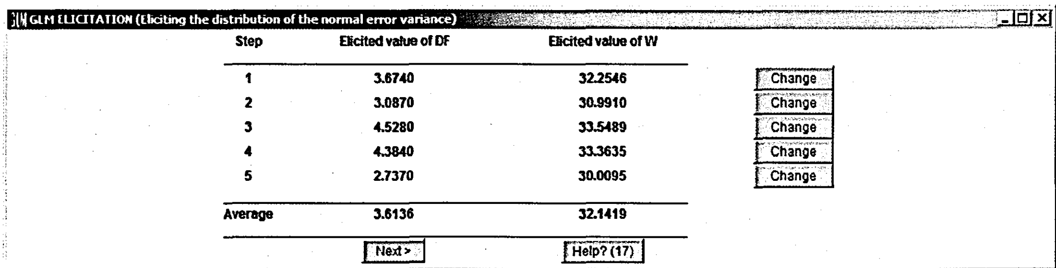
The upward arrows of the data points from the previous set of hypothetical data, z_1 and z_2, are shown in the green color, while the upward arrows of the new generated data points, z_3 and z_4, are shown in the black color.

Given the virtual data set (displayed as arrows), the expert is asked to assess her current median value q_3 by clicking on the horizontal line between the two short red lines. These are the lower limit a_3 computed as in (5.20) with $\min(\nu) = 1$ and the initial assessment q_0. The expert's median must lie between the red boundaries, otherwise she will get a warning message asking her to re-assess her median and satisfy this condition.

To assess q_3, the expert has two obvious strategies. The first strategy (the black one) is to look at the black line that shows her initial assessment q_0, and decide where to revise this value in the light of the new information given by the black downward arrow that shows the median of the whole hypothetical data set $z_1, \cdots, z_4$. The other strategy (the green one) is for the expert to look at the green line that shows her most recent median assessment which has been based on the hypothetical data set in green arrows z_1 and z_2. She then decides where to revise this median assessment in the light of the new generated points z_3 and z_4 shown as the black arrows.

With both of these strategies, if the expert is confident about her previous assessment, then her new median assessment should be near to this value rather than near to the new hypothetical data. When the expert gives her new median assessment q_3, its value is first used by the software to compute ν_3 from (5.19), and then to compute w_3 from (5.16) using ν_3.

The final output of the procedure, as illustrated in Figure 5.3, gives the five different elicited pairs of ν and w, together with the geometric mean of ν and its corresponding value of w. The expert is asked to check whether the different elicited values are close to each other and represents her opinion well. If not, she has the option to change any of them by going back to reassess a specific q_j through pressing the corresponding 'Change' button for this step, see Figure 5.3.


GLM ELICITATION (Eliciting the distribution of the normal error variance)

Step	Elicited value of DF	Elicited value of W	
1	3.6740	32.2546	Change
2	3.0870	30.9910	Change
3	4.5280	33.5489	Change
4	4.3840	33.3635	Change
5	2.7370	30.0095	Change
Average	3.6136	32.1419	
	Next >	Help? (17)	

Figure 5.3: The output table showing the elicited hyperparameters

After the expert has finished making any revision, the hyperparameters ν and w are set equal to the two values in the last row of the table illustrated in Figure 5.3.

5.3 Eliciting a prior distribution for the scale parameter in gamma GLMs

In this section, we propose a novel method for eliciting a lognormal prior distribution for the scale parameter of a gamma GLM. It is well-known that the scale parameter of a gamma GLM, which is the reciprocal of the dispersion parameter, is in fact the shape parameter of the gamma distribution. Our new method is a valid means of eliciting the shape parameter of any gamma distribution once the distribution's mean has been elicited (or the mean is assumed to be known).

Bayesian methods have been developed for analyzing data to estimate the shape parameters of a gamma distribution, or the scale parameters of a gamma GLM. Miller (1980) proposed a general conjugate class of priors for the two parameters of the gamma distribution, but he gave no method of eliciting its hyperparameters. Sweeting (1981) introduced some suggestions for the Bayesian estimation of the scale parameters in exponential families. The problem of unknown scale parameters in GLMs was examined by West (1985). In his work, he discussed general ideas concerning scale parameters and variance functions in non-normal models including gamma GLMs, (see also West *et al.* (1985)). However, there does not seem to be a good method of eliciting a prior distribution for such parameters. Ibrahim and Laud (1991) suggested a Jeffreys's prior for the regression coefficients and an independent marginal informative prior on the scale parameter of gamma GLM, but they did not suggest any family of distributions for this informative prior. The method of Bedrick *et al.* (1996), which is considered as the first elicitation method of informative prior distributions for GLMs, assumed the scale parameter to be known and elicited priors only for the regression coefficients. Chen and Ibrahim (2003) proposed a novel class of conjugate priors for GLMs. They also discussed elicitation issues and strategies of these conjugate priors. Their proposed prior structure involves the dispersion parameter as well. However, no explicit elicitation method was introduced for the dispersion parameter.

5.3.1 GLMs with a gamma distributed response variable

For a continuous, positive, skewed distributed response variable ζ in a GLM of the form,

$$Y = g(\mu) = g(E(\zeta|\underline{X})) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m, \tag{5.27}$$

the observations are often assumed to follow a gamma distribution, say

$$\zeta \sim \text{Gamma}(\lambda, \theta),$$

where λ and θ depend on $\underline{X}$. Its pdf is

$$f(\zeta|\lambda,\theta) = \frac{1}{\Gamma(\lambda)}\theta^{\lambda}\zeta^{\lambda-1}e^{-\theta\zeta}, \quad \zeta, \lambda, \theta > 0, \tag{5.28}$$

where λ is the shape parameter, θ is the rate parameter or the inverse of the scale parameter. It is well-known that

$$\mu = E(\zeta) = \lambda/\theta, \qquad \sigma_\zeta^2 = \text{Var}(\zeta) = \lambda/\theta^2. \tag{5.29}$$

For the gamma GLM in (5.27), with any monotone increasing link function $g(.)$, the methods discussed in Chapters 3 and 4 can be used to elicit the prior distribution of the regression coefficients

$$\underline{\beta} = (\alpha, \ \beta_1, \ \beta_2, \ \cdots, \ \beta_m), \tag{5.30}$$

which represents the prior distribution of μ, i.e. reflects the prior knowledge about the ratio λ/θ. We assume that the prior distribution of this ratio has already been elicited as

$$g(\lambda/\theta) \sim N(\underline{X}_0'\underline{b}, \underline{X}_0'\Sigma\underline{X}_0), \tag{5.31}$$

where $\underline{b} = E(\underline{\beta})$, $\Sigma = \text{Var}(\underline{\beta})$, have been assessed using methods given in the previous chapters, and the vector $\underline{X}_0$ denotes all explanatory variables to be at their reference points.

Having elicited this prior for the ratio λ/θ, the prior expert's opinion about one of the hyperparameters λ and θ must be quantified to complete the prior structure of the gamma GLM model. In what follows, expert opinion about the scale parameter λ is modelled by a lognormal prior distribution and we propose an assessment method for determining the hyperparameters of this distribution. As discussed before, the proposed method can be also used to elicit a shape parameter λ of any gamma distribution.

We base our method on a gamma distribution with λ as the only unknown parameter, assuming μ to be already assessed or completely known. For gamma GLMs, the elicited vector $\underline{b}$ can be used to obtain a single value of μ, say μ_0, from (5.31). As we assume that the link function $g(.)$ is monotonic increasing, the median value of λ/θ is then

$$\mu_0 = g^{-1}(\underline{X}_0'\underline{b}). \tag{5.32}$$

We take the gamma distributed random variable ζ defined in (5.28) and change parameters by putting $\theta = \lambda/\mu$ as in 5.29. This gives

$$f(\zeta|\lambda,\mu) = \frac{1}{\Gamma(\lambda)}\left(\frac{\lambda}{\mu}\right)^{\lambda}\zeta^{\lambda-1}e^{-\left(\frac{\lambda}{\mu}\right)\zeta}. \quad \zeta,\lambda,\mu > 0, \tag{5.33}$$

We let

$$W = \frac{\zeta}{\mu}, \tag{5.34}$$

and then the pdf of W will depend only on λ, i.e. $W \sim \text{Gamma}(\lambda,\lambda)$. This has the form

$$f(w|\lambda) = \frac{1}{\Gamma(\lambda)}\lambda^{\lambda}w^{\lambda-1}e^{-\lambda w}, \quad w,\lambda > 0. \tag{5.35}$$

Our aim now is to find some meaningful strictly monotonic function in λ, such that the expert can quantify her opinion about this function effectively. The expert cannot answer questions about λ directly, as a gamma distribution parameter has little meaning to an expert because it is not an observable quantity. Instead, the expert should be asked about an observable quantity that directly relates to the observable gamma variate, and which can be monotonically transformed to λ. The expert can thus be asked about any quantile of the gamma distribution as an observable quantity, provided that it is a strictly monotonic function in λ. In what follows, we show that quantifying the expert opinion about the lower quartile of the gamma distribution in (5.35) will lead to a full prior distribution for λ, and that this quartile is a strictly monotonic function in λ.

To check the monotonicity of different quantiles in λ, let $F(w, \lambda, \lambda)$ be the cdf of W, then it can be written in the form of a regularized gamma function as follows

$$F(w, \lambda, \lambda) = \frac{\gamma(\lambda, \lambda, w)}{\Gamma(\lambda)}, \tag{5.36}$$

where $\gamma(\lambda, \lambda, w)$ is a form of the lower incomplete gamma function,

$$\gamma(\lambda, \lambda, w) = \int_{t=0}^{w} \lambda^{\lambda} t^{\lambda-1} e^{-\lambda t} dt. \tag{5.37}$$

Note that it differs from the usual lower incomplete gamma function $\gamma(\lambda, w)$ in that the latter does not contain λ^{λ} in the integrand.

It is clear that the function $F(w, \lambda, \lambda)$, as a cdf of W, is strictly monotonic increasing in w. But, as a function in λ, the usual cdf

$$F(w, \lambda) = \frac{\gamma(\lambda, w)}{\Gamma(\lambda)}, \tag{5.38}$$

as a regularized gamma function is strictly monotonic decreasing in λ. The proof of this fact is given in Tricomi (1952), see also Gautschi (1998).

We next show that the same type of monotonicity is true for the function $F(w, \lambda, \lambda)$ in (5.36). This helps in finding a range of quantiles that are monotonic functions in λ.

In fact, following the note of Koornwinder (2008) for $F(w, \lambda)$, we can write

$$F(w, \lambda, \lambda) = \frac{\gamma(\lambda, \lambda, w)}{\Gamma(\lambda)} = \frac{\gamma(\lambda, \lambda, w)}{\gamma(\lambda, \lambda, w) + \Gamma(\lambda, \lambda, w)}, \tag{5.39}$$

where $\gamma(\lambda, \lambda, w)$ takes the form in (5.37), and $\Gamma(\lambda, \lambda, w)$ is a form of the upper incomplete gamma function, i.e.

$$\Gamma(\lambda, \lambda, w) = \int_{t=w}^{\infty} \lambda^{\lambda} t^{\lambda-1} e^{-\lambda t} dt. \tag{5.40}$$

Differentiating (5.39) with respect to λ, we have

$$\frac{\partial F(w, \lambda, \lambda)}{\partial \lambda} = \frac{-1}{\Gamma^2(\lambda)} \left\{ \gamma(\lambda, \lambda, w) \frac{\partial \Gamma(\lambda, \lambda, w)}{\partial \lambda} - \Gamma(\lambda, \lambda, w) \frac{\partial \gamma(\lambda, \lambda, w)}{\partial \lambda} \right\}. \tag{5.41}$$

The quantity in curly braces can be written, after getting the derivatives as,

$$\lambda^{2\lambda} \left\{ \int_{t=w}^{\infty} \int_{u=0}^{w} (tu)^{\lambda-1} e^{-\lambda(t+u)} \left[\log(t/u) - (t-u) \right] du dt \right\}. \tag{5.42}$$

So, the function $F(w, \lambda, \lambda)$ is monotonic decreasing in λ if $\log(t/u) - (t-u) > 0$ in the integration domain, i.e. if

$$\int_{t=w}^{\infty} te^{-t}dt > \int_{t=0}^{w} te^{-t}dt. \tag{5.43}$$

Apparently, the above condition is fulfilled if

$$w < \text{median of Gamma}(2, 1) = 1.678. \tag{5.44}$$

Hence, from the positive skewness of a gamma distribution, and for all $0 < \alpha < 0.5$,

$$w_\alpha < w_{0.5} < E(W) = 1 < 1.678, \qquad \forall \lambda > 0, \tag{5.45}$$

where w_α is the α-quantile of W.

From (5.44) and (5.45) we can see that $F(w, \lambda, \lambda)$ is strictly monotonic decreasing in λ for all quantiles w, such that $w < w_{0.5}$. However, we believe that the expert can efficiently quantify her opinion about quartiles more easily by using the bisection method, see for example Pratt *et al.* (1995). So, we choose the lower quartile, $w_{0.25}$, as a monotonic function in λ since the function $F(w_{0.25}, \lambda, \lambda)$ is decreasing in λ. Note that the opposite is not true, i.e. if $w > w_{0.5}$ then w is not necessarily greater than 1.678, and no monotonicity is guaranteed for $w_{0.75}$, for example.

Another reason for choosing the lower quartile and not the upper quartile, beside monotonicity as discussed above, is that the lower quartile is more sensitive than the upper quartile to changes in the the shape parameter λ at any fixed value of the mean. Figure 5.4 illustrates this fact; it shows the changes in both the lower and upper quartiles of gamma distributions due to the change of its parameter value λ, for different fixed mean values at 0.5, 5, 50, and 500. It can be seen from Figure 5.4 that the lower quartile is more sensitive than the upper quartile to the changes in λ at fixed mean values.

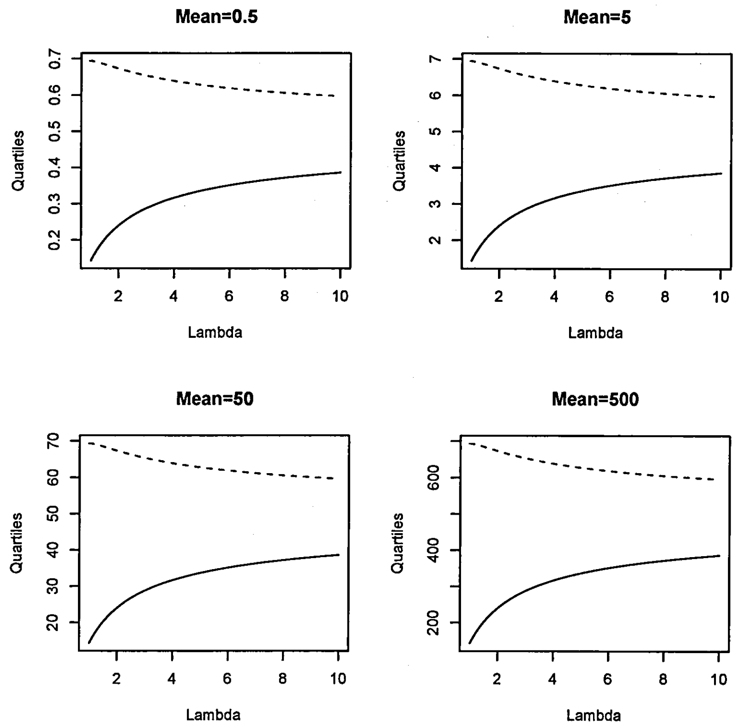


Figure 5.4: Changes in quartile values with the change of λ at different mean values.

Now, since $F(w, \lambda, \lambda)$ is strictly monotonic increasing in w and strictly monotonic decreasing in λ, for $w < w_{0.5}$, then fixing $F(w, \lambda, \lambda) = 0.25$, the lower quartile $w_{0.25}$ is an implicit monotonic increasing function in λ, say

$$w_{0.25} = h^*(\lambda). \tag{5.46}$$

Hence, from (5.34), we have

$$Q_1 = \mu h^*(\lambda) = h(\lambda), \tag{5.47}$$

where Q_1 is the lower quartile of ζ, and $h(.)$ is a monotonic increasing function of λ.

The expert will be asked to assess three quartiles of her prior distribution for Q_1. Then, from the monotonicity of $h(.)$ in (5.47), these quartiles can be transformed into the corresponding three quartiles of λ. We assume that the prior distribution of λ is a lognormal distribution, and use the three transformed quartiles to solve for the two parameters of the lognormal distribution. The required assessment tasks to implement this method using interactive graphical software are detailed in the next section.

5.3.2 Assessment tasks

The expert is questioned about the lower quartile of the gamma distribution, Q_1 say. However, she is not simply asked to give a point estimate of Q_1 - she is asked to give assessments that quantify her uncertainty about it. Specifically, she is asked to give her lower and upper quartiles for Q_1 in addition to her median assessment of its value. Questions that make this a meaningful task that an expert can reasonably be asked to perform are suggested later.

- Three quartiles of Q_1 will be assessed by the expert, say $Q_{1,1}$, $Q_{1,2}$ and $Q_{1,3}$, where the median $Q_{1,2}$ is a point estimate of Q_1, and $Q_{1,3} - Q_{1,1}$ is its interquartile range. Details on how to ask about these quartiles are given later.

- Under the monotonicity of $h(.)$ in (5.47), the three assessed quartiles $Q_{1,1}$, $Q_{1,2}$ and $Q_{1,3}$ of Q_1 can be transformed to the three corresponding quartiles of $\lambda|\mu$, say $Q_{\lambda,1}$, $Q_{\lambda,2}$ and $Q_{\lambda,3}$, respectively.

- Hence, we obtain the three quartiles $Q_{\lambda,1}$, $Q_{\lambda,2}$ and $Q_{\lambda,3}$ of the prior distribution of λ given μ, as

$$Q_{\lambda,i} = h^{-1}(Q_{1,i}), \quad i = 1, 2, 3, \tag{5.48}$$

  where $h^{-1}(.)$ can be implemented by numerically inverting the incomplete gamma function $F(w, \lambda, \lambda)$ via a simple search procedure.

- From (5.47) and (5.48), if the three assessed values $Q_{1,1}$, $Q_{1,2}$ and $Q_{1,3}$ are the three quartiles of Q_1, then $Q_{\lambda,1}$, $Q_{\lambda,2}$ and $Q_{\lambda,3}$ are the three corresponding quartiles of $\lambda|\mu$, respectively. Clearly

$$\begin{aligned} \Pr\{Q_1 < Q_{1,i}\} &= \Pr\{(\lambda|\mu) < h^{-1}(Q_{1,i})\} \\ &= \Pr\{(\lambda|\mu) < Q_{\lambda,i}\} = 0.25(i), \quad i = 1, 2, 3. \end{aligned} \tag{5.49}$$

- We assume that the prior distribution of λ given μ is a lognormal distribution with two hyperparameters a and b of the form

$$f(\lambda|\mu) = \frac{1}{\lambda b\sqrt{2\pi}} \exp\left[\frac{-(\ln\lambda - a)^2}{2b^2}\right], \quad \lambda, b > 0. \tag{5.50}$$

The properties of the normal distribution are used to estimate a and b from the transformed assessments $Q_{\lambda,i}$, $i = 1, 2, 3$.

- Since, from the assumed lognormal prior distribution in (5.50), we have

$$(\ln \lambda|\mu) \sim \mathrm{N}(a, b), \tag{5.51}$$

and using the fact that $b = \mathrm{IQR}/1.349$, then clearly

$$a = \ln(Q_{\lambda,2}), \quad b = \frac{\ln(Q_{\lambda,3}) - \ln(Q_{\lambda,1})}{1.349}. \tag{5.52}$$

- The prior structure of the gamma GLM parameters take the form

$$f(\mu, \lambda) = f(\mu) \times f(\lambda|\mu), \tag{5.53}$$

where $f(\mu)$ can be obtained from (5.31), and $f(\lambda|\mu)$ is given as lognormal(a, b).

This elicitation method has been implemented in graphical user-friendly software that automatically estimates the two hyperparameters of the lognormal distribution. The software has been developed as an add-on to the PEGS-GLM (Correlated Coefficients) software for eliciting the scale parameter λ of the gamma GLM. It is also freely available at http://statistics.open.ac.uk/elicitation as a stand alone version, PEGS-Gamma, for eliciting the shape parameter λ of a gamma distribution with a known mean.

In the former case, the median μ_0 and and the lower quartile Q_1 of the response variable ζ at the reference point have already been elicited, see (5.32). For the latter case, the expert is asked, in a dialogue box, to assess her mean value μ_0 and the lower quartile Q_1 of the gamma random variable. In both cases, these two assessments represents the first assessment step, from which the software suggests reasonable initial values for the other two required assessments.

The median value $Q_{1,2}$ is set equal to the assessed value of Q_1, while the other two quartile values $Q_{1,1}$ and $Q_{1,3}$ are suggested as

$$Q_{1,1}^0 = Q_{1,2} - \frac{1}{3}\min(Q_{1,2}, \mu_0 - Q_{1,2}), \tag{5.54}$$

$$Q^0_{1,3} = Q_{1,2} + \frac{1}{3}\min(Q_{1,2}, \mu_0 - Q_{1,2}). \tag{5.55}$$

These initial suggested values are used in (5.47) and (5.49) to get the three quartiles $Q^0_{\lambda,1}$, $Q_{\lambda,2}$ and $Q^0_{\lambda,3}$ of the parameter λ, respectively. The inversion of (5.47) is done by the software through a simple search procedure.

As in (5.52), these quartiles are used to compute the two hyperparameters a and b of the assumed lognormal distribution of λ. Using a and b, the mean value of λ, say μ_λ, is computed from the lognormal distribution of λ:

$$\mu_\lambda = \exp(a + \frac{1}{2}b^2). \tag{5.56}$$

Then μ_λ is used with the assessed mean value μ_0 to draw the pdf graph of the gamma distribution, Gamma$(\mu_\lambda, \mu_\lambda/\mu_0)$. A main panel is presented to the expert showing this pdf graph; see the upper graph of Figure 5.5. The thick black line on this graph represents the mean value μ_0.

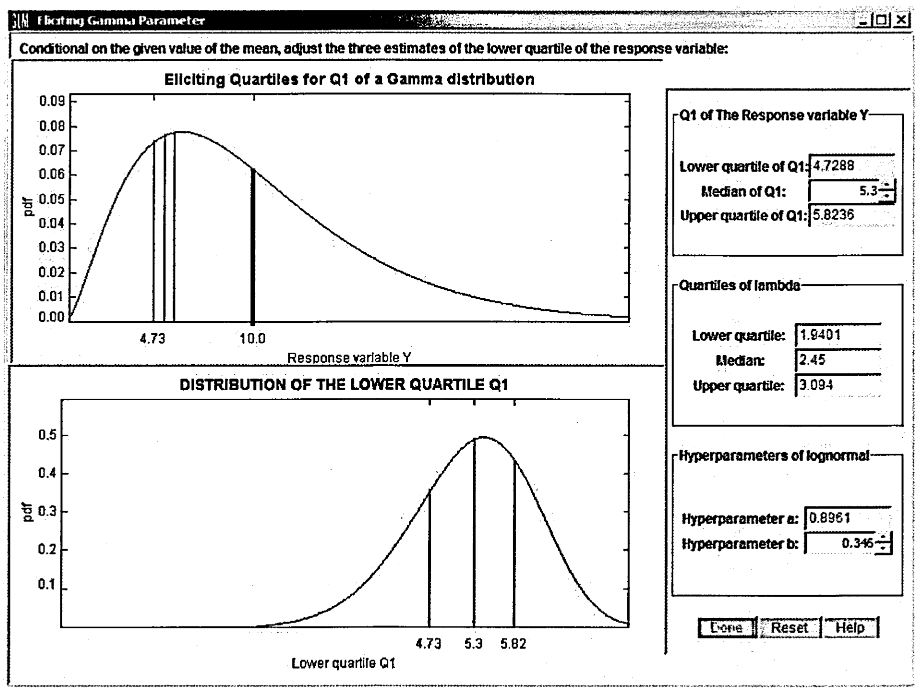


Figure 5.5: The main software panel for assessing gamma parameter

For statistical coherence of the assumed normal distribution of $\ln(\lambda)$, the two normal quartiles $\ln(Q^0_{\lambda,1})$ and $\ln(Q^0_{\lambda,3})$ should be symmetrical around the normal mean, $a = \ln(Q_{\lambda,2})$.

To attain this, we assume that the expert is always more confident in assessing the median value, than assessing the other two quartiles. So we treat her original and transformed medians $Q_{1,2}$ and $Q_{\lambda,2}$, respectively, as being correct. Then we suggest two coherent sets of quartiles $Q_{1,1}$, $Q_{1,3}$ and $Q_{\lambda,1}$, $Q_{\lambda,3}$ to replace the initial assessments $Q^0_{1,1}$, $Q^0_{1,3}$ and $Q^0_{\lambda,1}$, $Q^0_{\lambda,3}$, respectively, as follows. First, $Q_{\lambda,1}$, $Q_{\lambda,3}$ are computed as the actual first and third quartiles, respectively, of a lognormal distribution with the two elicited parameters a and b. Then $Q_{1,1}$, $Q_{1,3}$ are computed from $Q_{\lambda,1}$, $Q_{\lambda,3}$, respectively, using (5.47) and (5.49).

The first group of values in the right-hand side panel of Figure 5.5 gives the values of the three suggested coherent quartiles $Q_{1,1}$, $Q_{1,2}$ and $Q_{1,3}$. These quartiles are also drawn as the three blue lines in the upper and lower pdf graphs of Figure 5.5. The second group of values gives the three quartiles of λ, $Q_{\lambda,1}$, $Q_{\lambda,2}$ and $Q_{\lambda,3}$. The elicited values of a and b are shown as the third group of values in the same panel.

The lower graph in Figure 5.5 represents the elicited distribution of the lower quartile Q_1, with the three vertical blue lines representing $Q_{1,1}$, $Q_{1,2}$ and $Q_{1,3}$. The graph is intended to help the expert check that the distribution is a reasonable representation of her prior knowledge of Q_1. Although we do not assume any specific family of distributions for Q_1, the pdf graph is drawn using pointwise numerical derivatives of the cdf of Q_1. This cdf is obtained as in (5.49), not only for the three quartile points, but also for a sufficiently large number of points. A set of 1000 points covering the whole range of Q_1 has been used.

Hence, Figure 5.5 shows all the assessed and suggested quartiles of Q_1 and λ, with the two corresponding values of a and b. The two pdf graphs of λ and Q_1 are also presented to the expert to show her the impact of these quartile values and hyperparameters on the two distributions. The main assessment task that the expert is asked to perform uses the following type of question. Let us suppose that the variable that has the gamma distribution is the period of time that a patient with some medical disorder may stay in hospital. Then the expert will be asked to consider the length of time that a hypothetical patient, John, will spend in hospital. She is told, "John has this disorder and will spend a time in hospital.

Suppose he is fortunate and does not spend as long as most people in hospital. Specifically, suppose exactly 25% of patients with John's disorder spend a shorter time in hospital than John. Give your median assessment for the length of time that John spends in hospital. Now give your lower and upper quartiles for this length of time."

The expert will be shown suggested coherent assessments and graphs. If she finds the suggestions a reasonable representation of her opinion, she can accept them, which finishes the assessment procedure. If they do not represent her opinion adequately, she has the option of directly reviewing the median value $Q_{1,2}$ of Q_1, or indirectly reviewing the quartiles $Q_{1,1}$ and $Q_{1,3}$ by changing the value of the hyperparameter b. As discussed before, for statistical coherence, changes must be made first to the value of b and then transformed into corresponding coherent changes in $Q_{1,1}$ and $Q_{1,3}$.

In principal, the expert can change $Q_{1,2}$ to any value in $(0, \mu_0)$, and she can change b to any positive value. However, to get a unimodal distribution for Q_1, some restrictions must be imposed on the values of a and b, as detailed below.

Although the relation between Q_1 and λ, as given in (5.47), is strictly monotonic increasing for all $\lambda > 0$, the numerical second derivative of $h(\lambda)$ reveals a critical point of zero at $\lambda = 0.5045$. Therefore, the pdf of Q_1 is not guaranteed to be unimodal if the elicited values of a and b lead to a non-neglectable probability of $\lambda < 0.5045$.

To avoid an undesirable appearance of the pdf of Q_1, we restrict the elicited lognormal hyperparameters a and b to satisfy

$$\frac{a + 0.684}{b} > 3.11. \tag{5.57}$$

This condition insures (from the standard normal distribution) that

$$\Pr\left\{\frac{\ln(\lambda) - a}{b} < \frac{-0.684 - a}{b}\right\} < 0.001, \tag{5.58}$$

i.e. it guarantees that

$$\Pr(\lambda < 0.5045) < 0.001. \tag{5.59}$$

If condition (5.57) is not satisfied, the right hand side panel on Figure 5.5 will only allow the expert to increase the value of $Q_{1,2}$, hence increasing $a = \ln(Q_{1,2})$, or directly decreasing the value of b.

A 'Reset' button is available for the expert to return at any time to the initial coherent set of suggestions and graphs and review them again if she needs to. When the expert is happy with the quartile values and the corresponding pdf graphs, she clicks 'Done' and obtains the two corresponding hyperparameters a and b as the output of her assessments.

5.4 Concluding comments

To elicit an informative prior distribution for normal and gamma GLMs, expert opinion must be quantified about both the regression coefficients and the extra parameters in these models. In this chapter, two elicitation methods have been proposed to quantify expert's opinion about a prior distribution of the random error variance in normal GLMs, and a prior distribution for the scale parameter in gamma GLMs.

A method of assessing a conjugate inverted chi-squared prior distribution for the error variance in normal models has been proposed. The method quantifies an expert's opinions through assessments of a median and conditional medians of the absolute difference between two observations of the response variable at the same design point. Conditional assessments have been based on various sets of hypothetical future samples. These assessments depend only on the random error and have been used to elicit the inverted chi-squared distribution. A computer program that implements the method is available as an option in the PEGS-GLM (Correlated Coefficient) software and also as an add-on to any other elicitation software for normal models, PEGS-Normal. Both versions are freely available at http://statistics.open.ac.uk/elicitation.

A novel method for eliciting a lognormal prior distribution for the scale parameter of a gamma GLM, or the shape parameter of any gamma distribution, has also been proposed. The method depends only on quantifying an expert's opinion about the lower quartile of a gamma distributed random variable. This lower quartile is itself a random variable; for which the expert assesses a median value as a point estimate and an interquartile range. An example of questions that can be addressed to the expert has been given. The interactive graphical PEGS-Gamma software implementing this method is user-friendly. It gives coherent suggestions for all the required assessments and presents instant graphical feedback. To the best of the author's knowledge, this is the first piece of interactive software that is designed for eliciting a prior distribution of the shape parameter of a gamma distribution or the scale parameter of a gamma GLM.

Chapter 6

Eliciting Dirichlet priors for multinomial models

6.1 Introduction

Multinomial models, consisting of items that belong to a number of complementary and mutually exclusive categories, arise in many scientific disciplines and industrial applications. For example, they are frequently encountered in geology for different compositions of rocks, in microeconomics for patterns of consumer selection preferences, in political science for voting behavior. Other application areas include medicine, psychology and biology.

For mathematical coherence, the probabilities of each category must be non-negative and satisfy a unit-sum constraint. The multinomial distribution describes this model as a direct generalization of the binomial distribution to more than two categories.

It is well-known that the Dirichlet distribution is a conjugate prior for the parameters of multinomial models. The distribution preserves the unit sum constraint of multinomial probabilities and imposes a simple Dirichlet pattern of dependency between them. This structure gives negative correlations between the probabilities of categories, as will be shown later.

A different way of thinking about prior distributions for multinomial models is to use the multivariate normal distribution as a large sample approximation to the Dirichlet distribution or to the distribution of the log contrasts of the multinomial probabilities. Another option is to estimate the exact distribution of log contrasts using a Monte Carlo sample. Generalized, nested or mixed forms of the Dirichlet distribution have been also introduced and suggested as suitable priors for multinomial models. For more details on possible prior distributions for multinomial models see, for example, O'Hagan and Forster (2004).

Eliciting parameters of multivariate distributions is not, in general, an easy task. It is even more complex when the variates are not independent, in which case summaries of the marginal distributions should be assessed, together with effective and reliable summaries of the dependence structure of the joint distribution [O'Hagan *et al.* (2006)]. In this chapter, our proposed method makes use of assessments of marginal beta distributions. Decomposition of the Dirichlet elicitation process into the assessment of several marginal beta distributions helps reduce the complexity of eliciting a multivariate distribution.

In Section 6.2, we develop a method of quantifying opinion about a beta prior distribution by the assessment of three quartiles. The method will be generalized to elicit a Dirichlet distribution in Section 6.3. The elicited beta univariate distribution will also be used to construct more flexible distributions in the next chapter, including the generalized Dirichlet prior and a Gaussian copula function for the prior distribution.

6.2 Eliciting beta parameters using quartiles

6.2.1 Introduction

The beta distribution is widely used in Bayesian analysis as a conjugate prior for the probability of success in Bernoulli trials. The domain of definition for the beta distribution of the first type is the interval [0,1], which is appropriate for the probability parameter of Bernoulli and binomial distributions. Moreover, the beta distribution is also a conjugate prior for Bernoulli and binomial sampling distributions, so that the posterior distribution is obtained through simple arithmetic. The wide range of valid values of the two hyperparameters of the beta prior gives it great flexibility and its pdf has varied shapes. In this sense, the beta distribution is more likely to be a reasonable model of the expert's opinion compared with other priors such as the uniform distribution over the interval [0,1] or the triangular distribution suggested by van Dorp and Kotz (2002).

It seems that eliciting beta parameters is the most studied elicitation problem to date, whether it is a beta prior for Bernoulli or Binomial sampling distributions, a distribution of a probability of an event, or a proportion that ranges between zero and one. There are many methods available in the literature for eliciting beta distribution parameters. A comprehensive literature review may be found in Hughes and Madden (2002), Jenkinson (2007) or O'Hagan *et al.* (2006).

The available methods for beta elicitation can be classified into two general classes of elicitation methods, variable interval and fixed interval. In the variable interval methods, the probability is fixed and the expert assesses an interval that gives this probability. In the fixed interval methods, the interval is fixed and the expert assesses the probability that the event of interest will be in that interval. Asking about quartiles is an example of the first methods, while assessing probabilities is an example of the second class of methods.

Beta elicitation methods vary in the quantities that the expert must assess. She may be asked to assess a location value such as the mean, the median or the mode. Also, a scale value must be assessed, such as the probability of being in an interval, the boundaries of an interval, or the mean absolute deviation about a location value. These quantities may be converted into the hyperparameters in exact forms or through numerical approximation.

Regarding the number of required assessments, most of the available methods use only two assessed quantities, usually one for location and the other for scale. These give estimates of the two beta parameters. Although only two assessments are mathematically needed to elicit two unique parameters, some methods use over-fitting through assessing three or more quantities, followed by some sort of averaging or reconciliation.

In this section we propose a new method of eliciting the parameters of a beta prior distribution for the binomial success probability. Assessments of the median and two quartiles are elicited. A compromise is needed to reconcile these three assessments into two unique parameters. We use a normal approximation to the beta distribution to estimate initial values of the beta parameters, followed by a least-squares technique to optimize the two initial values. According to the classifications given above, the proposed method is a variable interval method that uses three assessments, a median and two quartiles.

We believe that it is better to elicit a median as a location value and quartiles for scale, than, say, to elicit a mean and other quantiles. The median and quartiles are easier for an expert to assess as they are obtained by the first two steps of equally likely subdivisions (bisection method). The expert can be asked about the median as the value that the probability of success is equally likely to be above or below. Then we ask the expert to sub-divide the interval above the median into two equally likely intervals for the probability; her assessed value is her upper quartile. The same concept is used for the interval below the median in order to obtain her lower quartile.

van Dorp and Mazzuchi (2000, 2003, 2004) introduced a numerical algorithm and software to specify the parameters of the beta distribution and its Dirichlet extension using quantiles. They used the median as a measure of central tendency with any other single quantile as a measure of dispersion. Although they proved that this suffices mathematically for the existence of a unique solution for beta parameters, it is more useful in elicitation contexts to use over-fitting as a means towards better representation of an expert's opinion.

6.2.2 Normal approximations for beta elicitation

To estimate the two parameters of the beta distribution using three assessed quartiles, we propose a two step approach. In the first step, a normal approximation for the beta distribution is used to transform and reconcile the three assessed quartiles as two initial values for the beta parameters. In the second step, a numerical least-squares method is applied to the initial parameter values so as to optimize them. The aim is to find parameter values that give nominal quartiles that are as close as possible to the assessed values. This section is devoted to the proposed normal approximation, while the least-squares optimization is discussed in Section 6.2.3 below.

A method that directly fits a beta distribution to the assessed median and two quartiles is given in Pratt *et al.* (1995). They used a normal approximation for the beta distribution together with averaging. The method was also used as the main assessment method in a study of the effect of feedback and learning on the assessment of subjective probability distributions (Staël von Holstein, 1971). Our proposed method adopts the technique of Pratt *et al.* (1995), but with a different normal approximation and a new compromise to get initial parameter values. We also add a least-squares optimization technique. In what follows, we summarize the argument of Pratt *et al.* (1995) and then propose a different normal approximation and a different compromise.

Let p be the success probability of concern, and assume that p has a conjugate standard beta prior distribution of the form

$$f(p) = \frac{1}{\beta(a,b)} p^{a-1}(1-p)^{b-1}, \quad 0 \le p \le 1,\, a > 0,\, b > 0. \tag{6.1}$$

Pratt *et al.* (1995) stated that the transformation

$$Z = 2\left\{[p(b-1/3)]^{1/2} - [(1-p)(a-1/3)]^{1/2}\right\} \tag{6.2}$$

has approximately a standard normal distribution. Let q_i be the ith quartile of p that is assessed by the expert, for $i = 1, 2, 3$. Using the assessed lower quartile q_1 and the assessed median q_2, we get the following two equations from (6.2):

$$\Pr\left\{Z < 2\left\{[q_1(b-1/3)]^{1/2} - [(1-q_1)(a-1/3)]^{1/2}\right\}\right\} = 0.25, \tag{6.3}$$

$$\Pr\left\{Z < 2\left\{[q_2(b-1/3)]^{1/2} - [(1-q_2)(a-1/3)]^{1/2}\right\}\right\} = 0.5. \tag{6.4}$$

Solving (6.3) and (6.4) for a and b gives

$$a_1 = c_1 q_2 + \frac{1}{3} \tag{6.5}$$

and

$$b_1 = c_1(1-q_2) + \frac{1}{3}, \tag{6.6}$$

where

$$c_1 = 0.112\left\{[q_2(1-q_1)]^{1/2} - [q_1(1-q_2)]^{1/2}\right\}^{-2}.$$

Similarly, the assessed upper quartile, q_3, gives the equation

$$\Pr\left\{Z < 2\left\{[q_3(b-1/3)]^{1/2} - [(1-q_3)(a-1/3)]^{1/2}\right\}\right\} = 0.75. \tag{6.7}$$

Solving (6.4) and (6.7) for a and b gives

$$a_2 = c_2 q_2 + \frac{1}{3} \tag{6.8}$$

and

$$b_2 = c_2(1-q_2) + \frac{1}{3}, \tag{6.9}$$

where

$$c_2 = 0.112\left\{[q_2(1-q_3)]^{1/2} - [q_3(1-q_2)]^{1/2}\right\}^{-2}.$$

The compromise of Pratt *et al.* (1995) is simply to estimate a and b as the average of (6.5), (6.6), (6.8) and (6.9), i.e.

$$\begin{aligned} a &= \frac{a_1 + a_2}{2}, \\ b &= \frac{b_1 + b_2}{2}. \end{aligned} \tag{6.10}$$

However, Pratt *et al.* (1995) did not mention the theoretical derivation of the approximation in (6.2), nor its accuracy. So, we tried to use another approximation that is still mathematically tractable, but whose justification and accuracy have been investigated. Patel and Read (1982) give a good review of some accurate normal approximations to beta variables. They describe the following normal approximation as a simple yet accurate approximation.

If p has a beta distribution of the form in (6.1), then the transformation

$$Z = 2\left\{[p(b-1/4)]^{1/2} - [(1-p)(a-1/4)]^{1/2}\right\}, \tag{6.11}$$

has an approximate standard normal distribution. The absolute error of this approximation is of order

$$O\left(\frac{1}{\sqrt{a+b}}\right).$$

We adopt the approximation (6.11) to propose a new elicitation method for the beta parameters a and b using the three assessed quartiles q_i, $i = 1, 2, 3$.

Instead of direct averaging, we introduce a new compromise, making use of the characteristics of the normal distribution. In fact, it is well-known that

$$z_{0.75} - z_{0.25} = 1.349, \tag{6.12}$$

where $z_{0.25}$ and $z_{0.75}$ are the lower and upper quartiles of the standard normal distribution, respectively.

In view of the approximation (6.11), we have

$$[q_2(b-1/4)]^{1/2} - [(1-q_2)(a-1/4)]^{1/2} = 0. \tag{6.13}$$

$$z_{0.25} = 2\left\{[q_1(b-1/4)]^{1/2} - [(1-q_1)(a-1/4)]^{1/2}\right\}, \tag{6.14}$$

$$z_{0.75} = 2\left\{[q_3(b-1/4)]^{1/2} - [(1-q_3)(a-1/4)]^{1/2}\right\}. \tag{6.15}$$

Substituting with (6.14) and (6.15) in (6.12) we get the new compromise between q_1 and q_3 as

$$\begin{aligned}&\left\{[q_3(b-1/4)]^{1/2} - [(1-q_3)(a-1/4)]^{1/2}\right\} - \\ &\left\{[q_1(b-1/4)]^{1/2} - [(1-q_1)(a-1/4)]^{1/2}\right\} = \frac{1.349}{2}.\end{aligned} \tag{6.16}$$

Solving (6.13) and (6.16) for a and b, we get

$$a = cq_2 + \frac{1}{4} \tag{6.17}$$

and

$$b = c(1-q_2) + \frac{1}{4}, \tag{6.18}$$

where

$$\begin{aligned}c = \frac{(1.349)^2}{4} \{ &[q_2(1-q_1)]^{1/2} - [q_1(1-q_2)]^{1/2} + \\ &[q_3(1-q_2)]^{1/2} - [q_2(1-q_3)]^{1/2}\}^{-2}.\end{aligned}$$

We argue that our method preserves the assessed median value and the only compromise is between the two quartiles. We believe this will represent the expert's opinion better. The expert usually assesses her median with more certainty and less bias than her lower and upper quartiles. By using the new compromise of quartiles in (6.16) and keeping the median equation (6.13) fixed, we reflect the probable greater accuracy of the median assessment.

According to the accuracy of the normal approximation, the proposed initial values of the beta parameters, given in (6.17) and (6.18), lead to nominal values for the beta quartiles that are close to the assessed quartiles. However, they are not guaranteed to be the parameter values that minimize the differences between nominal and assessed quartile values. This is not ideal, so we just treat equations (6.17) and (6.18) as giving initial parameter values that can be improved upon.

6.2.3 Least-squares optimizations for beta parameters

Oakley (2010) gave a least-squares method for choosing beta parameters a and b that minimize

$$Q = [F(q_1, a, b) - 0.25]^2 + [F(q_2, a, b) - 0.5]^2 + [F(q_3, a, b) - 0.75]^2 , \qquad (6.19)$$

where $F(x, a, b)$ is the cdf of a beta distribution with parameters a and b at the point x.

The same approach has been implemented in the SHELF elicitation framework developed in Oakley and O'Hagan (2010). They introduced an R package of templates and software for conducting elicitation, within which minimizing Q in (6.19) was used to estimate beta parameters from assessed quartiles. However, they do not use any explicit normal approximation to a beta distribution when deriving the initial estimates of the beta parameters. Instead, they just transform the assessed beta quartiles into the mean and variance of a normal distribution, as if the quartiles were assessed for a normal distribution. The mean and variance are then assumed to be those of a beta distribution, from which initial values for the parameters can be computed.

Our accompanying elicitation software, PEGS-Dirichlet, implements programs written by Flanagan (2011) for the Java scientific library. These numerically minimize (6.19), which cannot be minimized analytically. They use a multidimensional technique called the downhill simplex method. The method was introduced by Nelder and Mead (1965) as a quick multidimensional minimization method that uses only function evaluations, not derivatives.

To constrain beta parameters to be positive, we transform them to a logarithmic scale. Hence we actually minimize

$$\begin{aligned} Q = \{F[q_1, \exp(a^*), \exp(b^*)] - 0.25\}^2 + \{F[q_2, \exp(a^*), \exp(b^*)] - 0.5\}^2 \\ + \{F[q_3, \exp(a^*), \exp(b^*)] - 0.75\}^2 , \end{aligned}$$

for a^* and b^*, with initial values as in (6.17) and (6.18), but on the logarithmic scale, i.e. $\log(a)$ and $\log(b)$. The final resulting beta parameter values are thus $\exp(a^*)$ and $\exp(b^*)$.

Our elicitation software, PEGS-Dirichlet, presents an interactive graph to the expert showing the previously assessed probability medians of all categories. The expert is asked to assess a lower and an upper probability quartile for each category by clicking on the graph. Once the two required quartiles are assessed for any single category, the proposed method of beta parameter elicitation is implemented by the software on the probability of this category. A pop up window opens showing the pdf graph of the elicited beta distribution with the location of the three assessed quartiles. This gives instant feedback to the expert, see Figure 6.1.

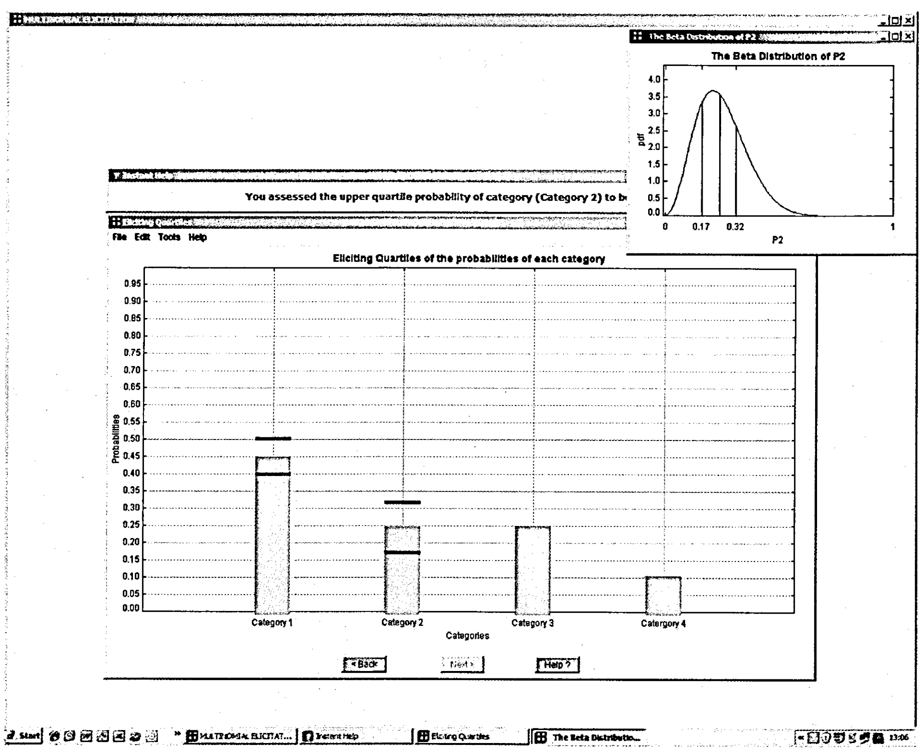


Figure 6.1: Assessing probability quartiles of each category

If she is not satisfied with the fitted beta distribution, the expert can simply change her assessments of the two quartiles. The whole elicitation process is applied again whenever the expert changes her quartile assessments. The pdf curve is interactively changing to show the direct impact of changing quartiles.

On finishing the elicitation process for all categories, the beta parameters are then compromised to estimate the Dirichlet hyperparameter vector as discussed in Section 6.3, below.

6.3 Eliciting a Dirichlet prior for a multinomial model

6.3.1 Introduction

A limited number of attempts have been made to develop elicitation methods for Dirichlet parameters, see Chapter 2 for more details. Jenkinson (2007) and O'Hagan *et al.* (2006) discussed two methods for Dirichlet elicitation. Namely, the method of Dickey *et al.* (1983) and that of Chaloner and Duncan (1987).

The elicitation method suggested by Dickey *et al.* (1983) starts by assessing the probability of each category directly from the expert. She will then be given a hypothetical future sample of a fixed size and told the number of items in each category. She is asked to re-assess the probabilities given this hypothetical sample. The equivalent sample size that corresponds to her prior knowledge can thus be estimated using Bayes' theorem.

Chaloner and Duncan (1983) give a method for eliciting a beta distribution. Chaloner and Duncan (1987) generalize this method and give an interactive graphical tool for Dirichlet elicitation. This is based on assessing the sample size and the modal values of Dirichlet variates, and then giving feedback to adjust the parameter values.

As mentioned before, van Dorp and Mazzuchi (2003, 2004) introduced a numerical algorithm that yields the Dirichlet parameters from quantile assessments. Their algorithm uses k quantile assessments to estimate all the parameters of a k-dimension Dirichlet distribution. However, we believe that it is better to assess more than k quantiles and then apply some form of reconciliation to estimate the parameters.

Assuming a Dirichlet prior for the success probabilities is one way of reconciling separate marginal beta prior distributions. Eliciting a Dirichlet prior by using assessed beta marginal distributions was outlined in Bunn (1978, 1979). However, his elicitation method used the hypothetical future sample technique. He stated that the application of the usual univariate quantile methods may generally be difficult and tedious in practice because of the multivariate nature of the Dirichlet distribution. However, the availability of interactive graphs and efficient computing enables us to use the quantile method in an elicitation method that is easy for the assessor and quick.

In what follows, we propose some reconciliation methods, based on the Dirichlet distribution, of combining beta marginals that have already been assessed using the method introduced in Section 6.2.

6.3.2 The multinomial and Dirichlet distributions

Let the random vector $\mathbf{X} = (X_1, \ X_2, \ \cdots, \ X_k)$ be multinomially distributed with k categories, n trials and a vector of probabilities $\underline{p} = (p_1, \ p_2, \ \cdots, \ p_k)$, taking the form

$$f(x_1, x_2, \cdots, x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \tag{6.20}$$

$$0 \leq x_i \leq n, \quad \sum x_i = n, \quad 0 \leq p_i \leq 1, \quad \sum p_i = 1,$$

or, equivalently, in the form

$$f(x_1, x_2, \cdots, x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_{k-1}^{x_{k-1}} (1 - p_1 - p_2 - \cdots - p_{k-1})^{x_k}, \tag{6.21}$$

$$0 \leq x_i \leq n, \quad \sum x_i = n, \quad 0 \leq p_i \leq 1, \quad \sum p_i \leq 1.$$

A conjugate prior for the parameter vector $\underline{p}$ is the Dirichlet distribution, which has the form

$$\pi(p_1, p_2, \cdots, p_k) = \frac{\Gamma(N)}{\Gamma(a_1)\Gamma(a_2)\cdots\Gamma(a_k)} p_1^{a_1-1} p_2^{a_2-1} \cdots p_k^{a_k-1}, \tag{6.22}$$

$$0 \leq p_i \leq 1, \quad \sum p_i = 1, \quad a_i > 0, \quad N = \sum a_i,$$

or, equivalently, the form

$$\pi(p_1, p_2, \cdots, p_{k-1}) = \frac{\Gamma(N)}{\Gamma(a_1)\Gamma(a_2)\cdots\Gamma(a_k)} p_1^{a_1-1} p_2^{a_2-1} \cdots (1 - p_1 - p_2 \cdots - p_{k-1})^{a_k-1}, \tag{6.23}$$

$$0 \leq p_i \leq 1, \quad \sum p_i < 1, \quad a_i > 0, \quad N = \sum a_i.$$

It is well-known that the expectations, variances and covariances of the Dirichlet variates p_i, for $i = 1, 2, \cdots, k$, are given by

$$E(p_i) = \frac{a_i}{N}, \tag{6.24}$$

$$\mathrm{Var}(p_i) = \frac{a_i(N - a_i)}{N^2(N+1)}, \tag{6.25}$$

$$\mathrm{Cov}(p_i, p_j) = \frac{-a_i a_j}{N^2(N+1)}. \tag{6.26}$$

To elicit the vector of hyperparameters $\underline{a} = (a_1, \quad a_2, \quad \cdots, \quad a_k)$, we use the direct relation between the Dirichlet distribution and its special univariate case, the beta distribution. We have already developed, in Section 6.2, a method of eliciting the two hyperparameters of a beta distribution. The hyperparameters of the Dirichlet distribution can be induced from those of the univariate beta distributions through some form of reconciliation. This can be done using either the standard marginal beta distributions of the multinomial probabilities, or the conditional scaled beta distribution of each of them. In what follows, these two proposed approaches are given in detail.

6.3.3 The marginal approach

Consider the form in (6.20) for the multinomial distribution with the conjugate prior Dirichlet distribution in (6.22). It is well-known that, from (6.20), the marginal distribution of each x_i is a binomial distribution with the two parameters

$$n_i = n, \quad p_i, \quad i = 1, 2, \cdots, k.$$

It is straightforward to show, using the Dirichlet pdf in (6.22), that the marginal distribution of each p_i is a beta distribution:

$$p_i \sim \mathrm{beta}(\alpha_i, \beta_i), \quad \text{for } i = 1, 2, \cdots, k, \tag{6.27}$$

where

$$\begin{aligned} \alpha_i &= a_i, & \text{for } i = 1, 2, \cdots, k, \\ \beta_i &= \sum_{j \neq i}^{k} a_j, & \text{for } i = 1, 2, \cdots, k. \end{aligned} \tag{6.28}$$

Assessment tasks

Exploiting the beta marginal distributions, the elicitation process may be divided into k steps. At each step, the expert will be asked to assess three quartiles for p_i, the binomial probability of category i ($i = 1, 2, \cdots, k$). See Figure 6.1, where the lower and upper quartiles have already been elicited for the first two categories. These quartiles can then be used to estimate the two hyperparameters α_i and β_i of the beta prior distribution of p_i, as proposed in Section 6.2. Since we use the marginal approach, the categories here are interchangeable. It does not matter where to start assessing nor the order of the categories.

To reconcile these separate marginal beta distribution into a Dirichlet distribution, we use a least-squares technique as follows.

Least-squares techniques

It is clear that the system of equations in (6.28) does not have a consistent solution, $\underline{a} = (a_1,\ a_2,\ \cdots,\ a_k)$. From (6.28), each marginal step of the elicitation process provides estimates of a_i and N_i, namely

$$a_i = \alpha_i, \qquad \text{for } i = 1, 2, \cdots, k, \tag{6.29}$$

and

$$N_i = \alpha_i + \beta_i = \sum_{j=1}^{k} a_j, \qquad \text{for } i = 1, 2, \cdots, k. \tag{6.30}$$

The estimated hyperparameters must fulfill the unit sum constraint of the probability expectations, i.e. they must satisfy

$$\sum_{i=1}^{k} \mu_i = 1,$$

where

$$\mu_i = \frac{a_i}{N_i}, \quad i = 1, 2, \cdots, k. \tag{6.31}$$

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. Following their approach, we propose the following options for reconciling different incoherent estimates of μ_i and N, yielding coherent estimates μ_i^* and N^*, respectively.

Options for μ_i^*:

1. Normalize each μ_i, as required for the Dirichlet distribution, giving

$$\mu_i^* = \frac{\mu_i}{\sum_{j=1}^{k} \mu_j}, \quad i = 1, 2, \cdots, k. \tag{6.32}$$

2. Minimize the sum of squares of differences between μ_i^* and μ_i, $i = 1, 2, \cdots, k$, subject to the constraint $\sum_{i=1}^{k} \mu_i^* = 1$. This can be done using Lagrangian multipliers to minimize Q as follows.

$$\text{Minimize } Q = \sum_{i=1}^{k} (\mu_i^* - \mu_i)^2 + \lambda(\sum_{i=1}^{k} \mu_i^* - 1). \tag{6.33}$$

Solve for μ_i^*, giving

$$\mu_i^* = \mu_i + \frac{1 - \sum_{j=1}^{k} \mu_j}{k}, \quad i = 1, 2, \cdots, k. \tag{6.34}$$

However, the values of μ_i^* computed here using Lagrangian optimization are not guaranteed to be positive. If negative values are found, we replace the Lagrangian multipliers method with a numerical restricted minimization technique. The downhill simplex method of Nelder and Mead (1965) can also perform restricted minimization as follows.

$$\text{Minimize } Q = \sum_{i=1}^{k} (\mu_i^* - \mu_i)^2, \tag{6.35}$$

such that

$$0 < \mu_i^* < 1, \quad i = 1, 2, \cdots, k, \tag{6.36}$$

$$\sum_{i=1}^{k} \mu_i^* = 1. \tag{6.37}$$

To solve this restricted optimization problem, for μ_i^*, $i = 1, 2, \cdots, k$, our elicitation software, PEGS-Dirichlet, implements a program for minimization written by Flanagan (2011). The initial values for this method are obtained from (6.32).

3. The option in (6.34) changes each value of μ_i by adding a fixed amount. However, the precision of each estimate, i.e. the inverse of its variance, can be used as a weight

to reflect the expert's confidence in each of her assessments [Lindley *et al.* (1979)]. A constrained weighted least-squares procedures can be formulated as follows.

$$\text{Minimize } Q = \sum_{i=1}^{k} w_i(\mu_i^* - \mu_i)^2 + \lambda(\sum_{i=1}^{k} \mu_i^* - 1). \tag{6.38}$$

where

$$w_i = [\mathrm{Var}(p_i)]^{-1} = \left[\frac{\alpha_i\beta_i}{(\alpha_i+\beta_i+1)(\alpha_i+\beta_i)^2}\right]^{-1}, \quad i = 1, 2, \cdots, k. \tag{6.39}$$

Solving for μ_i^* gives

$$\mu_i^* = \mu_i + \frac{1 - \sum_{j=1}^{k}\mu_j}{w_i \sum_{j=1}^{k} w_j^{-1}}, \quad i = 1, 2, \cdots, k.$$

Again, the minimization method implementing the restricted downhill simplex method is used if negative values of μ_i^* are found:

$$\text{Minimize } Q = \sum_{i=1}^{k} w_i(\mu_i^* - \mu_i)^2, \tag{6.40}$$

under the same constraints given by (6.36) and (6.37), using initial values as in (6.32).

Options for N^*:

1. Since no constraints are imposed on N^*, minimizing the sum of squares

$$\text{Minimize } Q = \sum_{i=1}^{k} (N^* - N_i)^2,$$

gives the average

$$N^* = \frac{\sum_{i=1}^{k} N_i}{k}. \tag{6.41}$$

2. Using the same weights as in (6.39) gives the weighted average

$$N^* = \frac{\sum_{i=1}^{k} w_i N_i}{\sum_{j=1}^{k} w_j}, \tag{6.42}$$

as a solution of

$$\text{Minimize } Q = \sum_{i=1}^{k} w_i(N^* - N_i)^2.$$

Estimating μ_i^* and N^*, using any of the options listed above, makes it easy to estimate a_i by a_i^*, where

$$a_i^* = \mu_i^* N^*, \quad i = 1, 2, \cdots, k.$$

Implementation and feedback

We use three different combinations of the options given above as follows:

1. Direct normalization of μ_i^* as in (6.32) and the average N^* in (6.41).

2. Least-squares optimization for μ_i^* as in (6.33) or (6.35), and for N^* as in (6.41).

3. Weighted least-squares optimization for μ_i^* as in (6.38) or (6.40), and for N^* as in (6.42).

The software elicits three hyperparameter vectors of the Dirichlet distribution, one vector for each of the above combinations. Each vector is then used to compute the corresponding pairs of marginal beta parameters as given in (6.28). Three quartiles for each beta marginal are computed numerically for each different 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. The vector with the selected set of quartiles will be taken as the final elicited hyperparameter vector of the Dirichlet prior. See Figure 6.2, where the first two combinations are shown and the expert has selected the second one.

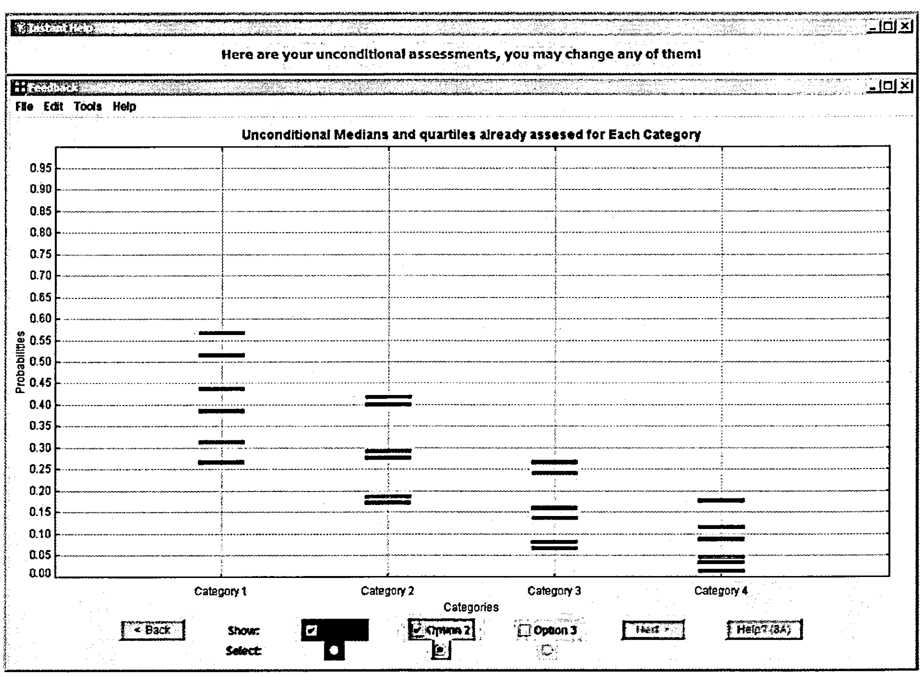


Figure 6.2: A feedback screen showing 2 different quartile options

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 as in Section 6.2, and the final Dirichlet hyperparameter vector is computed according to equations (6.29) - (6.32) and (6.41).

6.3.4 The conditional approach

Consider the form of multinomial distribution given in (6.21), with the form of conjugate Dirichlet distribution given in (6.23).

If

$$\underline{p}_{k-1} = (p_1 \quad p_2 \quad \cdots \quad p_{k-1}) \sim \text{Dirichlet}(a_1, a_2, \cdots, a_k),$$

then it can be shown [e.g. Wilks (1962)] that the marginal distribution of any subset of $\underline{p}_{k-1}$ is again a Dirichlet distribution, e.g.

$$\underline{p}_r = (p_1 \quad p_2 \quad \cdots \quad p_r) \sim \text{Dirichlet}(a_1, a_2, \cdots, a_r, \sum_{i=r+1}^{k} a_i), \quad 1 \leq r \leq k-1.$$

For $1 < r \leq k-1$, we can get the following conditional scaled beta distributions

$$\pi(p_r|p_1, p_2, \cdots, p_{r-1}) = \frac{\pi(\underline{p}_r; a_1, a_2, \cdots, a_r, \sum_{i=r+1}^{k} a_i)}{\pi(\underline{p}_{r-1}; a_1, a_2, \cdots, a_{r-1}, \sum_{i=r}^{k} a_i)} \tag{6.43}$$

$$= \frac{1}{\beta(a_r, \sum_{i=r+1}^{k} a_i)} \frac{p_r^{a_r-1}}{(1-\sum_{i=1}^{r-1} p_i)^{a_r}} \left(1 - \frac{p_r}{1-\sum_{i=1}^{r-1} p_i}\right)^{\sum_{i=r+1}^{k} a_i - 1}, \tag{6.44}$$

which are the scaled beta distributions over the intervals $(0, 1 - \sum_{i=1}^{r-1} p_i)$. The distributions in (6.44) are also known as three parameter beta distributions, i.e.

$$(p_r|p_1, p_2, \cdots, p_{r-1}) \sim \text{beta}(a_r, \sum_{i=r+1}^{k} a_i, 1 - \sum_{i=1}^{r-1} p_i), \qquad \text{for } 1 < r \leq k-1.$$

Applying the transformation

$$p_r^* = \begin{cases} p_1, & \text{for } r = 1, \\ \dfrac{p_r}{1 - \sum\limits_{i=1}^{r-1} p_i}, & \text{for } r = 2, 3, \cdots, k-1, \end{cases}$$

gives

$$p_r^* \sim \text{beta}(a_r, \sum_{i=r+1}^{k} a_i), \quad \text{for } r = 1, 2, \cdots, k-1. \tag{6.45}$$

Assessment tasks

The elicitation process is conducted as follows:

- The expert chooses the most convenient category to start with; we denote its probability as p_1.
- The expert assesses three quartiles for p_1, which are then converted into estimates of the two hyperparameters α_1 and β_1 of the beta distribution, beta(α_1, β_1).
- The expert is asked to assume that the median value she gave in the first step is the correct value of p_1, and she then assesses three quartiles for p_2. Figure 6.3 shows the graph after the median and lower quartile of the second category have been assessed by the expert, given the median of the first category as shown by the red bar.

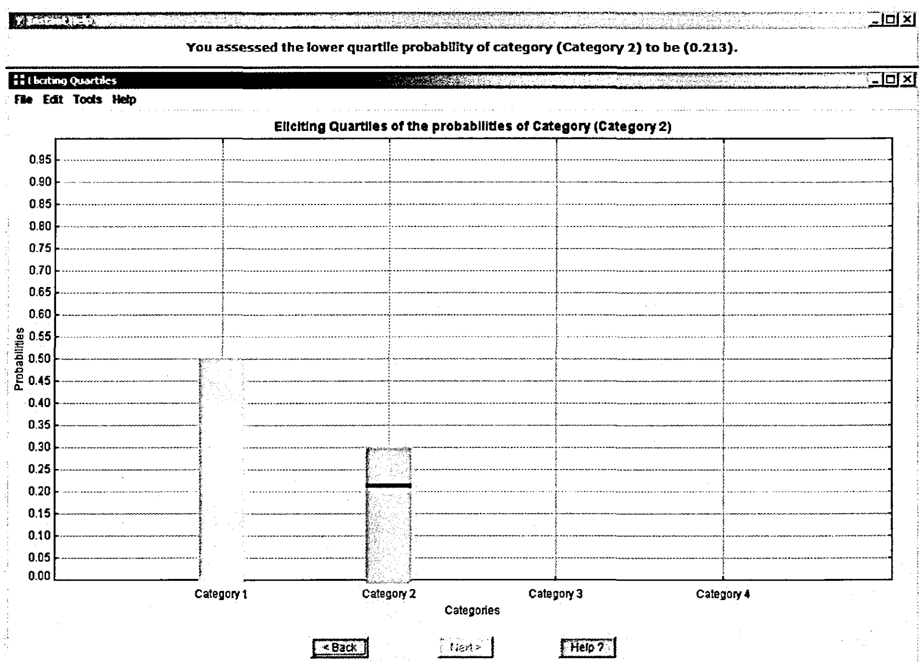


Figure 6.3: Assessing conditional quartiles for Dirichlet elicitation

- Dividing each of the three quartiles of p_2 by $1 - p_1$, we get the quartiles of p_2^*. Hence we obtain estimates of the hyperparameters α_2 and β_2 of the marginal beta distribution of p_2^*.

- The process is repeated for each category except for the last one. For $r = 3, 4 \cdots, k-1$, the expert gives quartiles for $(p_r|p_1, p_2, \cdots, p_{r-1})$. Dividing by $1 - \sum_{i=1}^{r-1} p_i$ gives the three quartiles of p_r^*, which are used to estimate the two hyperparameters α_r and β_r of its marginal distribution. (We do not require the marginal distribution of p_k.)

- To help the expert during this task, the software presents an interactive graph showing the pdf curve of the conditional beta distribution of $(p_r|p_1, \cdots, p_{r-1})$, for $r = 2, 3, \cdots, k-1$, see Figure 6.4. The expert is able to change her assessed conditional quartiles of p_r until she finds the conditional pdf curve an acceptable representation of her opinion.

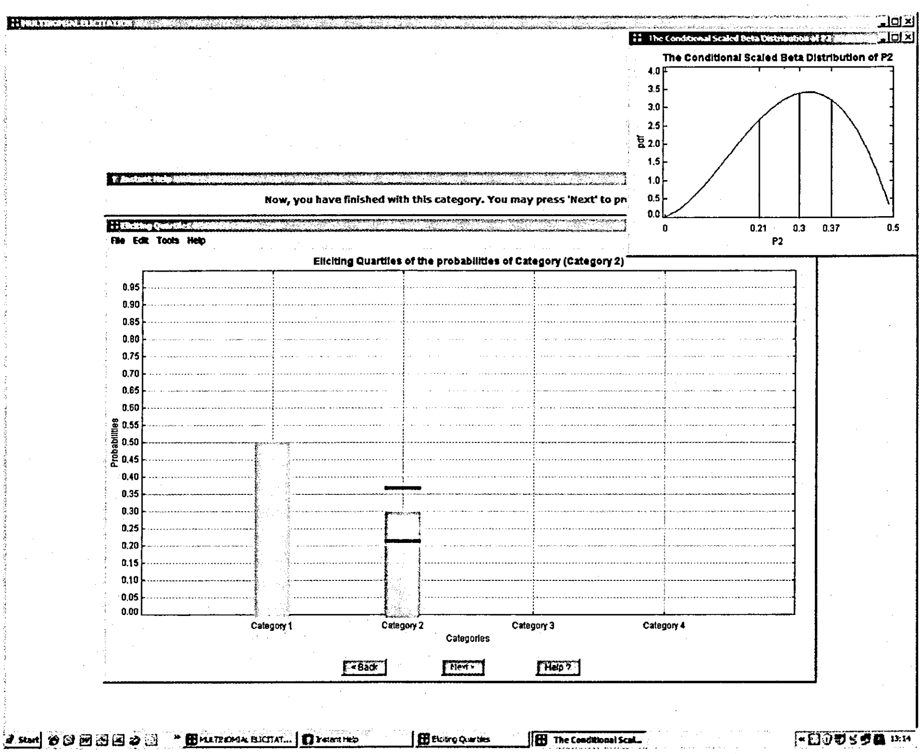


Figure 6.4: Assessing conditional quartiles with scaled beta feedback

Eliciting the hyperparameter vector

Using (6.45), we get the following system of equations

$$
\begin{aligned}
\alpha_r &= a_r, && \text{for } r = 1, 2, \cdots, k-1, \\
\beta_r &= \sum_{i=r+1}^{k} a_i, && \text{for } r = 1, 2, \cdots, k-1.
\end{aligned} \tag{6.46}
$$

Each elicited beta distribution has its own different estimate of N, given by

$$N_r = \sum_{i=1}^{r-1} \alpha_i + \alpha_r + \beta_r, \tag{6.47}$$

based on α_i, $i = 1, 2, \cdots, r-1$, which has been estimated in previous steps.

The system of equations in (6.46), as in the marginal approach, might not be consistent nor have a unique solution for $\underline{a} = (a_1, \ a_2, \ \cdots, \ a_k)$. So, we try to find a way of averaging this system to get a vector of estimates $\underline{a}^* = (a_1^*, \ a_2^*, \ \cdots, \ a_k^*)$ that is a good representation of the expert's opinion. We believe that keeping the mean value fixed, where possible, while moving from different beta distributions to a Dirichlet distribution may be a sensible approach.

Using (6.24), put

$$\mu_r \equiv E(p_r) = \frac{a_r}{N_r}.$$

Hence, in view of (6.46) and (6.47)

$$\mu_r = \begin{cases} \dfrac{\alpha_r}{\sum_{i=1}^{r} \alpha_i + \beta_r}, & r = 1, 2, \cdots, k-1, \\[2ex] \dfrac{\beta_{k-1}}{\sum_{i=1}^{k-1} \alpha_i + \beta_{k-1}}, & r = k. \end{cases} \tag{6.48}$$

Since, for the Dirichlet Distribution, it is required that

$$\sum_{r=1}^{k} \mu_r = 1,$$

we normalize the set of μ_r, for $r = 1, 2, \cdots, k$, to get

$$\mu_r^* = \frac{\mu_r}{\sum_{i=1}^{k} \mu_i}, \quad r = 1, 2, \cdots, k.$$

Moreover, let

$$N^* \equiv \sum_{r=1}^{k} a_r^*,$$

and take

$$a_r^* = \mu_r^* N^*, \quad r = 1, 2, \cdots, k.$$

It remains now to find a proper estimate of N^*. We take this as the average of all the denominators in (6.48):

$$N^* = \frac{\sum_{r=1}^{k-1}\left[\sum_{i=1}^{r}\alpha_i + \beta_r\right] + \sum_{i=1}^{k-1}\alpha_i + \beta_{k-1}}{k}.$$

Changing the expert's selection of the first category, as well as the order of conditioning categories at each step, will lead to different estimates of $\underline{a}$. To overcome this, one possibility is to repeat the whole process several times, using different starting categories and orderings. This will give sets of estimates $\underline{a}^*$'s, for which a simple averaging might give a suitable choice for $\underline{a}^*$. However, showing the marginal quartiles of the marginal beta distributions as a feedback to the expert and offering her the option of changing them seems another sensible option.

Feedback

The feedback process for the conditional approach is similar to that for the marginal approach. The main difference is in the relationship between the Dirichlet hyperparameters and beta parameters in the two approaches. To present the quartiles of each probability p_i, $i = 1, 2, \cdots, k$, as feedback to the expert after applying the conditional elicitation approach, we must first compute the parameters of the marginal beta distributions.

The two parameters α_i and β_i of each marginal beta distribution of p_i, $i = 1, 2, \cdots, k$, can be simply computed from the already elicited hyperparameter vector $\underline{a}^*$ of the Dirichlet distribution:

$$\alpha_i = a_i^*,$$
$$\beta_i = \sum_{j \neq i}^{k} a_j^* = N^* - a_i^*.$$

These can be used to compute numerically the three quartiles of each beta marginal distribution. The computed quartiles are then presented to the expert as feedback, see Figure 6.5.

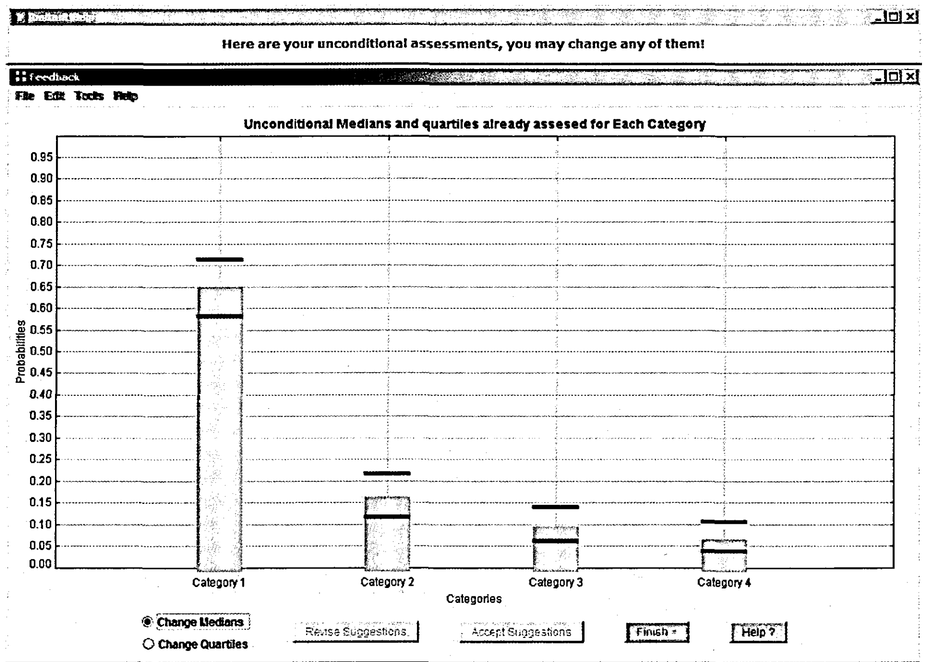


Figure 6.5: The feedback graph presenting marginal quartiles

The expert is asked to change any of the quartiles that do not satisfactorily represent her opinion. If any (or all) of these marginal quartiles are changed by the expert, we apply the marginal approach to re-elicit the Dirichlet hyperparameters as follows.

The new set of modified marginal beta quartiles are used to elicit new pairs of beta parameters as proposed in Section 6.2. Using these new parameters, together with equations (6.29), (6.30) and (6.31), we apply the first combination proposed in the marginal approach in Section 6.3.3. We implement the first combination that uses simple averaging as a quick and straightforward way to recompute the Dirichlet hyperparameter vector using the new set of modified quartiles. The whole process can be continuously applied until the expert is satisfied with the quartiles in the feedback.

6.4 Concluding comments

A reasonable method for eliciting beta parameters using quartiles has been proposed. The method combines two different approaches that have been used separately in the literature. A normal approximation was used to compute initial parameter values, which have then been optimized using a least-squares technique. In order to elicit the hyperparameter vector of the Dirichlet distribution, we made use of both the marginal and conditional beta distributions in two different approaches. The two approaches are programmed in the PEGS-Dirichlet software that is freely available at http://statistics.open.ac.uk/elicitation.

As it is the simplest conjugate prior distribution for multinomial models, the Dirichlet distribution is very tractable. However, its lack of flexibility limits its usefulness as a prior distribution. In the next chapter, we discuss the drawbacks of the Dirichlet distribution and propose new elicitation methods that give more flexible prior distributions for multinomial models.

Chapter 7

Eliciting more flexible priors for multinomial models

7.1 Introduction

Being a conjugate prior for the multinomial models, the standard Dirichlet distribution is widely used for its tractability and mathematical simplicity. However, the Dirichlet distribution in its standard form has been criticized as insufficiently flexible to represent prior information about the parameters of multinomial models [e.g. Good (1976), Aitchison (1986), O'Hagan and Forster (2004), Wong (2007)].

The main criticisms of the Dirichlet distribution can be summarized as follows.

1. It has a limited number of parameters. A k-variate Dirichlet distribution is only specified with k parameters. These determine all the k means, k variances and the $k(k-1)/2$ covariances, as given in (6.24)-(6.26).

2. The relative magnitudes of each a_i determine the prior mean, while only the overall magnitude $N = \Sigma a_i$ determines all the variances and covariances if the means are kept fixed.

3. Consequently, the dependence structure between Dirichlet variates cannot be determined independently of its mean values.

4. Dirichlet variates are always negatively correlated, as can be seen from the covariances formulae in (6.26), which may not represent prior belief.

5. Dirichlet variates that have the same mean necessarily have equal variances.

Motivated by these deficiencies, many authors have been interested in constructing new families of distributions for proportions to allow more general dependence structures [e.g. Leonard (1975), Aitchison (1982), Albert and Gubta (1982), Krzysztofowicz and Reese (1993), Rayens and Srinivasan (1994), Tian *et al.* (2010)].

Some of these new distributions are direct generalizations of the standard Dirichlet distribution [e.g. Dickey (1968, 1983), Connor and Mosimann (1969), Grunwald *et al.* (1993), Hankin (2010)]. We select one of them and develop a method of eliciting its hyperparameters as a prior distribution for the multinomial model. The selected generalized Dirichlet distribution shares some of the desirable properties of the standard Dirichlet distribution. It is conjugate, reasonably tractable and can be elicited via the beta elicitation procedure proposed in Chapter 6. The method of eliciting a generalized Dirichlet distribution is given in Section 7.2 and an example illustrating its use is given in Section 7.3. A Gaussian copula function is proposed in Section 7.4 as a flexible multivariate distribution that combines the marginal beta distributions that an expert has assessed.

7.2 Eliciting a generalized Dirichlet prior for a multinomial model

7.2.1 Connor-Mosimann distribution

Connor and Mosimann (1969) introduced a form of the generalized Dirichlet distribution that is also known as Connor-Mosimann distribution. It has a more general covariance structure than the standard Dirichlet distribution and a larger number of parameters, $2(k-1)$.

Its properties have been investigated by Lochner (1975) and Wong (1998), who used it as a prior distribution in a real life engineering application in Wong (2005) and addressed its maximum likelihood estimation in Wong (2010). The density function can be written in the form [Connor and Mosimann (1969)],

$$\pi(p_1, p_2, \cdots, p_k) = \prod_{i=1}^{k-1} \left[\frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)}\, p_i^{a_i - 1} \left(\sum_{j=i}^{k} p_j \right)^{b_{i-1} - (a_i + b_i)} \right] p_k^{b_{k-1} - 1}, \tag{7.1}$$

$$0 \le p_i \le 1, \quad \sum p_i = 1, \quad a_i > 0, \quad b_i > 0, b_0 \text{ is arbitrary.}$$

Or, equivalently, in the form [Lochner (1975)]

$$\pi(p_1, p_2, \cdots, p_{k-1}) = \prod_{i=1}^{k-1} \left[\frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)}\, p_i^{a_i - 1} (1 - p_1 - p_2 - \cdots - p_i)^{\gamma_i} \right], \tag{7.2}$$

$$0 \le p_i \le 1, \quad \sum p_i \le 1, \quad a_i > 0, \quad b_i > 0,$$

where $\gamma_i = b_i - (a_{i+1} + b_{i+1})$, for $i = 1, 2, \cdots, k-2$, and $\gamma_{k-1} = b_{k-1} - 1$.

The standard Dirichlet distribution is a special case of the Connor-Mosimann distribution when $b_i = a_{i+1} + b_{i+1}$, for $i = 1, 2, \cdots, k-2$. Moreover, it is also a conjugate prior to the multinomial distribution. See, for example, Wong (1998).

This generalized Dirichlet distribution can be obtained by transforming $(k-1)$ independent beta variates $Z_1, Z_2, \cdots, Z_{k-1}$, each with parameters a_i and b_i, for $i = 1, 2, \cdots, k-1$, as follows

$$p_j = \begin{cases} Z_1, & \text{for } j = 1, \\ Z_j \prod_{i=1}^{j-1} (1 - Z_i), & \text{for } j = 2, \cdots, k-1. \end{cases} \tag{7.3}$$

The remaining variable p_k can be also given, in terms of $Z_1, Z_2, \cdots Z_k$, as

$$p_k = Z_k \prod_{i=1}^{k-1} (1 - Z_i), \tag{7.4}$$

where, by definition, $Z_k = 1$.

The inverse transformations are given by

$$Z_j = \begin{cases} p_1, & \text{for } j = 1, \\ \dfrac{p_j}{1 - \sum_{i=1}^{j-1} p_i}, & \text{for } j = 2, \cdots, k. \end{cases} \tag{7.5}$$

The first two moments of the generalized Dirichlet variates can be computed, in view of (7.3) and (7.4), as

$$S_j = E(p_j) = \begin{cases} E(Z_1), & \text{for } j = 1, \\ E(Z_j) \prod_{i=1}^{j-1} E(1 - Z_i), & \text{for } j = 2, \cdots, k, \end{cases} \tag{7.6}$$

and

$$T_j = E(p_j^2) = \begin{cases} E(Z_1^2), & \text{for } j = 1, \\ E(Z_j^2) \prod_{i=1}^{j-1} E(1 - Z_i)^2, & \text{for } j = 2, \cdots, k. \end{cases} \tag{7.7}$$

Hence, using well-known formulae for the first two moments of the standard beta distribution, and since $Z_k = 1$, we write

$$S_j = \begin{cases} \dfrac{a_1}{a_1 + b_1}, & \text{for } j = 1, \\ \dfrac{a_j}{a_j + b_j} \displaystyle\prod_{i=1}^{j-1} \dfrac{b_i}{a_i + b_i}, & \text{for } j = 2, \cdots, k-1, \\ \displaystyle\prod_{i=1}^{k-1} \dfrac{b_i}{a_i + b_i}, & \text{for } j = k, \end{cases} \tag{7.8}$$

$$T_j = \begin{cases} \dfrac{a_1(a_1+1)}{(a_1+b_1)(a_1+b_1+1)}, & \text{for } j = 1, \\ \dfrac{a_j(a_j+1)}{(a_j+b_j)(a_j+b_j+1)} \displaystyle\prod_{i=1}^{j-1} \dfrac{b_i(b_i+1)}{(a_i+b_i)(a_i+b_i+1)}, & \text{for } j = 2, \cdots, k-1, \\ \displaystyle\prod_{i=1}^{k-1} \dfrac{b_i(b_i+1)}{(a_i+b_i)(a_i+b_i+1)}, & \text{for } j = k. \end{cases} \tag{7.9}$$

and

$$\operatorname{Var}(p_j) = T_j - S_j^2, \qquad \text{for } j = 1, 2, \cdots, k. \tag{7.10}$$

Regarding covariances, Connor and Mosimann (1969) showed that

$$\operatorname{Cov}(p_1, p_j) = -\frac{E(p_j)}{E(1-p_1)} \operatorname{Var}(p_1), \qquad \text{for } j = 2, \cdots, k, \tag{7.11}$$

$$\begin{aligned} \operatorname{Cov}(p_j, p_{j+1}) = E(Z_{j+1}) E[Z_j(1-Z_j)] \prod_{i=1}^{j-1} E[(1-Z_i)^2] \\ - E(p_j)E(p_{j+1}), \qquad \text{for } j = 2, \cdots, k-1, \end{aligned} \tag{7.12}$$

and

$$\operatorname{Cov}(p_j, p_m) = \frac{E(Z_m)}{E(Z_{j+1})} \left[\prod_{i=j+1}^{m-1} E(1 - Z_i) \right] \operatorname{Cov}(p_j, p_{j+1}), \qquad \text{for } 1 < j < m \leq k. \tag{7.13}$$

Therefore, p_1 is always negatively correlated with all other variates. However, any other two successive variates can be positively correlated, as can be seen from equation (7.12). Moreover, the correlation between any p_j and p_m, for $1 < j < m \leq k$, has the same sign as that of $\operatorname{Cov}(p_j, p_{j+1})$. In this sense, the generalized Dirichlet distribution has a more flexible dependence structure than the standard Dirichlet, which always imposes negative correlations between all pairs of variables, as mentioned before. Similar results were found by Lochner (1975), while Wong (2005) used these properties to select a generalized Dirichlet prior for sorting probabilities of microelectronic chips that tend to be positively correlated.

As in the case of the standard Dirichlet distribution, the conditional distributions of the generalized Dirichlet variates are still scaled beta distributions. This can be shown, using the marginal distributions of the generalized Dirichlet distribution, as follows.

If $\underline{p}_{k-1} = (p_1, p_2, \cdots, p_{k-1})$ has a generalized Dirichlet distribution of the form (7.2), then the marginal distribution of any subset from $\underline{p}_{k-1}$, say $\underline{p}_r = (p_1, p_2, \cdots, p_r)$, $r = 2, 3, \cdots, k-1$, is again a generalized Dirichlet distribution with the corresponding parameters [e.g. Wong (1998)].

The conditional distributions of $p_r|p_1, p_2, \cdots, p_{r-1}$, for $r = 2, 3, \cdots, k-1$, can be computed from (7.2) as follows

$$\pi(p_r|p_1, p_2, \cdots, p_{r-1}) = \frac{\pi(\underline{p}_r; a_1, a_2, \cdots, a_{r-1}, b_1, b_2, \cdots, b_{r-1})}{\pi(\underline{p}_{r-1}; a_1, a_2, \cdots, a_{r-2}, b_1, b_2, \cdots, b_{r-2})} \tag{7.14}$$

$$= \frac{1}{\beta(a_r, b_r)} \frac{p_r^{a_r - 1}}{(1 - \sum_{i=1}^{r-1} p_i)^{a_r}} \left(1 - \frac{p_r}{1 - \sum_{i=1}^{r-1} p_i}\right)^{b_r - 1}, \tag{7.15}$$

which are scaled beta distributions over the intervals $(0, 1 - \sum_{i=1}^{r-1} p_i)$. They are also known as three parameter beta distributions, i.e.

$$(p_r|p_1, p_2, \cdots, p_{r-1}) \sim \text{beta}(a_r, b_r, 1 - \sum_{i=1}^{r-1} p_i), \qquad \text{for } r = 2, 3, \cdots, k-1.$$

As in Section 6.3.4, applying the transformation

$$p_r^* = \begin{cases} p_1, & \text{for } r = 1, \\ \dfrac{p_r}{1 - \sum\limits_{i=1}^{r-1} p_i}, & \text{for } r = 2, 3, \cdots, k-1, \end{cases}$$

gives

$$p_r^* \sim \text{beta}(a_r, b_r) \quad \forall r = 1, 2, \cdots, k-1. \tag{7.16}$$

7.2.2 Assessment tasks

The elicitation process given before in the conditional approach for the standard Dirichlet case in Section 6.3.4 is still valid here. The main difference in the current case is that the generalized Dirichlet hyperparameters $(a_1, a_2, \cdots, a_{k-1}, b_1, b_2, \cdots, b_{k-1})$ are exactly the parameters (a_i, b_i) of the beta distribution of p_r^* in (7.16), for $r = 1, 2, \cdots, k-1$. Hence, the generalized Dirichlet hyperparameters are directly estimated using beta parameters that can be elicited using conditional assessments as in Section 6.3.4. Note that no compromise or averaging is needed here, since the total number of hyperparameters that are elicited is equal to the number of hyperparameters in the generalized Dirichlet distribution, namely, $2(k-1)$. This extended number of parameters does not eliminate the benefits of feedback, but it gives the generalized Dirichlet distribution a more flexible structure than the standard one.

Positive correlations can occur in this generalized case, as discussed before, making it more useful and practical in quantifying expert's opinion. However, Aitchison (1986) criticized the class of generalized Dirichlet distributions as being intractable, particularly with respect to statistical analysis. He also noted that, despite having a more general dependence structure than the standard Dirichlet, the class still retains a strong independence structure.

7.2.3 Marginal quartiles of the generalized Dirichlet distribution

It is always useful to give feedback to the expert based on her elicited hyperparameters. This feedback makes the elicited quantities a better representation of the expert's opinion. For the generalized Dirichlet prior, where the assessed probability quartiles are all conditional except for the first category, it is helpful to inform the expert of the corresponding marginal probability quartiles of each category. She should be given the opportunity to modify them so that they are closer to her opinion, and the elicitation method should change the hyperparameter vector according to these modifications.

Unfortunately, marginal distributions of the generalized Dirichlet are not directly of the beta type. However, we make use of the independent beta random variables given in (7.5) to approximate the distribution of each p_j, $j = 1, 2, \cdots, k$, as a standard beta distribution. Detail is given in the remainder of this section.

An approximate distribution for the product of independent beta variates

Fan (1991) introduced a beta approximation to the product of a finite number of independent beta random variables. His method is described in Johnson *et al.* (1994) and Gupta and Nadarajah (2004), who report favorably on the method based on Fan's comparison of the first ten approximate and exact moments. The method equates the first two moments of the approximate beta distribution to the corresponding product moments of the independent beta random variables.

In what follows, we use the method of Fan (1991) to derive the marginal approximate beta distribution of each p_j, $j = 1, 2, \cdots, k$, from which the marginal quartiles are computed. The method can also be inverted to give a new elicited hyperparameter vector of the generalized Dirichlet distribution, based on the marginal quartiles, if any have been modified by the expert.

For $j = 1, 2, \cdots, k$, using the method of Fan (1991), the distribution of each p_j can be approximated by

$$p_j \sim \text{beta}(\alpha_j, \beta_j), \tag{7.17}$$

where

$$\alpha_j = \frac{S_j(S_j - T_j)}{T_j - S_j^2},$$

and

$$\beta_j = \frac{(1 - S_j)(S_j - T_j)}{T_j - S_j^2},$$

with S_j and T_j as given by equations (7.8) and (7.9), respectively.

Feedback

The three quartiles of the distributions in (7.17) are numerically computed and presented to the expert. She is invited to modify some or all of them as she thinks necessary, in which case the modified quartiles are converted in the same manner as proposed in Section 6.2, to give modified pairs of parameters (α_j^*, β_j^*).

The modified two moments of each p_j, for $j = 1, 2, \cdots, k$, are computed as follows

$$S'_j = \frac{\alpha^*_j}{\alpha^*_j + \beta^*_j}, \tag{7.18}$$

and

$$T'_j = \frac{\alpha^*_j(\alpha^*_j + 1)}{(\alpha^*_j + \beta^*_j)(\alpha^*_j + \beta^*_j + 1)}. \tag{7.19}$$

After obtaining S'_j and T'_j, they are transformed into normalized values S^*_j and T^*_j, respectively, such that $\sum_{j=1}^{k} S^*_j = 1$.

In the manner of (7.8) and (7.9), we can write the two modified moments of each Z_j, denoted by $U_j = E^*(Z_j)$ and $W_j = E^*(Z^2_j)$, for $j = 1, 2, \cdots, k-1$, as

$$U_j = \frac{a^*_j}{a^*_j + b^*_j} = \begin{cases} S^*_1, & \text{for } j = 1, \\ \dfrac{S^*_j}{\prod_{i=1}^{j-1} \dfrac{b^*_i}{a^*_i + b^*_i}}, & \text{for } j = 2, \cdots, k-1, \end{cases}$$

and

$$W_j = \frac{a^*_j(a^*_j + 1)}{(a^*_j + b^*_j)(a^*_j + b^*_j + 1)} = \begin{cases} T^*_1, & \text{for } j = 1, \\ \dfrac{T^*_j}{\prod_{i=1}^{j-1} \dfrac{b^*_i(b^*_i + 1)}{(a^*_i + b^*_i)(a^*_i + b^*_i + 1)}}, & \text{for } j = 2, \cdots, k-1, \end{cases}$$

The above system of equations can be recursively solved for the modified hyperparameters of the generalized Dirichlet distribution, a^*_j and b^*_j, for $j = 1, 2, \cdots, k-1$, to give

$$a^*_j = \frac{U_j(U_j - W_j)}{W_j - U^2_j},$$
$$b^*_j = \frac{(1 - U_j)(U_j - W_j)}{W_j - U^2_j}.$$

These modified hyperparameters of the generalized Dirichlet distribution represent the final output of the method.

7.3 Example: Obesity misclassification

Obesity and being overweight are serious public health problems whose adverse consequences can include diabetes, high blood pressure and cardiovascular disease. Obesity can be measured using the Body Mass Index (BMI) of adults, which is defined as body weight (in kilograms) divided by body height (in meters) squared. Obesity is defined as a BMI of over 30 and overweight is a BMI over 25. Looking at the situation in Europe it is estimated that 50% of adults between 35 and 65 years of age are overweight, of whom 10-25% are obese.

Malta reportedly has one of the highest levels of overweight people in Europe. According to the European Health Interview Survey (EHIS), November 2011, Malta recorded the highest proportion of obese men (24.7%) and women (21.1%) amongst the 19 EU Member States for which data are available. The EHIS reports 36.3% of adults in Malta being overweight and a further 22.3% being obese. Obesity in Malta is indeed a major public health challenge and it is targeted as a priority action in Malta's Strategy for Sustainable Development.

In interview surveys, the heights and weights of participating subjects are not measured. Self-reported values of these variables are normally used instead. However, self-reported values are less precise and have no guarantee of accuracy, specially when they are converted into BMI (Shields *et al.* (2008)). Indeed, the prevalence of overweight and obesity are generally underestimated when calculated from self-reported data as compared with measured data. Adults have been shown to systematically overestimate their height, and underestimate their weight. The extent of weight underreporting increases with increasing measured weight (Shields *et al.* (2008)). As a result, significant misclassification occurs when BMI categories are estimated from self-reported data. Correcting interview data for this misclassification bias is desirable but data to estimate the bias is lacking. Instead, quantifying expert opinion might be used to estimate the bias.

One aspect of the obesity misclassification problem in Malta was formulated in a multinomial model as follows. It relates to Maltese adults (16+) who self-report themselves as having a normal weight ($18.5<\text{BMI}\leq 25$). Their actual clinical BMI classification may fall in one of the following multinomial categories: Underweight ($\text{BMI}\leq 18.5$), Normal ($18.5<\text{BMI}\leq 25$), Overweight ($25<\text{BMI}\leq 30$) or Obese ($\text{BMI}>30$). A health information expert, Dr. Neville Calleja, used our PEGS-Dirichlet elicitation software to quantify his opinion about this model, first giving two separate sets of assessments, each of which determines the parameters of a Dirichlet distribution, so that his opinion could be represented by a Dirichlet prior distribution. The second set of assessments was also used to determine the parameters of a generalized Dirichlet distribution, so that his opinion could be modelled by a more flexible prior distribution. Dr. Calleja has been responsible for all health surveys in Malta for the last 10 years. Currently, he is the director of the Department of Health Information and Research in the Ministry of Health, the Elderly and Community Care, Malta. His department leads the collection, analysis and delivery of health related information in Malta.

To elicit a Dirichlet prior based on unconditional beta marginals, the expert ordered the four categories as Normal, Overweight, Obese, Underweight. His unconditional median assessments for these categories were 0.65, 0.20, 0.10, 0.04, respectively. Then he gave his unconditional lower (upper) quartile assessments as 0.55, 0.15, 0.06, 0.02 (0.70, 0.30, 0.14, 0.07), respectively. See Figure 7.1. The four beta marginals were then reconciled into a Dirichlet distribution using three different ways; direct normalizing and averaging, least-squares optimization, and weighted least-squares. Since the expert's assessed medians nearly sum to one, the three different ways gave sets of reconciled quartiles that were very close to each other. He selected marginal medians and quartiles that were computed by direct normalizing and averaging. The elicited hyperparameters of the Dirichlet prior distribution were obtained as $a_1 = 13.23$, $a_2 = 4.71$, $a_3 = 2.18$, $a_4 = 1.08$, with their sum $N = 21.20$.

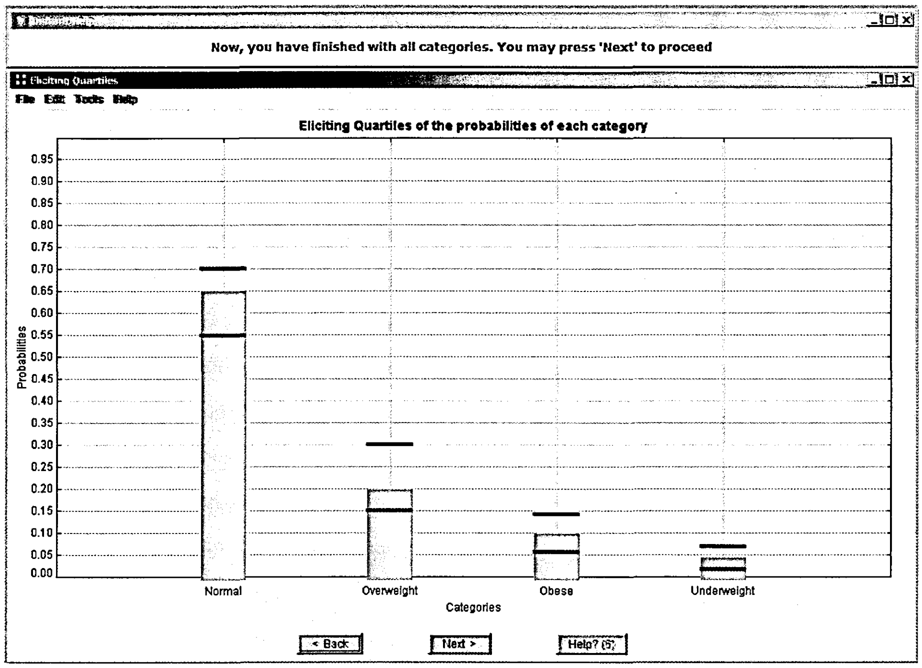


Figure 7.1: Medians and quartiles assessments

Based on conditional beta distributions, the expert quantified his opinion again to elicit another Dirichlet prior for the same problem, but using a different elicitation method. His three quartile assessments of the first category were 0.60, 0.65, 0.72. Then, he was asked to assume that the probability value of the first category is exactly 0.65; given this information he gave his three quartiles for the second category as 0.17, 0.20, 0.25. Finally, conditioning on the probabilities of the first two categories being 0.65, 0.20, he gave the three quartiles of the third category to be 0.07, 0.09, 0.15. The three quartiles of the fourth category were automatically computed and shown to the expert as 0.01, 0.06, 0.08.

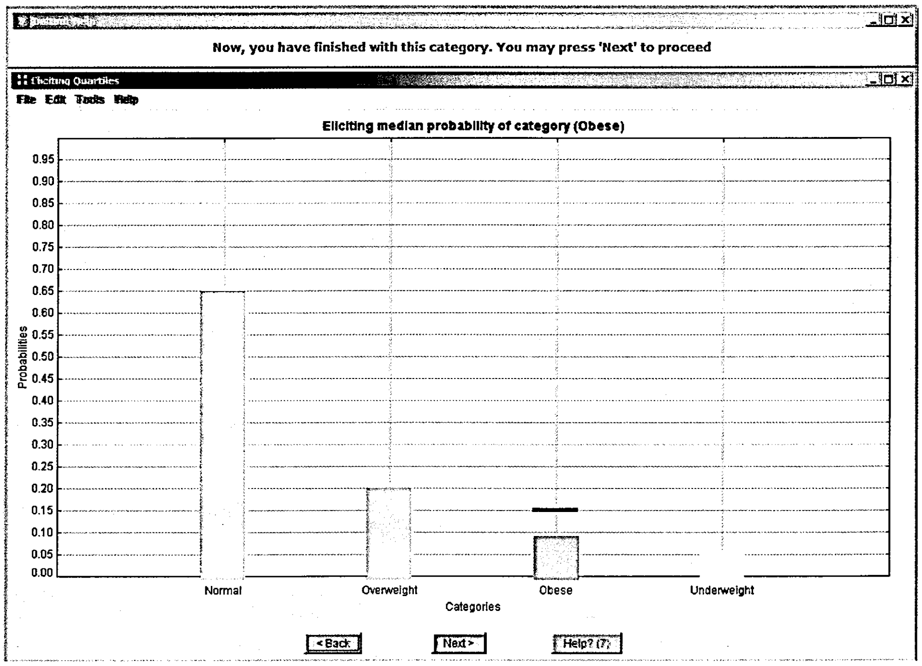


Figure 7.2: Assessing conditional medians

Figure 7.2 is a screen shot after the expert had assessed his median for the third category. The median probability of the third category is in blue (it was assessed), while the fourth median is in yellow (it was calculated from other assessments). Figure 7.3 shows the conditional quartiles that the expert assessed for the third category and the conditional quartiles that were calculated for the fourth category. The elicited hyperparameters of the Dirichlet distribution using this method were $a_1 = 19.91$, $a_2 = 5.00$, $a_3 = 1.11$, $a_4 = 0.65$, with a sum of $N = 26.67$.

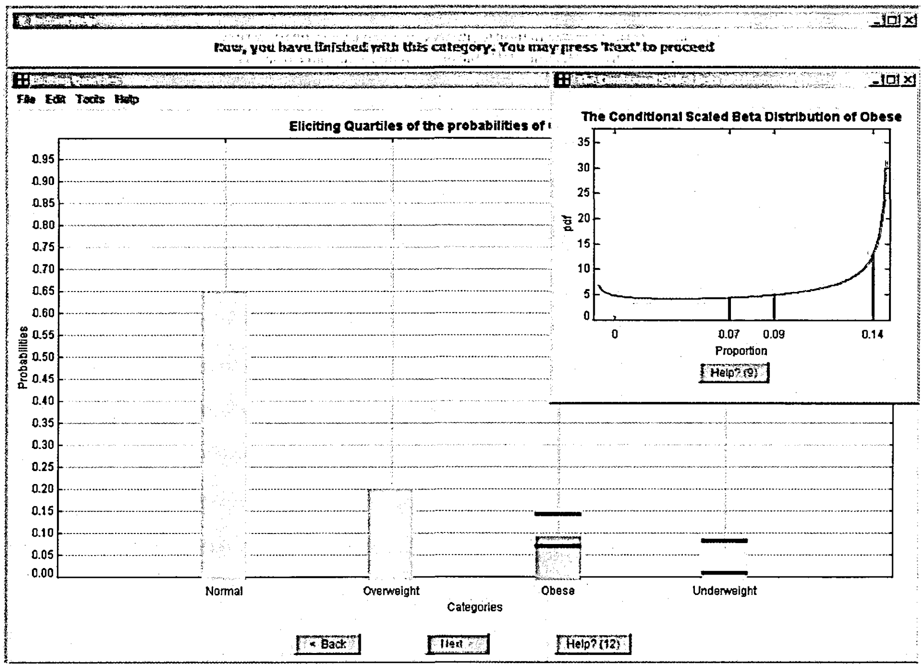


Figure 7.3: Assessing conditional quartiles

On finishing the elicitation process using conditional assessments, the expert was shown a software message offering him the possibility of using the same conditional assessments to elicit a generalized Dirichlet distribution. The expert chose to elicit this more general distribution as well. The following hyperparameters of the generalized Dirichlet prior distribution were elicited, $a_1 = 19.29$, $a_2 = 4.41$, $a_3 = 0.91$, $b_1 = 10.23$, $b_2 = 3.15$, $b_3 = 0.54$.

To compare the three prior distributions elicited in this example, expected values and variances of multinomial probabilities were computed for each distribution as shown in Table 7.1. The means and variances of the Dirichlet distribution were computed using the elicited values of the hyperparameters in formulae (6.24) and (6.25), respectively. The same was done for the generalized Dirichlet using formulae (7.6) to (7.10).

Table 7.1: Probability assessments for different elicited priors

	Marginal assessments			Conditional assessments			Generalized Dirichlet	
	Median	$E(p_i)$	$V(p_i)$	Median	$E(p_i)$	$V(p_i)$	$E(p_i)$	$V(p_i)$
p_1	0.65	0.624	0.012	0.65	0.746	0.007	0.653	0.008
p_2	0.20	0.222	0.008	0.20	0.187	0.006	0.202	0.006
p_3	0.10	0.103	0.004	0.09	0.042	0.001	0.091	0.004
p_4	0.04	0.051	0.002	0.06	0.024	0.001	0.054	0.003

It can be seen from Table 7.1 that the first Dirichlet prior, which was elicited using marginal assessments, and the generalized Dirichlet prior both gave expected values of the multinomial probabilities that are close to the assessed medians. The second Dirichlet prior that was elicited using conditional assessments gave a relatively higher mean value for the first probability than its assessed median, combined with a reduction in the expected values of all other probabilities. This is a little surprising as the generalized Dirichlet utilized the conditional assessments that give the second Dirichlet distribution, yet its hyperparameters are similar to the method that uses marginal assessments. The two elicited values of the hyperparameter N were relatively close to each other, 21.20 and 26.67, in the two elicited standard Dirichlet priors. (There is no single value for N with the generalized Dirichlet.) Moreover, variances of the multinomial probabilities were all small and also close to each other in the three elicited prior distributions.

After eliciting each of the three Dirichlet prior distributions discussed above, the software showed the suggested marginal medians and quartiles of each pair to the expert. He accepted the suggested marginal quartile values, saying that the suggested values were very close to his initial beliefs. Keeping the unit sum constraint in his mind, the expert remarked that assessing conditional medians and quartiles was easier than assessing marginal quartiles. He stated that he could not think about marginal assessments for each category independently of the others. However, he noted at the same time that the elicited generalized Dirichlet distribution may be the most flexible prior of the three.

7.4 Constructing a copula function for the prior distribution

Using the marginal elicitation process given before, we obtain a number of marginal beta distributions. Rather than assume these stem from a Dirichlet distribution, we would like to allow a more flexible dependence structure via their joint distribution, with the aim of better representing the expert's opinion. A flexible tool for this task is given by the copula function, which allows us to choose the marginal distributions independently from the dependence structure between them. The latter structure is given by the copula.

A copula is best described as a multivariate distribution function that is used to bind together marginal distribution functions so as to form a joint distribution. The copula parameterizes the dependence between the marginals, while the parameters of each marginal distribution function can be assessed separately. See for example, Joe (1997), Nelsen (1999) and Kurowicka and Cooke (2006).

There are many types and classes of copula functions, but the most intuitive ones use inverted distribution functions as arguments in known multivariate distributions [Nelsen (1999)]. The general inversion form of a copula function C is given by

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

where G_i are the known marginal distribution functions, $F_{(1,\cdots,k)}$ and F_i $(i = 1, \cdots, k)$ are the assumed joint and marginal distribution functions, respectively. The copula function C works as the cdf of the multivariate distribution that "couples" the given marginal distributions.

7.4.1 Gaussian copula function

The best-known example of the inversion method is the Gaussian copula [Clemen and Reilly (1999)], which is given by

$$C[G_1(x_1), \cdots, G_k(x_k)] = \Phi_{k,R} \left\{ \Phi^{-1}[G_1(x_1)], \cdots, \Phi^{-1}[G_k(x_k)] \right\}. \tag{7.20}$$

Here $\Phi_{k,R}$ is the cdf of a k-variate normal distribution with zero means, unit variances, and a correlation matrix R that reflects the desired dependence structure. Φ is the marginal standard univariate normal cdf.

Since $\Phi_{k,R}$ and Φ are differentiable, the Gaussian copula density function can be simply obtained by differentiating (7.20) with respect to x_i, $i = 1, 2, \cdots, k$, giving

$$f(x_1, x_2, \cdots, x_k|R) = \frac{g_1(x_1) \times \cdots \times g_k(x_k)}{|R|^{1/2}} \exp\{-\frac{1}{2}\underline{Y}_k'(R^{-1} - I_k)\underline{Y}_k\}. \tag{7.21}$$

where

$$\underline{Y}_k' = (\Phi^{-1}[G_1(x_1)], \quad \Phi^{-1}[G_2(x_2)], \quad \cdots, \quad \Phi^{-1}[G_k(x_k)]),$$

$g_i(.)$ is the density function corresponding to $G_i(.)$, $i = 1, 2, \cdots, k$, and I_k is the identity matrix of order k.

To construct a Gaussian copula function in the case of a multinomial model, we can think of each marginal distribution as a beta distribution whose two hyperparameters have been assessed. Then we can construct a Gaussian copula function for the multivariate distribution of $p_1, p_2, \cdots, p_{k-1}$. According to the unit sum constraint, the remaining variable, $p_k = 1 - \sum_{i=1}^{k-1} p_i$, can be treated as a redundant variable that may be removed from the multivariate distribution to avoid singularity problems. Using the Gaussian copula function, the dependence structure of the multivariate distribution will have high flexibility rather than the limited dependence structure imposed by the Dirichlet distribution.

The Gaussian copula function is indexed by the correlation matrix R, which needs to be elicited effectively and must be a positive-definite matrix. In what follows we introduce a method, inspired by Kadane *et al.* (1980), to elicit the correlation matrix R that is sure to be positive-definite.

Let $G_i(p_i)$ be the cdf of the beta distribution of p_i with hyperparameters α_i and β_i, $i = 1, 2, \cdots, k-1$, and assume that the joint density of $p_1, p_2, \cdots, p_{k-1}$ is given by a Gaussian copula density, such that

$$f(p_1, p_2, \cdots, p_{k-1}|R) = \frac{g_1(p_1) \times \cdots \times g_{k-1}(p_{k-1})}{|R|^{1/2}} \exp\{-\frac{1}{2}\underline{Y}_{k-1}'(R^{-1} - I_{k-1})\underline{Y}_{k-1}\}. \tag{7.22}$$

where

$$\underline{Y}'_{k-1} = (\Phi^{-1}[G_1(p_1)], \quad \Phi^{-1}[G_2(p_2)], \quad \cdots, \quad \Phi^{-1}[G_{k-1}(p_{k-1})]),$$

and $g_i(.)$ is the beta density of p_i, $i = 1, 2, \cdots, k-1$.

Note that the marginal distributions of this joint density are still the desired beta marginals. Since the hyperparameters of each beta distribution of p_i, $i = 1, 2, \cdots, k-1$, have already been elicited, the prior distribution is totally known except for the matrix R. Although the above density is not multivariate normal for $p_1, p_2, \cdots, p_{k-1}$ and the matrix R is not their correlation matrix, we can still use the multivariate normal properties to elicit a positive-definite matrix R by considering the following normalizing transformations,

$$Y_i = \Phi^{-1}[G_i(p_i)], \quad i = 1, 2, \cdots, k. \tag{7.23}$$

We should stress that with this copula function, the marginal distributions of the p_i are beta distributions that can be fixed independently of R. Thus the ability to specify R gives added flexibility. The aim is to choose R so as to model the expert's opinion about the dependence between the p_i.

According to the main assumption of the Gaussian copula construction, and from (7.23), the vector $\underline{Y}'_{k-1} = (Y_1, \quad Y_2, \quad \cdots, \quad Y_{k-1})$ has a multivariate normal distribution with zero means, unit variances and a correlation matrix R, i.e.

$$\underline{Y}_{k-1} \sim \text{MVN}(\underline{0}, R).$$

Following this assumption, together with the unit sum constraint of the elements of $\underline{p}$, the full vector $\underline{Y}' = (Y_1, \quad Y_2, \quad \cdots, \quad Y_k)$ has what is known as a singular multivariate normal distribution, which will be discussed in more detail in the next chapter. However, we will be interested, during the rest of this chapter, in eliciting a non-singular correlation matrix R for the Gaussian copula function only for $p_1, p_2, \cdots, p_{k-1}$.

Keeping in mind that the Pearson correlation coefficients, as elements of R, are not transformation respecting, i.e. they are not invariant even under strictly monotone increasing transformations as in (7.23). We do not attempt to elicit any correlations between the elements of $\underline{p}$. Even if a correlation matrix for $\underline{p}$ has been elicited it may be of no use in estimating R as no explicit relationship between the two matrices is available. Moreover, the density function in (7.22) is indexed by R, the correlation matrix of $\underline{Y}_{k-1}$, not the correlation matrix of $\underline{p}$.

An alternate method of estimating R that has been proposed in the literature was reviewed in Chapter 2. In that approach, a transformation that respects non-parametric measure of correlation, such as Kendall's τ or Spearman's ρ, is computed for $\underline{p}$. The monotonicity of a transformation like (7.23) is then used to impose the same correlations on $\underline{Y}_{k-1}$. Pearson's correlations are calculated using approximate relations between different correlation coefficients for the normal distribution. For more details see, for example, Clemen and Reilly (1999), Palomo *et al.* (2007) or Daneshkhah and Oakley (2010).

In our proposed approach, the matrix R is elicited as a covariance or correlation matrix of a multivariate normal random vector $\underline{Y}_{k-1}$. However, we still utilize the monotone increasing property of the transformations in (7.23). We may assess conditional quartiles of $\underline{p}$, then transform them into those of $\underline{Y}$ using (7.23). Correlation coefficients between the elements of $\underline{Y}_{k-1}$ can then be estimated using their conditional quartiles and utilizing the properties of the multivariate normal distribution. This is described in Sections 7.4.2 and 7.4.3.

Although the elicitation method of Kadane *et al.* (1980) has been designed to elicit the 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 eliciting positive-definite matrices [Garthwaite *et al.* (2005)]. The method is modified here to elicit the correlation matrix R of the Gaussian copula function.

7.4.2 Assessment tasks

Since the transformations in (7.23) are strictly monotonic increasing from $\underline{p}$ to $\underline{Y}$, we can establish a one-to one correspondence between medians and quartiles of these two vectors.

The required assessments are as follows.

Assessing initial medians and quartiles

1. To elicit each marginal beta distribution, the expert has already assessed a lower quartile, a median and an upper quartile for p_i, $i = 1, 2, \cdots, k$, say $L^*_{i,0}$, $m^*_{i,0}$ and $U^*_{i,0}$, respectively. The method proposed in Section 6.2 can be used to determine the two parameters α_i and β_i of each marginal beta distribution, for $i = 1, 2, \cdots, k$.

2. To help the expert assess the medians and quartile in (1), the PEGS-Copula software presents an interactive graph showing the pdf curve of the beta distribution of p_r, for $r = 1, 2, \cdots, k$. The expert is able to change her assessed quartiles of p_r until its pdf curve represents her opinion to her satisfaction, see Figure 6.1.

3. To attain the unit sum constraint, the mean values of the elicited beta marginals must sum to one. The elicited parameters α_i and β_i are thus modified to fulfill this condition, as follows.

   The mean values μ_i are computed as

   $$\mu_i = \frac{\alpha_i}{\alpha_i + \beta_i}, \qquad \text{for } i = 1, 2, \cdots, k.$$

   The normalized mean values μ_i^* are given by

   $$\mu_i^* = \frac{\mu_i}{\sum_{j=1}^{k} \mu_j}, \quad i = 1, 2, \cdots, k. \tag{7.24}$$

   We keep the variances fixed as

   $$\sigma_i^2 = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2 (\alpha_i + \beta_i + 1)}, \qquad \text{for } i = 1, 2, \cdots, k. \tag{7.25}$$

   Equations (7.24) and (7.25) give the modified set of parameters α_i^* and β_i^*, for $i = 1, 2, \cdots, k$:

   $$\alpha_i^* = \mu_i^* \left[\frac{\mu_i^*(1 - \mu_i^*)}{\sigma_i^2} - 1 \right],$$

   $$\beta_i^* = (1 - \mu_i^*) \left[\frac{\mu_i^*(1 - \mu_i^*)}{\sigma_i^2} - 1 \right].$$

4. Before going further, the modified parameters of each marginal beta distribution are used to compute the corresponding quartiles numerically. These quartiles are presented as feedback to the expert, who is still able to change some or all of them, in which case the process is repeated again until the modified sets of quartiles are accepted by the expert.

Assessing conditional quartiles

5. To estimate the correlation matrix R, the expert is asked to assume that $p_1 = m^*_{1,0}$ and gives a lower quartile L^*_2 and an upper quartile U^*_2 for p_2. For each remaining p_j, $j = 3, \cdots, k-1$, she assesses the two quartiles L^*_j and U^*_j given that $p_1 = m^*_{1,0}$, $p_2 = m^*_{2,0}$, ..., $p_{j-1} = m^*_{j-1,0}$. Figure 7.4 shows the process of assessing conditional quartiles, where the expert has already assessed the lower quartile of the third category, conditional on the median values of the first two categories, which are shown by the red bars.

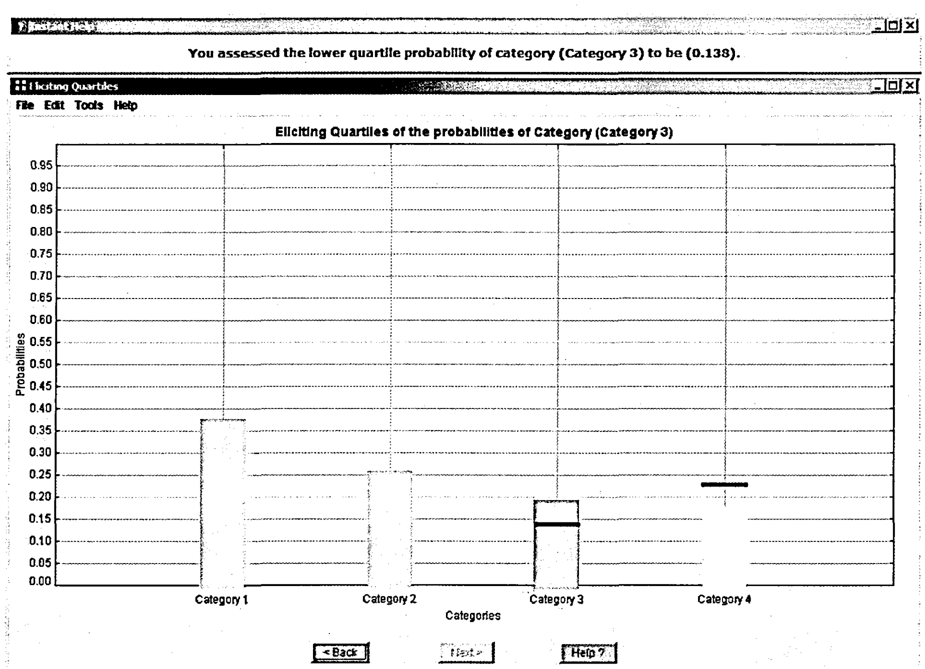


Figure 7.4: Assessing conditional quartiles for copula elicitation

6. The lower (upper) quartile L_k^* (U_k^*) of p_k will be 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}^*. See Figure 7.4, where the software has shown the upper quartile of the fourth category after the expert assessed the lower quartile of the third category. In fact, L_k^* (U_k^*) is the lower (upper) quartile of $(p_k|p_1 = m_{1,0}^*, \cdots, p_{k-2} = m_{k-2,0}^*)$ instead of $(p_k|p_1 = m_{1,0}^*, \cdots, p_{k-1} = m_{k-1,0}^*)$, as the two quartiles in the latter case should be just equal to $m_{k,0}^*$, because of the unit sum constraint.

Assessing conditional medians

7. Here we assume that the median of p_1 has been changed from $m_{1,0}^*$ into $m_{1,1}^* = m_{1,0}^* + \eta_1^*$. Given this information, the expert will be asked to change her previous medians $m_{j,0}^*$ of each p_j to be $m_{j,1}^*$. We put

$$m_{j,1}^* = m_{j,0}^* + \theta_{j,1}^*, \quad \text{for } j = 2, \cdots, k. \tag{7.26}$$

8. In each successive step i, for $i = 2, 3, \cdots, k-2$, the expert will be asked to suppose that the median values of p_1, p_2, ..., p_i are $m_{1,1}^* = m_{1,0}^* + \eta_1^*, m_{2,2}^* = m_{2,1}^* + \eta_2^*, \cdots, m_{i,i}^* = m_{i,i-1}^* + \eta_i^*$, respectively. Given this information, she will be asked to update her assessed medians from the most recent previous step $m_{i+1,i-1}^*$, $m_{i+2,i-1}^*, \cdots, m_{k,i-1}^*$. The updated assessments are $m_{i+1,i}^* = m_{i+1,i-1}^* + \theta_{i+1,i}^*$, $m_{i+2,i}^* = m_{i+2,i-1}^* + \theta_{i+2,i}^*$, $\cdots$, $m_{k,i}^* = m_{k,i-1}^* + \theta_{k,i}^*$, respectively. In other words, for $i = 1, 2, \cdots, k-2$, $j = i+1, i+2, \cdots, k$, we can write

$$m_{j,i}^* = m_{j,i-1}^* + \theta_{j,i}^* \text{ is the median of } (p_j|p_1 = m_{1,1}^*, \cdots, p_i = m_{i,i}^*). \tag{7.27}$$

On an interactive graph produced by the PEGS-Copula software, see Figure 7.5, the conditioning set of median values are shown as red bars. The conditional medians of the remaining categories at the most recent previous step are shown as black lines. The expert is asked to assess how her new median values will change based on the new conditioning set.

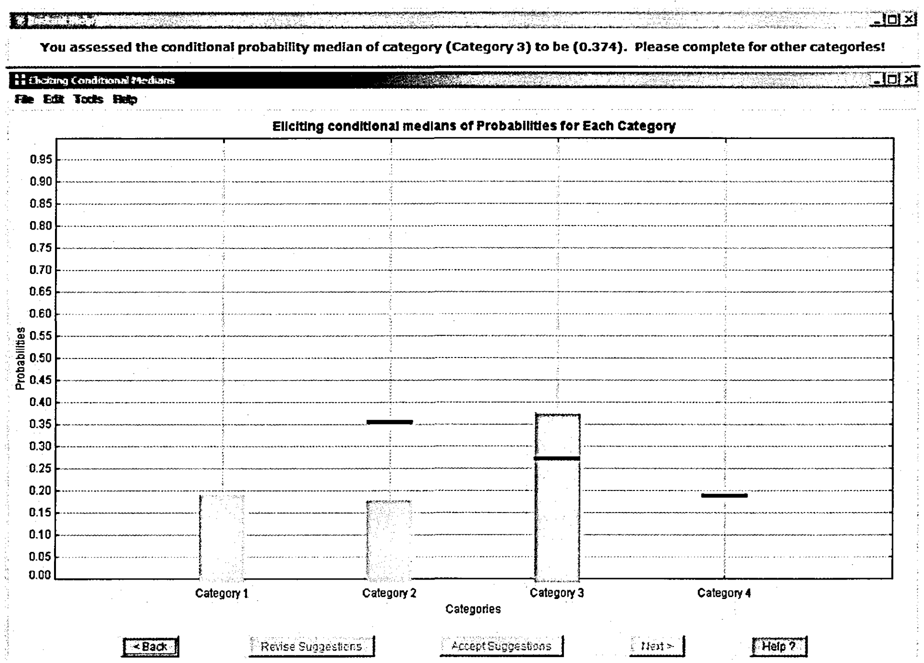


Figure 7.5: Assessing conditional medians for copula elicitation

9. For mathematical coherence, as will be proved in Lemma 7.1, we require

$$\sum_{j=1}^{i} m^*_{j,j} + \sum_{j=i+1}^{k} m^*_{j,i} = 1, \qquad i = 1, 2, \cdots, k-2.$$

The expert has the option of changing her initial set of assessments $m'_{i+1,i}$, $m'_{i+2,i}$, $\cdots$, $m'_{k,i}$ until she feels that the suggested normalized set $m^*_{i+1,i}$, $m^*_{i+2,i}$, $\cdots$, $m^*_{k,i}$ gives an adequate representation of her opinion. The software suggests each normalized conditional median $m^*_{j,i}$, given by yellow bars in Figure 7.6, as

$$m^*_{j,i} = \left[\frac{1 - \sum_{r=1}^{i} m^*_{r,r}}{\sum_{r=i+1}^{k} m'_{r,i}} \right] m'_{j,i}, \qquad \text{for } i = 1, \cdots, k-2, \quad j = i+1, \cdots, k.$$

10. 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 $p_1, p_2, \cdots, p_{k-1}$ gives a

fixed value for p_k. In this case no upper or lower quartiles can be assessed for p_k, as mentioned before.

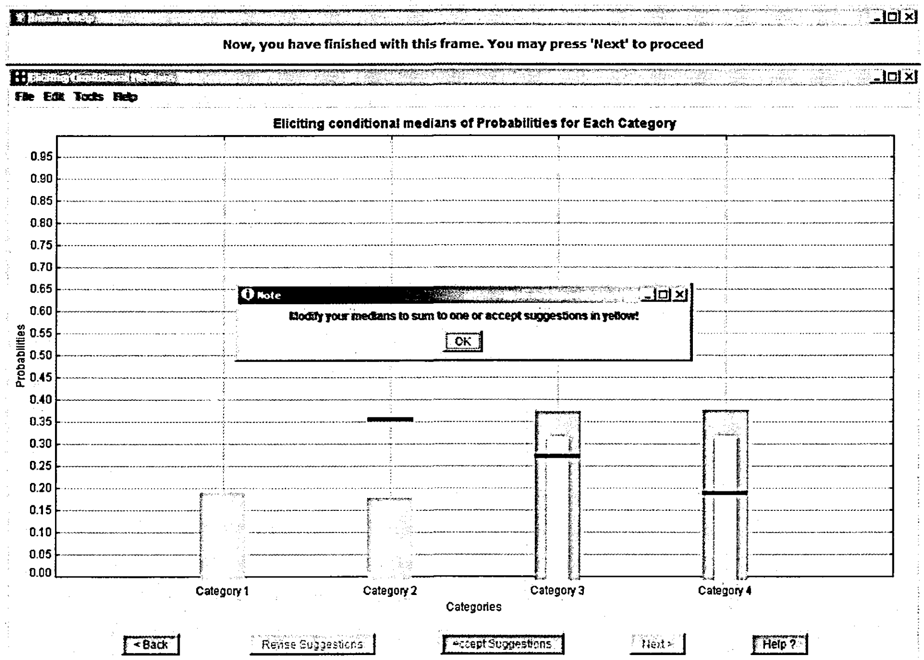


Figure 7.6: Software suggestions for conditional medians

7.4.3 Eliciting a positive-definite correlation matrix R

The normalizing one-to-one functions in (7.23) are used to transform the assessed conditional quartiles of $\underline{p}$ into conditional quartiles of $\underline{Y}$, and hence, into conditional expectations, variances and covariances of the multivariate normal variables. In particular, letting $M(X)$ denote the median function of the random variable X, we proceed as follows.

For $i = 1, 2, \cdots, k$, let $m_{i,0} = \Phi^{-1}[G_i(m^*_{i,0})]$.

For $i = 1, 2, \cdots, k-2$, and $j = i+1, \cdots, k-1$, let

$$m_{j,i} = E(Y_j | p_1 = m^*_{1,0} + \eta^*_1, p_2 = m^*_{2,1} + \eta^*_2, \cdots, p_i = m^*_{i,i-1} + \eta^*_i).$$

Then

$$m_{j,i} = \Phi^{-1}[G_j(m^*_{j,i})]. \tag{7.28}$$

For $i = 1, 2, \cdots, k-2$ define η_i by letting $\eta_i = Y_i - m_{i,i-1}$ when $p_i = m^*_{i,i-1} + \eta^*_i$. Then

$$m_{j,i} = E(Y_j | Y_1 = m_{1,0} + \eta_1, Y_2 = m_{2,1} + \eta_2, \cdots, Y_i = m_{i,i-1} + \eta_i), \tag{7.29}$$

and

$$\eta_i = \Phi^{-1}[G_i(m^*_{i,i-1} + \eta^*_i)] - \Phi^{-1}[G_i(m^*_{i,i-1})], \qquad \text{for } i = 1, 2, \cdots, k-2.$$

Analogous to $m^*_{i,i} = m^*_{i,i-1} + \eta^*_i$, define

$$m_{i,i} = m_{i,i-1} + \eta_i, \qquad \text{for } i = 1, 2, \cdots, k-2, \tag{7.30}$$

so that $Y_i = m_{i,i}$ when $p_i = m^*_{i,i}$.

For $i = 1, 2, \cdots, k-2$, and $j = i+1, \cdots, k-1$, analogous to $\theta^*_{j,i} = m^*_{j,i} - m^*_{j,i-1}$, define

$$\theta_{j,i} = m_{j,i} - m_{j,i-1},$$

so that

$$\theta_{j,i} = \Phi^{-1}[G_j(m^*_{j,i-1} + \theta^*_{j,i})] - \Phi^{-1}[G_j(m^*_{j,i-1})].$$

For $i = 1, 2, \cdots, k-2$, and $j = i+1, \cdots, k-1$, let

$$V_{j,i} = \mathrm{Var}(Y_j | Y_1 = m_{1,0}, Y_2 = m_{2,0}, \cdots, Y_i = m_{i,0}),$$

so that

$$V_{j,j-1} = \left[\frac{U_j - L_j}{1.349}\right]^2, \quad \text{for } j = 2, 3, \cdots, k-1, \tag{7.31}$$

with

$$U_j = \Phi^{-1}[G_j(U^*_j)], \quad L_j = \Phi^{-1}[G_j(L^*_j)].$$

Having defined the above quantities, we are ready now to state and prove the following lemma.

Lemma 7.1. *Under the unit sum constraint of $\underline{p}$, and the multivariate normality of $\underline{Y}$,*

$$\sum_{j=1}^{i} m^*_{j,j} + \sum_{j=i+1}^{k} m^*_{j,i} = 1, \qquad i = 1, 2, \cdots, k-2.$$

Proof

A property of conditional expectations of singular multivariate normal distributions is given by equation (8a.2.11) in (Rao, 2002, p 522). Using this property, for $i = 1, 2, \cdots, k-2$, we have

$$E[Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i}] = E[Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i},$$

$$Y_{i+1} = E(Y_{i+1}|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i}),$$

$$\cdots$$

$$Y_{k-1} = E(Y_{k-1}|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i})],$$

then, from equations (7.29) and (7.30)

$$M(Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i}) = M(Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i},$$

$$Y_{i+1} = m_{i+1,i}, \cdots, Y_{k-1} = m_{k-1,i}).$$

Hence

$$M\left\{\Phi^{-1}[G_k(p_k)]|p_1 = m^*_{1,1}, \cdots, p_i = m^*_{i,i}\right\} =$$

$$M\left\{\Phi^{-1}[G_k(p_k)]|p_1 = m^*_{1,1}, \cdots, p_i = m^*_{i,i}, p_{i+1} = m^*_{i+1,i}, \cdots, p_{k-1} = m^*_{k-1,i}\right\},$$

which, utilizing equations (7.26) and (7.27), gives

$$\Phi^{-1}[G_k(m^*_{k,i})] = \Phi^{-1}\{G_k[M(p_k|p_1 = m^*_{1,1}, \cdots, p_i = m^*_{i,i},$$

$$p_{i+1} = m^*_{i+1,i}, \cdots, p_{k-1} = m^*_{k-1,i})]\},$$

i.e.

$$m^*_{k,i} = M(p_k|p_1 = m^*_{1,1}, \cdots, p_i = m^*_{i,i}, p_{i+1} = m^*_{i+1,i}, \cdots, p_{k-1} = m^*_{k-1,i}).$$

Since the condition in the RHS of the above equation is on all the p_is except p_k, applying the unit sum constraint gives the conditional median in the form of the following complement

$$m^*_{k,i} = 1 - \sum_{j=1}^{i} m^*_{j,j} - \sum_{j=i+1}^{k-1} m^*_{j,i},$$

which ends the proof of Lemma 7.1.

- To elicit a positive-definite correlation matrix R, let

$$\underline{Y}_i = (Y_1, \ \ Y_2, \ \ \cdots, \ \ Y_i), \qquad i = 1, 2, \cdots, k-1,$$

and

$$R_i = \text{Var}(\underline{Y}_i), \qquad i = 1, 2, \cdots, k-1,$$

where $R_1 = \text{Var}(Y_1) = 1$ and the final matrix $R = R_{k-1}$.

- Suppose that R_{i-1} has been estimated as a positive-definite matrix, we aim now to elicit R_i, and show it is positive-definite. R_i can be partitioned as follows

$$R_i = \begin{bmatrix} R_{i-1} & R_{i-1}\underline{r}_i \\ \underline{r}_i' R_{i-1} & V_i \end{bmatrix}, \tag{7.32}$$

where

$$R_{i-1}\underline{r}_i = \text{Cov}(\underline{Y}_{i-1}, Y_i),$$

$$V_i = \text{Var}(Y_i).$$

Although the Gaussian copula function implies that $\text{Var}(Y_i) = 1$, we will find another estimate for V_i using the conditional variance of Y_i elicited in (7.31). The reason for this, as will be shown later, is to follow the approach of Kadane *et al.* (1980) so as to ensure the positive-definiteness of the matrix R_i. In what follows, we use the conditional median assessments to estimate $\underline{r}_i$.

- Using the partition (7.32), it is well-known from multivariate normal distribution theory, since $E(\underline{Y}) = \underline{0}$, that

$$E(Y_i|\underline{Y}_{i-1}) = \underline{Y}_{i-1}' R_{i-1}^{-1} R_{i-1}\underline{r}_i = \underline{Y}_{i-1}'\underline{r}_i. \tag{7.33}$$

Moreover, for $j \leq i-1$, taking the conditional expectation of both sides of (7.33), given that $\underline{y}_j = (m_{1,0} + \eta_1, m_{2,1} + \eta_2, \cdots, m_{j,j-1} + \eta_j)$, gives

$$E\left[E(Y_i|\underline{Y}_{i-1})|\underline{Y}_j = \underline{y}_j\right] = E(\underline{Y}_{i-1}'|\underline{Y}_j = \underline{y}_j)\, \underline{r}_i, \tag{7.34}$$

i.e

$$E(Y_i|\underline{Y}_j = \underline{y}_j) = \left(y_1, \ \cdots, \ y_j, \ E(Y_{j+1}|\underline{Y}_j), \ \cdots, \ E(Y_{i-1}|\underline{Y}_j)\right) \underline{r}_i. \tag{7.35}$$

From (7.29) and (7.35) we get

$$m_{i,j} = \left(m_{1,0} + \eta_1, \quad m_{2,1} + \eta_2, \quad \cdots, \quad m_{j,j-1} + \eta_j, \quad m_{j+1,j}, \quad \cdots, \quad m_{i-1,j}\right) \underline{r}_i.$$

Since $j = 1, 2, \cdots, i-1$, we end up with a system of $i-1$ equations of the form

$$T_i = Q_{i-1}\underline{r}_i, \tag{7.36}$$

where

$$T_i = \begin{bmatrix} m_{i,1} \\ m_{i,2} \\ \vdots \\ m_{i,i-1} \end{bmatrix},$$

and

$$Q_{i-1} = \begin{bmatrix} \eta_1 & m_{2,1} & m_{3,1} & \cdots & m_{i-1,1} \\ \eta_1 & m_{2,1} + \eta_2 & m_{3,2} & \cdots & m_{i-1,2} \\ \eta_1 & m_{2,1} + \eta_2 & m_{3,2} + \eta_3 & \cdots & m_{i-1,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta_1 & m_{2,1} + \eta_2 & m_{3,2} + \eta_3 & \cdots & m_{i-1,i-2} + \eta_{i-1} \end{bmatrix}.$$

Since $m_{i,j} - m_{i,j-1} = \theta_{i,j}$, $j = 1, 2, \cdots, i-1$, and $m_{i,0} = 0$, multiplying both sides of (7.36) from the left by the matrix

$$M_{i-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix},$$

the system can be written as

$$\begin{bmatrix} \theta_{i,1} \\ \theta_{i,2} \\ \vdots \\ \theta_{i,i-1} \end{bmatrix} = \begin{bmatrix} \eta_1 & \theta_{2,1} & \cdots & \theta_{i-1,1} \\ 0 & \eta_2 & \cdots & \theta_{i-1,2} \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \eta_{i-1} \end{bmatrix} \underline{r}_i.$$

Provided that

$$\eta_j \neq 0, \quad j = 1, 2, \cdots, i-1,$$

the upper diagonal matrix $M_{i-1}Q_{i-1}$ is non-singular. Hence

$$\underline{r}_i = \begin{bmatrix} \eta_1 & \theta_{2,1} & \cdots & \theta_{i-1,1} \\ 0 & \eta_2 & \cdots & \theta_{i-1,2} \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \eta_{i-1} \end{bmatrix}^{-1} \begin{bmatrix} \theta_{i,1} \\ \theta_{i,2} \\ \vdots \\ \theta_{i,i-1} \end{bmatrix}.$$

- Since

$$\mathrm{Var}(Y_i|\underline{Y}_{i-1}) = \mathrm{Var}(Y_i) - \underline{r}_i' R_{i-1} \underline{r}_i,$$

we can now use the assessed conditional variance given by $V_{i,i-1}$ in (7.31) to estimate the unconditional variance V_i as follows

$$V_i = V_{i,i-1} + \underline{r}_i' R_{i-1} \underline{r}_i.$$

Using the Schurr complement, the matrix R_i is positive-definite if and only if

$$V_i - \underline{r}_i' R_{i-1} \underline{r}_i > 0,$$

which is guaranteed from (7.31) since $V_{i,i-1} > 0$.

- Choosing the arbitrary values $\eta_j^* \neq 0$, $j = 1, 2, \cdots, i-1$, guarantees the existence of a unique solution for $\underline{r}_i$. It can be seen from the relation

$$\eta_j = \Phi^{-1}[G_j(m_{j,j-1}^* + \eta_j^*)] - \Phi^{-1}[G_j(m_{j,j-1}^*)],$$

that $\eta_j \neq 0$ as $\eta_j^* \neq 0$, $j = 1, 2, \cdots, i-1$.

- With the proposed method, R_i is a positive-definite matrix if R_{i-1} is positive-definite ($i = 2, 3, \cdots, k-1$). Since $R_1 = 1 > 0$, by mathematical induction, the full correlation matrix $R = R_{k-1}$ is guaranteed to be positive-definite.

- We have to note that, according to this method of elicitation, the variances on the main diagonal of R, say $r_{i,i}$, $i = 1, 2, \cdots, k-1$, will seldom equal one, except for the first element $r_{1,1}$. It is easy, however, to transform R into R^*, where R^* is a suitable correlation matrix for the Gaussian copula function, satisfying both the unit variances and positive-definiteness. R^* can be obtained from R using the transformation

$$R^* = ARA.$$

where

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \dfrac{1}{\sqrt{r_{2,2}}} & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \dfrac{1}{\sqrt{r_{k-1,k-1}}} \end{bmatrix}.$$

The unit variances in the correlation matrix R^* ensures that each marginal distribution $G_i(p_i)$ is still a beta distribution with the same marginal hyperparameters α_i and β_i that were elicited before ($i = 1, 2, \cdots, k$).

- The accompanying software outputs the elicited pairs of beta parameters α_i and β_i, for $i = 1, 2, \cdots, k$, together with the elicited covariance matrix, R^*.

7.5 Example: Waste collection

The Environmental Agency in the UK is currently interested in the fuel consumption of waste collection vehicles. It is thought that substantial quantities of fuel are used to collect recyclable waste and that local authorities are insufficiently aware of the amounts involved. In this example, a waste management expert, (Dr. Stephen Burnley, The Open University) used the PEGS-Copula elicitation software to quantify his opinion about the proportions of waste collection trips according to the type of recyclable waste. Dr. Burnley is a fellow of the Chartered Institution of Waste Management. He advised that two main types of the waste are considered; urban recycle and rural recycle. Each of them may contain bins, sacks, garden waste and recycle waste. Hence, each collection trip is arranged by the local authority for only one of eight different waste types. Considering the proportions of collection trips for waste in each category, the problem can be formulated in a multinomial model with eight categories. Our method and software were used to quantify the expert's opinion about a Gaussian copula prior for the parameters of this multinomial model.

After initializing the software and defining the model, the expert assessed his medians of the proportion of collection trips for each of the following 8 types of waste: urban-bins/ urban-sacks/ urban-garden/ rural-bins/ rural-sacks/ rural-garden/ urban-recycle/ rural-recycle. Then the expert assessed lower and upper quartiles for the proportion of each category. His assessed medians and quartiles are shown as blue bars and short dark blue horizontal lines, respectively, in Figure 7.7. These assessments are also given in Table 7.2 below.

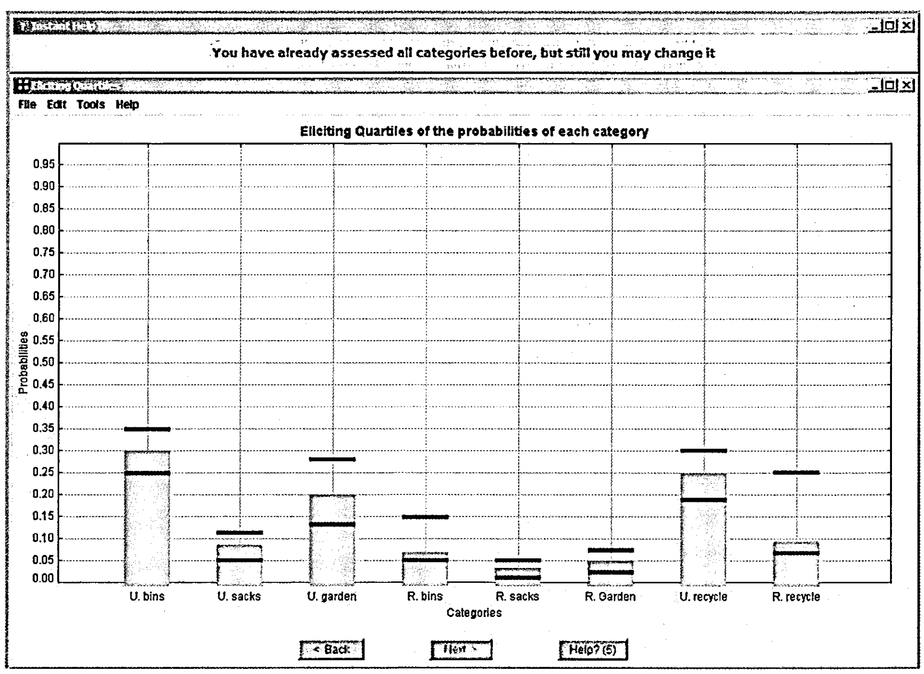


Figure 7.7: The initially assessed marginal medians and quartiles

Table 7.2: Expert's assessments of medians and quartiles

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
Lower quartile	0.25	0.05	0.13	0.05	0.01	0.02	0.18	0.07
Median	0.30	0.08	0.20	0.07	0.03	0.05	0.25	0.09
Upper quartile	0.35	0.12	0.28	0.15	0.05	0.07	0.30	0.25

These assessments were used to elicit a marginal beta prior distribution for the proportion of trips in each category. For mathematical coherence, the expected values of these elicited beta priors must sum to 1, so, the software used the initial assessments to elicit beta distributions that satisfy this condition. The median values and quartiles of the coherent beta distributions were computed and presented to the expert as feedback in Figure 7.8. During this feedback stage he was invited to accept or revise these quantities. The initial median values given by the expert have a sum that is nearly equal to one, so the coherent medians and quartiles suggested by the software in Figure 7.8 were close to his assessments and he naturally accepted them as representatives of his opinions.

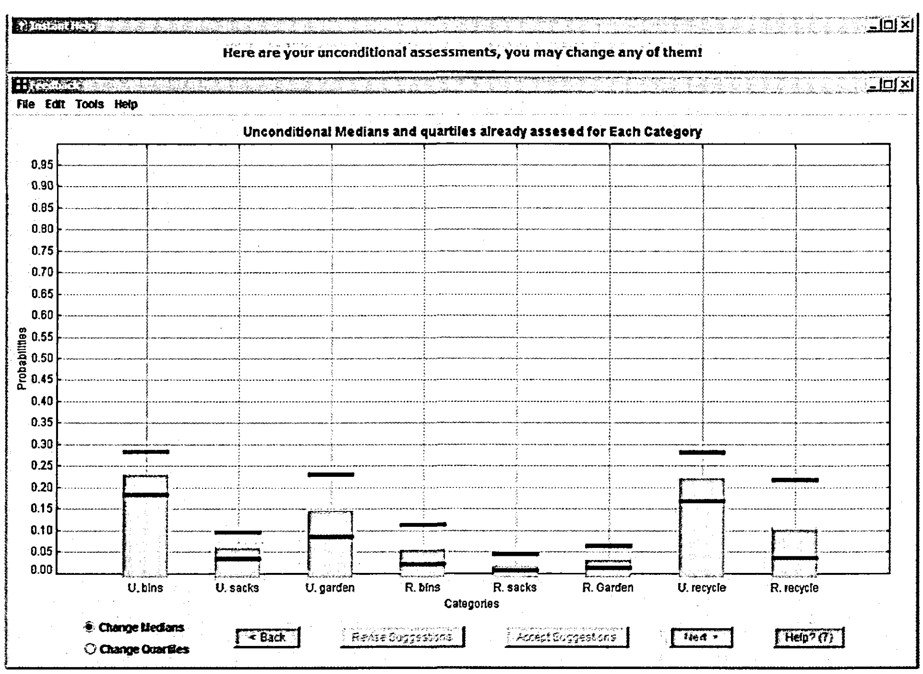


Figure 7.8: The coherent assessments suggested by the software

To elicit 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. To do that, he assessed conditional quartile values, under the condition that the assessed medians for the previous categories were actually the true values. For example, he assessed his conditional quartiles of the proportion for the fourth category, given that the median values for the first three categories equalled their true values. This is illustrated in Figure 7.9.

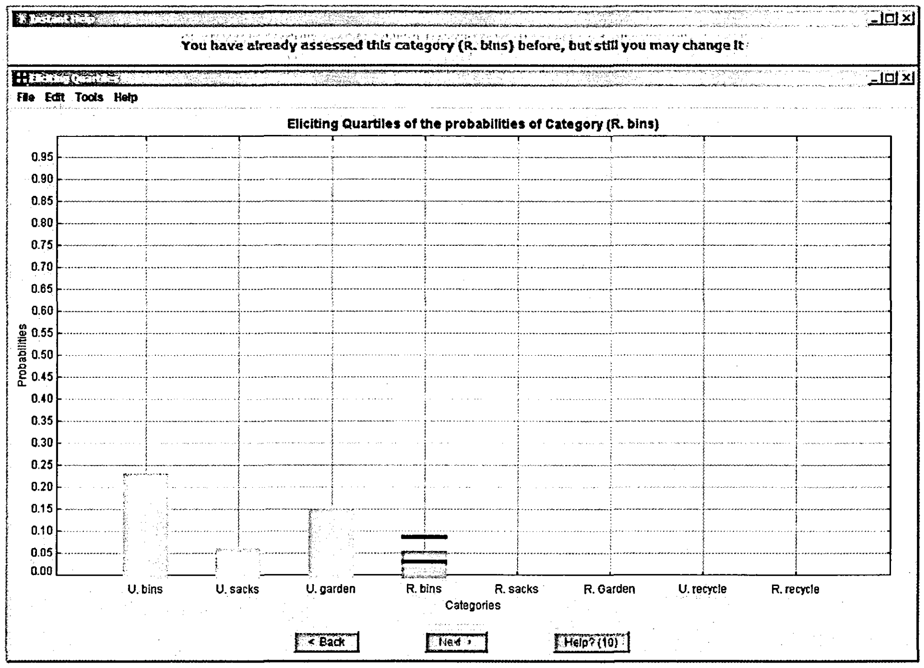


Figure 7.9: Assessing conditional quartiles

The expert's seven pairs of assessments for the lower and upper conditional quartiles are given in Table 7.3. The quartiles for the last category are shown in bold typeface in Table 7.3 as they were automatically computed by the software when the expert assessed two quartiles for the seventh category. This is also illustrated in Figure 7.10.

Table 7.3: Expert's assessments of conditional quartiles

p_2	p_3	p_4	p_5	p_6	p_7	p_8
0.03	0.10	0.03	0.01	0.02	0.20	**0.19**
0.13	0.23	0.08	0.04	0.08	0.28	**0.27**

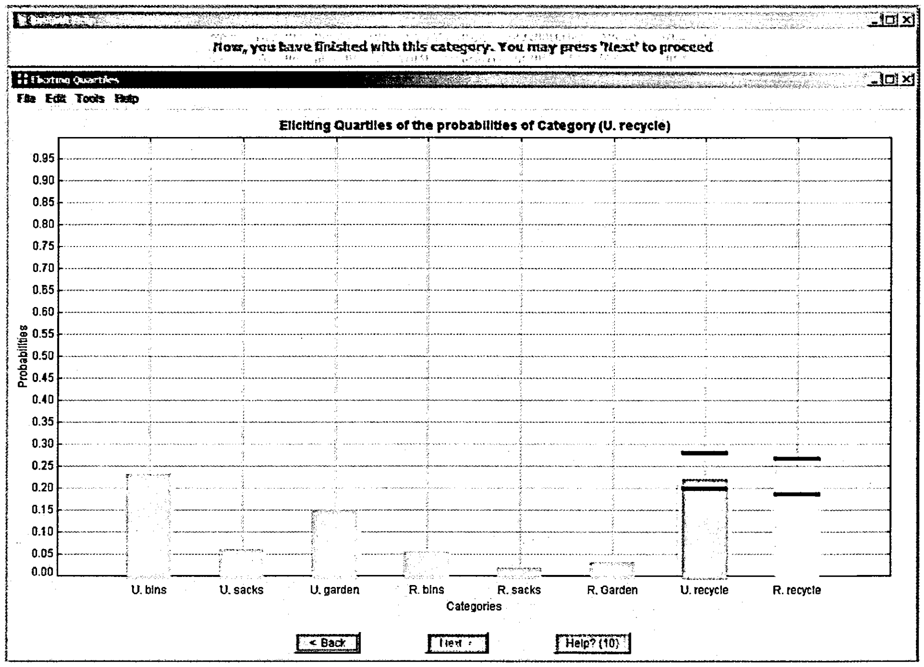


Figure 7.10: Assessing conditional quartiles for the last two categories

Next, conditional on the proportion for the first category being 0.12, the expert gave conditional median assessments for the proportions of the seven remaining categories. The number of conditions was then increased in stages. For example, in Figure 7.11, the expert has assessed the conditional medians for the last five categories given that the proportions for the first three categories are 0.12, 0.04 and 0.08, respectively. Table 7.4 gives all the conditional median assessments, where the underlined values constitute the conditioning set at each stage.

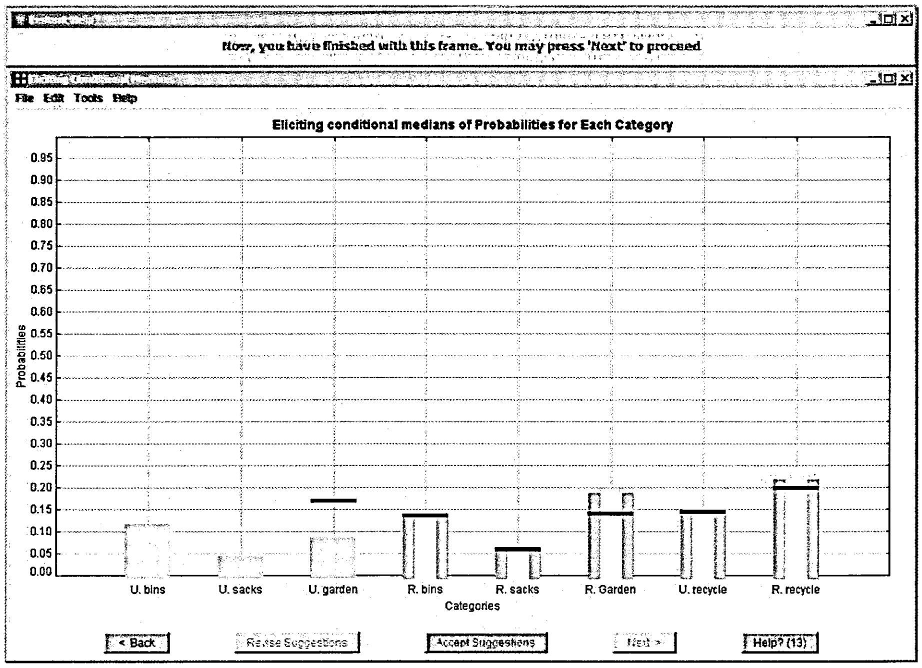


Figure 7.11: Assessing conditional medians

Table 7.4: Expert's assessments of conditional medians

p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
$\underline{0.12}$	0.09	0.16	0.10	0.06	0.12	0.15	0.14
$\underline{0.12}$	$\underline{0.04}$	0.16	0.14	0.06	0.14	0.14	0.2
$\underline{0.12}$	$\underline{0.04}$	$\underline{0.08}$	0.14	0.06	0.18	0.14	0.22
$\underline{0.12}$	$\underline{0.04}$	$\underline{0.08}$	$\underline{0.07}$	0.10	0.20	0.14	0.23
$\underline{0.12}$	$\underline{0.04}$	$\underline{0.08}$	$\underline{0.07}$	$\underline{0.05}$	0.22	0.15	0.26
$\underline{0.12}$	$\underline{0.04}$	$\underline{0.08}$	$\underline{0.07}$	$\underline{0.05}$	$\underline{0.11}$	0.22	0.33

This was the last assessment task, after which the software output the elicited hyperparameters of the marginal beta prior distributions as in Table 7.5. The dependence structure between these beta marginals was quantified as a multivariate Gaussian copula function with an elicited covariance matrix as given in Table 7.6.

Table 7.5: The elicited hyperparameters of marginal beta distributions

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
a	3.7607	1.0661	1.6536	0.6951	0.5731	0.6344	1.2489	0.4545
b	12.0047	14.8133	8.6742	8.6012	19.5493	14.8684	3.3578	3.3669

Table 7.6: The elicited covariance matrix of the Gaussian copula prior

	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
Y_1	1	-0.1279	-0.2601	-0.7773	-0.55	-0.6192	0.5414
Y_2	-0.1279	1	0.1328	0.1479	-0.4842	-0.3304	0.3326
Y_3	-0.2601	0.1328	1	-0.082	0.042	-0.03	-0.0618
Y_4	-0.7773	0.1479	-0.082	1	0.2358	0.4632	-0.4406
Y_5	-0.55	-0.4842	0.042	0.2358	1	0.5664	-0.5812
Y_6	-0.6192	-0.3304	-0.03	0.4632	0.5664	1	-0.8354
Y_7	0.5414	0.3326	-0.0618	-0.4406	-0.5812	-0.8354	1

The elicited matrix in Table 7.6 does not give covariances between the beta distributed proportions, $p_1, \cdots, p_8$. Instead, it gives the covariances between the transformed normal variates, $Y_1, \cdots, Y_7$. The eighth transformed normal variate is omitted so as to avoid the singularity of the elicited matrix, as discussed before. The Gaussian copula multivariate distribution is parameterized by both the marginal beta parameters and the covariance matrix in Table 7.6. The software produces a WinBUGS file with the Gaussian copula prior distribution. Marginal beta parameters can also be used to compute the expected value and variance of the proportions of each category. These are given in Table 7.7, where the expected values are very close to the coherent median assessments in Figure 7.8, and even closer to the initial median assessments in Table 7.2 and Figure 7.7.

The elicitation process took about an hour to complete. The expert stressed the importance of the convenient order of categories when conditioning. During the task of giving conditional assessments based on an increasing number of conditions, he commented that ordering the categories in a suitable sequence made it easier for him to think about these conditions according to his knowledge.

Table 7.7: Probability means and variances from marginal beta distributions

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
$E(p_i)$	0.239	0.067	0.160	0.075	0.028	0.041	0.271	0.119
$V(p_i)$	0.011	0.004	0.012	0.007	0.001	0.002	0.035	0.022

7.6 Concluding comments

The elicitation methods for beta parameters proposed in the previous chapter have been used in this chapter as the main tools for eliciting two more flexible prior distributions for multinomial models. A novel elicitation method for the generalized Dirichlet distribution has been introduced. The method makes use of the fact that the conditional distributions of the generalized Dirichlet variates are beta distributions. The method has been implemented in user-friendly software that is freely available as PEGS-Dirichlet at http://statistics.open.ac.uk/elicitation.

The elicitation of copula functions for multinomial models faces two obstacles, as noted in the literature. The usual correlations cannot be transformed through the assumed copula transformation, which is one obstacle, and the need to elicit a positive-definite variance-covariance matrix is the other. Our proposed elicitation method for the Gaussian copula prior has overcome both problems. The assessed conditional quartiles could be transformed through the normalizing one-to-one transformation, making it possible to elicit correlations. Moreover, the method of Kadane *et al.* (1980) has been modified to elicit a positive-definite variance-covariance matrix for the Gaussian copula. The method has been implemented in the user-friendly PEGS-Copula software that is freely available at http://statistics.open.ac.uk/elicitation.

Chapter 8

Eliciting logistic normal priors for multinomial models

8.1 Introduction

The logistic normal distribution has long been used as a multivariate distribution for proportions (Aitchison, 1986). The constrained proportions are obtained by transforming normally distributed unconstrained variables on the real space using some one-to-one transformation. Different multivariate logistic transformations are given in the literature, see for example Aitchison (1986). The most well-known and widely used logistic transformation, specially for multinomial logit models, is the additive logistic transformation.

We propose a method for quantifying opinion about a logistic normal prior for multinomial models. Our proposed method has been implemented in interactive graphical user-friendly software developed in Java. This is freely available as PEGS-Logistic at http://statistics.open.ac.uk/elicitation. The elicitation method proposed here is generalized in Chapter 9 to handle the case of multinomial models with covariates, or what are known as the multinomial logit models.

In Section 8.2 we define the logistic normal prior to be used and consider its assumptions. The required assessments with our structural procedure to elicit them using the software are given in Section 8.3. The use of these assessments to elicit the hyperparameters of the logistic normal prior distribution is proposed in Section 8.4. A method to obtain the prior's marginal quartiles, which are useful as feedback, is proposed in Section 8.5. We finish this chapter by giving an example in Sections 8.6 and some concluding comments in Section 8.7.

8.2 The additive logistic normal distribution

The additive logistic transformation from $\underline{Y}^*$ to $\underline{p}$ is defined by

$$p_i = \begin{cases} \dfrac{1}{1+\sum\limits_{j=2}^{k}\exp(Y_j)}, & \text{for } i=1, \\[2ex] \dfrac{\exp(Y_i)}{1+\sum\limits_{j=2}^{k}\exp(Y_j)}, & \text{for } i=2,3,\cdots,k, \end{cases} \tag{8.1}$$

with inverse transformation

$$Y_i = \log\left(\frac{p_i}{p_1}\right) = \log\left(\frac{p_i}{1 - p_2 - p_3 - \cdots - p_k}\right), \quad i = 2, 3, \cdots, k, \tag{8.2}$$

where

$$\underline{Y}^* = (Y_2, \quad Y_3, \quad \cdots, \quad Y_k) \sim \text{MVN}(\underline{\mu}_{k-1}, \Sigma_{k-1}). \tag{8.3}$$

- The transformation is one-to-one from the $k-1$ dimension random vector $\underline{Y}^*$ into the k dimension random vector $\underline{p}$. The definition of an extra random variable Y_1 will be given later.
- For any values $Y_2, \cdots, Y_k$, (8.1) gives $\sum_{i=1}^{k} p_i = 1$.
- The matrix Σ_{k-1} is non-singular.
- The transformation is not symmetric in the p_i, as we choose a fill-up variable

$$p_1 = 1 - p_2 - p_3 - \cdots - p_k.$$

- The transformation is used in the multinomial logit regression model when

$$Y_i = X'\underline{\beta}_i.$$

- If (8.3) applies, the elements of the vector $\underline{p}$ are said to have the multivariate logistic normal distribution. Their joint density has the form

$$f(\underline{p};\, \underline{\mu}_{k-1}, \Sigma_{k-1}) = \frac{1}{(2\pi)^{\frac{k-1}{2}} |\Sigma_{k-1}|^{\frac{1}{2}} (p_1 \times p_2 \times \cdots \times p_k)} \times$$
$$\exp\left\{ -\frac{1}{2} \quad [\log(\underline{p}_{k-1}/p_1) - \underline{\mu}_{k-1}]' \quad \Sigma_{k-1}^{-1} \quad [\log(\underline{p}_{k-1}/p_1) - \underline{\mu}_{k-1}] \right\},$$

where $\quad \underline{p}'_{k-1} = (p_2 \quad p_3 \quad \cdots \quad p_k), \quad 0 \le p_i \le 1, \quad \sum_{i=1}^{k} p_i = 1.$

- This additive logistic normal distribution is said to be permutation invariant. That is, whatever be the ordering of the elements of the vector $\underline{p}$, the density function given above is invariant. For a theoretical proof of this property see Aitchison (1986). Under the permutation invariance, any order of the elements of $\underline{p}$ can be considered. Consequently, the choice of the fill-up variable is arbitrary. Usually it is chosen as the probability of the most common category, the first category, or the last category. To elicit a logistic normal prior, we favour choosing the most common category as the first category and making p_1 the fill-up variable. This is more convenient for our method because of the order of conditioning we adopt later.

- For *sampling* compositional data, the problem of zero components has been reported by Aitchison (1986) as a critical irregular case that needs special attention in dealing with the logistic normal distribution. Clearly, the log transformation cannot be applied with zero components. However, we need not worry about this problem in our elicitation method, as categories with assessed zero probabilities can simply be removed from the analysis at the first early step without any loss.

We assume that prior opinion about $\underline{Y}^*$ can be represented by the multivariate normal distribution in (8.3). As will be shown later, for the assessments of $\underline{p}$ to be fully transformable to $\underline{Y}^*$, a further normalizing transformation must be defined on the fill-up variable p_1. We define an extra variable Y_1 such that

$$Y_1 = \log\left(\frac{p_1}{1-p_1}\right). \tag{8.4}$$

Based on the normality assumption of $\underline{Y}^*$ in (8.3) and the unit sum constraint of $\underline{p}$, the random variable e^{-Y_1} can be represented as a sum of $k-1$ lognormally distributed random variables, since

$$e^{-Y_1} = \frac{1-p_1}{p_1} = \sum_{i=2}^{k} \frac{p_i}{p_1}.$$

Although the sum of lognormal random variables has no simple exact distribution, it is common to approximate its distribution by another lognormal distribution. This is discussed in the next section.

8.2.1 Approximate distribution of the lognormal sum

Fenton (1960) considered the numerical convolution of lognormal distributions and showed that the sum of such distributions is a distribution that approximately follows the lognormal law. He added that the sum of two (or more) lognormal distributions can be assumed, as a first approximation, to have another lognormal distribution. Later, Schwartz and Yeh (1982) mentioned that 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.

Several approximations have been introduced for the sum of lognormal random variables. Although the idea of approximating their sum using another lognormal distribution has been common in many studies, methods differs in approximating the moments of the lognormal distribution of the sum. Fenton (1960) matches the first two moments of the sum of lognormal random variables to the first two moments of an equivalent lognormal random variable. Schwartz and Yeh (1982) follow the same approach but compute the exact first two moments for the sum of two lognormal random variables; the procedure is then iteratively applied for the sum of more than two lognormal random variables. Their method of computing the distribution of a sum of independent lognormal random variables was extended to the case of correlated lognormal random variables by Safak (1993).

Recently, based on approximating the distribution of the sum of lognormal random variables by another lognormal distribution, a lot of work have been devoted to giving various approximation methods. For example, Beaulieu and Xie (2004) uses a linearizing transform with a linear minimax approximation to determine an optimal lognormal approximation to a lognormal sum distribution. Tellambura and Senaratne (2010) use the classical complex integration techniques to approximate the moment generating function of the sum. Mahmoud (2010) approximates the characteristic function and the cumulative distribution function of the lognormal sum by exploiting the recent Hermit-Gauss quadrature-based approximation.

It is thus natural to approximate the distribution of Y_1 by a normal distribution with elicited mean and variance. (We do not require any approximations to obtain its parameters.) We can then state our main assumption:

$$\underline{Y}_k = (Y_1, \ \ Y_2, \ \ \ldots, \ \ Y_k)' \sim \text{MVN}(\underline{\mu}_k, \Sigma_k). \tag{8.5}$$

The unit sum constraint of $\underline{p}$ will always lead to a singular matrix Σ_k. However, we assume that there is only one condition on the elements of $\underline{p}$, namely the unit sum. In particular, we assume that there does not exist any subset of categories such that the sum of their probabilities is known with certainty.

Although no density function can be defined for the singular multivariate normal distribution, its theoretical properties and numerical results have been investigated in the literature. See, for example, Bland and Owen (1966), Kwong and Iglewicz (1996), Albajar and Fidalgo (1997) or Genz and Kwong (1999).

Usage of the singular normal is thus feasible and has been exploited in numerous multivariate methods. Khatri (1968) used the notion of a generalized inverse to utilize the singular normal distribution in multivariate regression. Styan (1970) discussed the distribution of quadratic forms in singular normal variables. West and Harrison (1997) defined the covariance matrix of the multivariate normal distribution as a non-negative definite matrix.

In Chapter 8 of his book on linear statistical inference, Rao (2002) did not use the density function to define the multivariate normal distribution. Instead, he characterized it by the property that every linear function of its elements has a univariate normal distribution. He could then list properties and characterizations of the multivariate normal distribution without using the pdf. The singular normal distribution is thus a special case of the standard normal distribution, and has similar properties, but with the usual inverse of the covariance matrix replaced by its generalized inverse. Conditional properties of the singular normal distribution have been extensively used in the current chapter for eliciting a logistic normal distribution.

To this end, using (8.5), we assume that the prior distribution of $\underline{p}$ is the logistic normal distribution induced by the vector

$$\underline{Y}^* = A\underline{Y}_k \sim \text{MVN}(\underline{\mu}_{k-1}, \Sigma_{k-1}).$$

where

$$A = \left[\underline{0} \,\vdots\, I_{k-1}\right], \tag{8.6}$$

$$\underline{\mu}_{k-1} = A\underline{\mu}_k, \tag{8.7}$$

$$\Sigma_{k-1} = A\Sigma_k A'. \tag{8.8}$$

We start by eliciting $\underline{\mu}_k$ and a matrix Σ_k of rank $k-1$. In our approach, we modify the method of Kadane *et al.* (1980) and add a special treatment for the kth row and column. This will give the $\underline{\mu}_{k-1}$ and Σ_{k-1} in equations (8.7) and (8.8). The matrix Σ_k is singular of rank $k-1$, given that no other constraint can be imposed on subsets of probabilities except the unit sum. However, the matrix Σ_{k-1} is shown to be positive-definite of full rank $k-1$, since it is simply Σ_k with its first row and column removed. A formal proof of the positive-definiteness of Σ_{k-1} will be given later in Section 8.4.2.

8.3 Assessment tasks

Since the transformations in (8.2) and (8.4) are strictly monotonic increasing from $\underline{p}$ to $\underline{Y}_k$, we can establish a one-to one correspondence between the medians and quartiles of these two vectors. The required assessments are detailed as follows.

8.3.1 Assessing initial medians

- The choice of a category to start with is arbitrary, as discussed earlier. Hence it may be chosen by the expert as the most common category and its probability is denoted p_1. A median value m_1 for p_1 will be assessed as a first step. Then the expert assesses median values m_j, $j = 2, \cdots, k$, for all the remaining categories. These assessed values are shown by the blue bars in Figure 8.1.

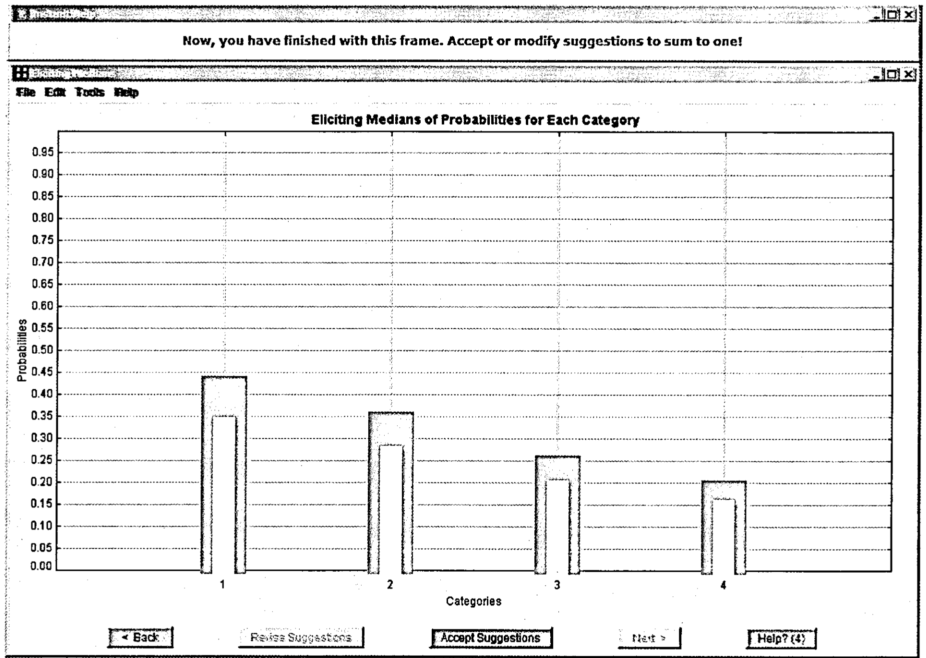


Figure 8.1: Assessing probability medians for logistic normal elicitation

- The normality assumption of $\underline{Y}_k$, together with the unit sum constraint of $\underline{p}$, can be used in Lemma 8.1 and Theorem 8.1 (which are given in Section 8.4) to show that the unit sum constraint must be also fulfilled by the m_j. That is $\sum_{j=1}^{k} m_j = 1$. To attain mathematical coherence, the software suggests a normalized set of assessments, given by the yellow bars in Figure 8.1, as follows. Suppose the initial assessments were $m'_1, m'_2, \cdots, m'_k$. Then the coherent assessments that are suggested for the m_j are given by

$$m_j = \frac{m'_j}{\sum\limits_{i=1}^{k} m'_i}, \quad \text{for } j = 1, 2, \cdots, k.$$

With our software, the expert can keep changing her assessed values until she is happy with the normalized values that are suggested.

8.3.2 Assessing conditional quartiles

- In this assessment task, the expert is asked to assess a lower quartile L_1^* and an upper quartile U_1^* for p_1. She is then asked to assume that $p_1 = m_1$ and gives a lower quartile L_2^* and an upper quartile U_2^* for p_2. For each remaining p_j, $j = 3, \cdots, k-1$, she assesses the two quartiles L_j^* and U_j^* given that $p_1 = m_1$, $p_2 = m_2$, $\cdots$, $p_{j-1} = m_{j-1}$. See Figure 8.2, 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 red bars.

- 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}, see Figure 8.2. The two quartiles L_k^* and U_k^* are also shown to the expert as a guide to help her choose L_{k-1}^* and U_{k-1}^*. In fact, L_k^* (U_k^*) is the lower (upper) quartile of $(p_k|p_1 = m_1, \cdots, p_{k-2} = m_{k-2})$ as $L_{k-1}^* + U_k^* = U_{k-1}^* + L_k^* = 1 - m_1 - \cdots - m_{k-2}$, from the unit sum constraint.

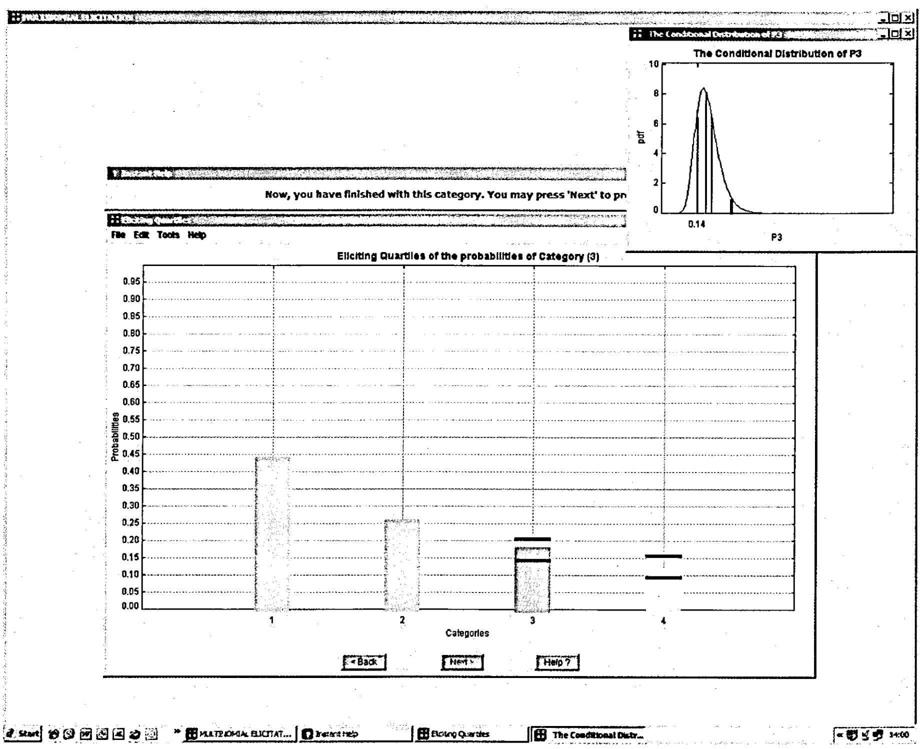


Figure 8.2: Assessing conditional quartiles with lognormal feedback

- To help the expert during this current task, the software presents an interactive graph showing the pdf curve of the lognormal distribution of $(p_j|p_1 = m_1, \cdots, p_{j-1} = m_{j-1})$, for $j = 2, 3, \cdots, k-1$, see Figure 8.2. 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. 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 red vertical line on the pdf graph of Figure 8.2. See Lemma 8.2 for the formal validity of the above results.

8.3.3 Assessing conditional medians

- Here, the expert is asked to assume that the median of p_1 has been changed from m_1 to $m^*_{1,1} = m_1 + \eta^*_1$. Given this information, the expert will be asked to change her previous medians m_j of each p_j. Her new assessment, $m^*_{j,1}$, may be written as

$$m^*_{j,1} = m_j + \theta^*_{j,1}, \quad \text{for } j = 2, \cdots, k. \tag{8.9}$$

- In each successive step i, for $i = 2, 3, \cdots, k-2$, the expert will be asked to suppose that the median values of p_1, p_2, $\cdots$, p_i are $m^*_{1,1} = m_1 + \eta^*_1, m^*_{2,2} = m^*_{2,1} + \eta^*_2, \cdots, m^*_{i,i} = m^*_{i,i-1} + \eta^*_i$, respectively, shown as red bars in Figure 8.3. Given this information, she will be asked to change her assessed medians of the most recent previous step $m^*_{i+1,i-1}$, $m^*_{i+2,i-1}, \cdots, m^*_{k,i-1}$, shown by black lines in Figure 8.3. Her new assessments are $m^*_{i+1,i} = m^*_{i+1,i-1} + \theta^*_{i+1,i}$, $m^*_{i+2,i} = m^*_{i+2,i-1} + \theta^*_{i+2,i}$, $\cdots$, $m^*_{k,i} = m^*_{k,i-1} + \theta^*_{k,i}$, respectively, which are shown as the blue bars in Figure 8.3. For $i = 2, 3, \cdots, k-2$, and $j = i+1, i+2, \cdots, k$, we can write

$$m^*_{j,i} = m^*_{j,i-1} + \theta^*_{j,i} \text{ is the median of } (p_j | p_1 = m^*_{1,1}, \cdots, p_i = m^*_{i,i}). \tag{8.10}$$

- For mathematical coherence, as will be proved in Lemma 8.3, we have to make sure that

$$\sum_{j=1}^{i} m^*_{j,j} + \sum_{j=i+1}^{k} m^*_{j,i} = 1, \qquad i = 1, 2, \cdots, k-2.$$

The expert has the option of changing her initial set of assessments $m'_{i+1,i}$, $m'_{i+2,i}$, $\cdots$, $m'_{k,i}$, the blue bars on Figure 8.3, until she feels that the suggested normalized set $m^*_{i+1,i}$, $m^*_{i+2,i}$, $\cdots$, $m^*_{k,i}$, shown as yellow bars on Figure 8.3, gives the best representation of her opinion. The software suggests each normalized conditional median $m^*_{j,i}$ as

$$m^*_{j,i} = \left[\frac{1 - \sum_{r=1}^{i} m^*_{r,r}}{\sum_{r=i+1}^{k} m'_{r,i}} \right] m'_{j,i}, \qquad \text{for } i = 1, \cdots, k-2, \quad j = i+1, \cdots, k.$$

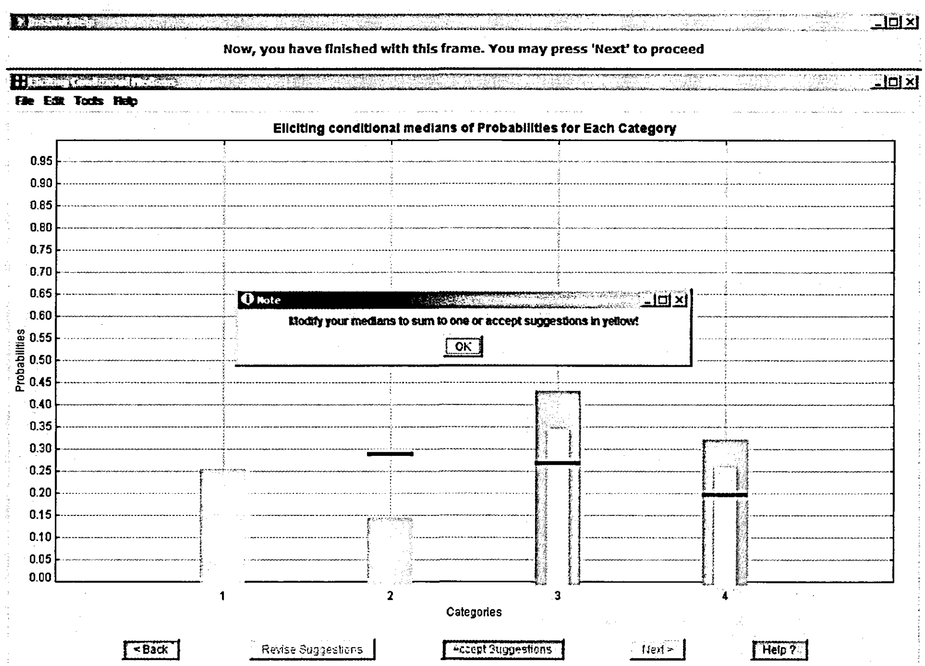


Figure 8.3: Assessing conditional medians for logistic normal elicitation

- 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. As the condition of summing to one must be fulfilled, conditioning on specific values of all $p_1, p_2, \cdots, p_{k-1}$ gives a fixed value for p_k. Then no upper or lower quartiles can be assessed for p_k, as mentioned before. Conditional medians of Y_k given specific values of $Y_1, Y_2, \cdots, Y_{k-1}$ can be automatically computed when needed, as will be shown later.

8.4 Eliciting prior hyperparameters

The normalizing one-to-one functions in equations (8.2) and (8.4) are used to transform the assessed conditional quartiles of $\underline{p}$ into conditional quartiles of $\underline{Y}_k$ and, hence, into conditional expectations, variances and covariances of the multivariate normal variables. In particular, letting $M(X)$ denote the median function of the random variable X, we proceed as follows.

Let

$$m^*_{j,0} = \begin{cases} m_1, & \text{for } j = 1, \\ M(p_j|p_1 = m_1), & \text{for } j = 2, 3, \cdots, k. \end{cases} \tag{8.11}$$

Since the normal variates, $Y_j = \log(p_j/p_1)$, $j = 2, \cdots, k$, depend on the fill-up probability p_1, eliciting prior hyperparameters for $\underline{Y}^*$ is tractable if we condition on p_1. That is why we define the extra normal variate Y_1 as in (8.4) and the conditional medians, $m^*_{j,0}$, as in (8.11). These conditional medians are required instead of the assessed unconditional medians, m_j, to elicit the hyperparameters of the logistic normal prior distribution. However, we chose to elicit the unconditional medians as they are easier to assess than conditional medians. Fortunately, under the normality assumption of $\underline{Y}^*$ and the unit sum constraint of $\underline{p}$, we will show in Theorem 8.1 below that the marginal unconditional medians, m_j, are identical to conditional medians, $m^*_{j,0}$, of p_j, for $j = 1, 2, \cdots, k$, respectively, provided the lognormal sum is adequately approximated by another lognormal random variable.

For $i = 1, 2, \cdots, k$, let

$$m_{i,0} = E(Y_i). \tag{8.12}$$

Remark 8.1

It is worth noting that $Y_1 = m_{1,0}$ when $p_1 = m^*_{1,0}$, but, $Y_i = m_{i,0}$ when both $p_i = m^*_{i,0}$ and $p_1 = m^*_{1,0}$, for $i = 2, 3, \cdots, k$.

Extensive use is made of the fact that each Y_i follows a symmetric distribution (each has a normal distribution), so $E(Y_i) = M(Y_i)$. This is a key assumption in proving the following lemma, which states an important result that is needed in the proof of Theorem 8.1.

Lemma 8.1. *Under the unit sum constraint of $\underline{p}$, and the multivariate normality of $\underline{Y}_k$,*

$$\sum_{i=1}^{k} m^*_{i,0} = 1.$$

Proof

As given by (Rao, 2002, p 522), the conditional distribution of any subset of singular normal random variables is normally distributed with the usual conditional mean and variance, but with generalized inverses of matrices. This property enables us to write, as in the non-singular case,

$$\begin{aligned} E(Y_k) &= E[Y_k|Y_1 = E(Y_1)] \\ &= E[Y_k|Y_1 = E(Y_1), Y_2 = E(Y_2), \cdots, Y_{k-1} = E(Y_{k-1})]. \end{aligned}$$

Then, replacing means by medians and using (8.12), we get

$$M(Y_k|Y_1 = m_{1,0}) = M(Y_k|Y_1 = m_{1,0}, Y_2 = m_{2,0}, \cdots, Y_{k-1} = m_{k-1,0}).$$

Hence, from Remark 8.1,

$$\begin{aligned} &M[\log(p_k) - \log(p_1)|p_1 = m^*_{1,0}] = \\ &\qquad M[\log(p_k) - \log(p_1)|p_1 = m^*_{1,0}, \cdots, p_{k-1} = m^*_{k-1,0}], \end{aligned}$$

which gives

$$\begin{aligned} &\log(m^*_{k,0}) - \log(m^*_{1,0}) = \\ &\qquad \log[M(p_k|p_1 = m^*_{1,0}, \cdots, p_{k-1} = m^*_{k-1,0})] - \log(m^*_{1,0}). \end{aligned}$$

i.e.

$$\begin{aligned} m^*_{k,0} &= M(p_k|p_1 = m^*_{1,0}, p_2 = m^*_{2,0}, \cdots, p_{k-1} = m^*_{k-1,0}) \\ &= 1 - \sum_{i=1}^{k-1} m^*_{i,0}. \end{aligned}$$

This is the unit sum constraint, which completes the proof of Lemma 8.1.

The main idea in Theorem 8.1 is that the fill-up category can be changed from the first category to any other category, and the same assumptions are still valid. We first give some relations and notations needed for the proof of the theorem.

Let $\underline{Y}_{(1)} = \underline{Y}^*$, and denote the mean vector and variance-covariance matrix of the multivariate normal distribution of $\underline{Y}_{(1)}$ by $\underline{\mu}_{(1)} = \underline{\mu}_{k-1}$, and $\Sigma_{(1)} = \Sigma_{k-1}$. We supposed $\underline{\mu}_{(1)}$ and $\Sigma_{(1)}$ have already been assessed. Moreover, let $Y_{1,1} = \log(p_1) - \log(1-p_1)$, with $E_1 = E(Y_{1,1})$ and $V_1 = \mathrm{Var}(Y_{1,1})$.

To change the fill-up category from the first category to any other category j, for $j = 2, 3, \cdots, k$, let

$$\underline{Y}_{(j)} = \begin{bmatrix} Y_{1,j} \\ \vdots \\ Y_{j-1,j} \\ Y_{j+1,j} \\ \vdots \\ Y_{k,j} \end{bmatrix} = \begin{bmatrix} \log(p_1) - \log(p_j) \\ \vdots \\ \log(p_{j-1}) - \log(p_j) \\ \log(p_{j+1}) - \log(p_j) \\ \vdots \\ \log(p_k) - \log(p_j), \end{bmatrix},$$

$$Y_{j,j} = \log(p_j) - \log(1 - p_j),$$

with

$$\underline{\mu}_{(j)} = E(\underline{Y}_{(j)}), \qquad \Sigma_{(j)} = \mathrm{Var}(\underline{Y}_{(j)}),$$

$$E_j = E(Y_{j,j}), \qquad V_j = \mathrm{Var}(Y_{j,j}),$$

and

$$\mu_{i,j} = E(Y_{i,j}), \qquad \sigma^2_{i,j} = \mathrm{Var}(Y_{i,j}), \qquad i, j = 1, 2, \cdots, k, \quad i \neq j.$$

It can easily be shown that, for $j = 2, \cdots, k$,

$$\underline{Y}_{(j)} = F_j \underline{Y}_{(j-1)}, \tag{8.13}$$

where F_j is the identity matrix of degree $k-1$ with the jth column replaced by a column of -1. From the normality assumption of $\underline{Y}_{(1)}$, and in view of (8.13), we have

$$\underline{Y}_{(j)} \sim \mathrm{MVN}(\underline{\mu}_{(j)}, \Sigma_{(j)}),$$

with

$$\underline{\mu}_{(j)} = F_j\, \underline{\mu}_{(j-1)},$$
$$\Sigma_{(j)} = F_j\, \Sigma_{(j-1)}\, F_j'.$$

Approximate normality of each $Y_{j,j}$, for $j = 2, 3, \cdots, k$, is thus induced from the normality assumption of $\underline{Y}_{(j)}$ in a manner exactly similar to that for $Y_{1,1}$. Hence, for each $j = 1, 2, \cdots, k$, we can also assume that the k random variables $Y_{i,j}$, for $i = 1, 2, \cdots, k$, are multivariate normally distributed. Moreover, using the normality assumption of $Y_{1,1}$, we assume that the $k+1$ random variables $Y_{1,1}$ and $Y_{i,j}$, for $i = 1, 2, \cdots, k$, are also multivariate normally distributed for each $j = 1, 2, \cdots, k$.

Theorem 8.1. *For any $j = 2, 3, \cdots, k$, under the unit sum constraint of $\underline{p}$, and the multivariate normality of $\underline{Y}_{(j)}$,*

$$m_j = M(p_j) = M(p_j | p_1 = m_{1,0}^*) = m_{j,0}^*.$$

Proof

Let

$$m_{i,(j)} = M[\log(p_i) - \log(p_j)], \quad i = 1, 2, \cdots, k, \quad i \neq j,$$

then

$$\begin{aligned} m_{i,(j)} =& E(Y_{i,j}) \\ =& E(Y_{i,j} | Y_{j,j} = E_j) \\ =& M[\log(p_i) - \log(p_j) | p_j = M(p_j)]. \end{aligned}$$

Hence, exponentiating both sides of the above relation, we get

$$M[p_i | p_j = M(p_j)] = M(p_j)\ \exp(m_{i,(j)}). \tag{8.14}$$

As in Lemma 8.1, we put

$$M(p_j) + \sum_{i \neq j}^{k} M[p_i | p_j = M(p_j)] = 1. \tag{8.15}$$

Solving (8.14) and (8.15) for $M(p_j)$, we get

$$M(p_j) = \frac{1}{1 + \sum_{i \neq j}^{k} \exp(m_{i,(j)})}. \tag{8.16}$$

On the other hand, for $j \neq 1$, since

$$\Pr\{p_j < m_{j,0}^* | p_1 = m_{1,0}^*\} = 0.5,$$

then

$$\Pr\{(p_j/p_1) < (m_{j,0}^*/m_{1,0}^*) | p_1 = m_{1,0}^*\} = 0.5,$$

and

$$\Pr\{\log(p_1/p_j) < \log(m_{1,0}^*/m_{j,0}^*) | Y_{1,1} = E_1\} = 0.5.$$

So, we can write

$$\begin{aligned} \log(m_{1,0}^*/m_{j,0}^*) &= M(Y_{1,j} | Y_{1,1} = E_1) \\ &= E(Y_{1,j} | Y_{1,1} = E_1) = E(Y_{1,j}) = m_{1,(j)}. \end{aligned} \tag{8.17}$$

Moreover, for $j \neq i \neq 1$, since

$$\begin{aligned} m_{i,(j)} =& E(Y_{i,j}) \\ =& E(Y_{i,j} | Y_{1,1} = E_1, Y_{j,j} = E(Y_{j,j} | Y_{1,1} = E_1)) \\ =& M[\log(p_i/p_j) | p_1 = m_{1,0}^*, p_j = m_{j,0}^*], \end{aligned}$$

we have that

$$\Pr\{\log(p_i/p_j) < m_{i,(j)} | p_1 = m_{1,0}^*, p_j = m_{j,0}^*\} = 0.5,$$

and

$$\Pr\{p_i < m_{j,0}^* \ \exp(m_{i,(j)}) | p_1 = m_{1,0}^*\} = 0.5.$$

So,

$$m_{i,0}^* = m_{j,0}^* \ \exp(m_{i,(j)}),$$

which gives

$$m_{i,(j)} = \log(m_{i,0}^*/m_{j,0}^*). \tag{8.18}$$

Substituting (8.17) and (8.18) into (8.16) shows that $M(p_j)$ is as stated in Theorem 8.1.

8.4.1 Eliciting a mean vector

To elicit a mean vector $\underline{\mu}'_k = (m_{1,0} \quad m_{2,0} \quad \cdots \quad m_{k,0})$, we put

$$m_{1,0} = E(Y_1) = M(Y_1) \tag{8.19}$$

$$= M(\log(p_1) - \log(1-p_1))$$

$$= \log(m^*_{1,0}) - \log(1 - m^*_{1,0}). \tag{8.20}$$

For $i = 2, 3, \cdots, k$, put

$$m_{i,0} = E(Y_i) = E[Y_i | Y_1 = E(Y_1)] \tag{8.21}$$

$$= M(Y_i | Y_1 = m_{1,0})$$

$$= M[\log(p_i) - \log(p_1) | p_1 = m^*_{1,0}]$$

$$= \log(m^*_{i,0}) - \log(m^*_{1,0}). \tag{8.22}$$

8.4.2 Eliciting a variance-covariance matrix

For $i = 1, 2, \cdots, k-2$, and $j = i+1, \cdots, k-1$, let

$$m_{j,i} = E(Y_j | p_1 = m^*_{1,0} + \eta^*_1, p_2 = m^*_{2,1} + \eta^*_2, \cdots, p_i = m^*_{i,i-1} + \eta^*_i).$$

Then

$$m_{j,i} = \log\left(\frac{m^*_{j,i}}{m^*_{1,1}}\right). \tag{8.23}$$

For $i = 1, 2, \cdots, k-2$, define η_i by letting $\eta_i = Y_i - m_{i,i-1}$ when $p_i = m^*_{i,i-1} + \eta^*_i$. Then

$$m_{j,i} = E(Y_j | Y_1 = m_{1,0} + \eta_1, Y_2 = m_{2,1} + \eta_2, \cdots, Y_i = m_{i,i-1} + \eta_i), \tag{8.24}$$

and

$$\eta_i = \begin{cases} \log\left(\dfrac{m^*_{1,0} + \eta^*_1}{1 - (m^*_{1,0} + \eta^*_1)}\right) - \log\left(\dfrac{m^*_{1,0}}{1 - m^*_{1,0}}\right), & \text{for } i = 1, \\ \log\left(\dfrac{m^*_{i,i-1} + \eta^*_i}{m^*_{1,1}}\right) - \log\left(\dfrac{m^*_{i,i-1}}{m^*_{1,1}}\right), & \text{for } i = 2, 3, \cdots, k-2. \end{cases}$$

Analogous to $m^*_{i,i} = m^*_{i,i-1} + \eta^*_i$, define

$$m_{i,i} = m_{i,i-1} + \eta_i, \qquad \text{for } i = 1, 2, \cdots, k-2, \tag{8.25}$$

so $Y_i = m_{i,i}$ when $p_i = m^*_{i,i}$.

For $i = 1, 2, \cdots, k-2$, and $j = i+1, \cdots, k-1$, analogous to $\theta^*_{j,i} = m^*_{j,i} - m^*_{j,i-1}$, define

$$\theta_{j,i} = m_{j,i} - m_{j,i-1},$$

so that

$$\theta_{j,i} = \log\left(\frac{m^*_{j,i-1} + \theta^*_{j,i}}{m^*_{1,1}}\right) - \log\left(\frac{m^*_{j,i-1}}{m^*_{1,1}}\right).$$

To elicit a (singular) variance-covariance matrix Σ_k of rank $k-1$, let

$$V_1 = \mathrm{Var}(Y_1) = \left[\frac{U_1 - L_1}{1.349}\right]^2, \tag{8.26}$$

where U_1 and L_1 are the upper and lower quartile of Y_1, respectively. We have that

$$U_1 = \log(U^*_1/1 - U^*_1),$$

$$L_1 = \log(L^*_1/1 - L^*_1).$$

For $i = 1, 2, \cdots, k-2$, and $j = i+1, \cdots, k-1$, let

$$V_{j,i} = \mathrm{Var}(Y_j | Y_1 = m_{1,0}, Y_2 = m_{2,0}, \cdots, Y_i = m_{i,0}),$$

so that

$$V_{j,j-1} = \left[\frac{U_j - L_j}{1.349}\right]^2, \quad \text{for } j = 2, 3, \cdots, k-1, \tag{8.27}$$

with

$$U_j = \log\left(\frac{U^*_j}{m^*_{1,0}}\right), \quad L_j = \log\left(\frac{L^*_j}{m^*_{1,0}}\right).$$

Having defined the above quantities, we are ready to state and prove the following two lemmas.

Lemma 8.2. *Under the assumptions of Lemma 8.1, for* $i = 2, \cdots, k-1$,

1. $$(p_i | p_1 = m^*_{1,0}, p_2 = m^*_{2,0}, \cdots, p_{i-1} = m^*_{i-1,0}) \sim \mathit{Lognormal}(\mu^*_i, V^*_i),$$

   where

   $$\mu^*_i = m_{i,0} + \log(m^*_{1,0}) = \log(m^*_{i,0}),$$

and

$$V_i^* = V_{i,i-1} = \left[\frac{U_i - L_i}{1.349}\right]^2.$$

2.
$$Pr\left\{p_i > 1 - \sum_{j=1}^{i-1} m_{j,0}^*\right\} < \alpha,$$

if and only if
$$\frac{U_i^*}{L_i^*} < \exp\left\{\frac{1.349}{z_{1-\alpha}}\left[\log\left(1 - \sum_{j=1}^{i-1} m_{j,0}^*\right) - \mu_i^*\right]\right\},$$

where z_α is the α quantile of the standard normal distribution.

Proof

From the normality of $\underline{Y}_k$ together with property (v) of the singular normal distribution in (Rao, 2002, p 522), we have

$$(Y_i|Y_1 = m_{1,0}, \cdots, Y_{i-1} = m_{i-1,0}) \sim \mathrm{N}(m_{i,0}, V_{i,i-1}).$$

Then for known fixed $m_{1,0}^*$,

$$(Y_i + \log(m_{1,0}^*)|Y_1 = m_{1,0}, \cdots, Y_{i-1} = m_{i-1,0}) \sim \mathrm{N}(m_{i,0} + \log(m_{1,0}^*), V_{i,i-1}).$$

The one-to-one transformations in (8.2) and (8.4) then give

$$\begin{aligned}&(\frac{p_i}{p_1} m_{1,0}^*|p_1 = m_{1,0}^*, \cdots, p_{i-1} = m_{i-1,0}^*)\\ &= (p_i|p_1 = m_{1,0}^*, \cdots, p_{i-1} = m_{i-1,0}^*) \sim \text{ Lognormal}(m_{i,0} + \log(m_{1,0}^*), V_{i,i-1}).\end{aligned}$$

Using equation (8.22), the first statement of the lemma is proved.

To prove the second statement, we use standard normal distribution theory and the first statement of this lemma to state that

$$\Pr\left\{\frac{\log(p_i) - \mu_i^*}{\sqrt{V_i^*}} > \frac{\log\left(1 - \sum_{j=1}^{i-1} m_{j,0}^*\right) - \mu_i^*}{\sqrt{V_i^*}}\right\} < \alpha,$$

if and only if

$$\sqrt{V_i^*} < \frac{\log\left(1 - \sum_{j=1}^{i-1} m_{j,0}^*\right) - \mu_i^*}{z_{1-\alpha}},$$

or, equivalently, if and only if

$$U_i - L_i < \frac{1.349}{z_{1-\alpha}} \left[\log\left(1 - \sum_{j=1}^{i-1} m_{j,0}^*\right) - \mu_i^* \right].$$

This proves the second statement.

Lemma 8.3. *Under the assumptions of Lemma 8.1,*

$$\sum_{j=1}^{i} m_{j,j}^* + \sum_{j=i+1}^{k} m_{j,i}^* = 1, \qquad i = 1, 2, \cdots, k-2.$$

Proof

Using equation (8a.2.11) of (Rao, 2002, p 522), for $i = 1, 2, \cdots, k-2$, we can state that

$$E[Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i}] = E[Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i},$$

$$Y_{i+1} = E(Y_{i+1}|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i}), \cdots, Y_{k-1} = E(Y_{k-1}|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i})].$$

Then, from definition (8.24) and (8.25)

$$M(Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i}) = M(Y_k|Y_1 = m_{1,1}, \cdots, Y_i = m_{i,i},$$

$$Y_{i+1} = m_{i+1,i}, \cdots, Y_{k-1} = m_{k-1,i}).$$

Hence

$$M[\log(p_k) - \log(p_1)|p_1 = m_{1,1}^*, \cdots, p_i = m_{i,i}^*] =$$

$$M[\log(p_k) - \log(p_1)|p_1 = m_{1,1}^*, \cdots, p_i = m_{i,i}^*, p_{i+1} = m_{i+1,i}^*, \cdots, p_{k-1} = m_{k-1,i}^*],$$

which, utilizing equations (8.9) and (8.10), gives

$$\log(m_{k,i}^*) - \log(m_{1,1}^*) =$$

$$\log[M(p_k|p_1 = m_{1,1}^*, \cdots, p_i = m_{i,i}^*, p_{i+1} = m_{i+1,i}^*, \cdots, p_{k-1} = m_{k-1,i}^*)] - \log(m_{1,1}^*),$$

i.e.

$$m_{k,i}^* = M(p_k|p_1 = m_{1,1}^*, \cdots, p_i = m_{i,i}^*, p_{i+1} = m_{i+1,i}^*, \cdots, p_{k-1} = m_{k-1,i}^*).$$

Since the condition in the RHS of the above equation relates to all ps except p_k, applying the unit sum constraint gives the conditional median in the form of the following complement:

$$m^*_{k,i} = 1 - \sum_{j=1}^{i} m^*_{j,j} - \sum_{j=i+1}^{k-1} m^*_{j,i},$$

which ends the proof of Lemma 8.3.

Now, we modify the method of Kadane *et al.* (1980) to show that the quantities in (8.24)-(8.27) are sufficient to elicit a positive-definite variance-covariance matrix V_{k-1} for $\underline{Y}_{k-1} = (Y_1, \cdots, Y_{k-1})$. Then, based on the condition of $\sum_{i=1}^{k} p_i = 1$, and assuming that it is the only constraint on sums of these probabilities, we add a kth row and column to get Σ_k as a singular variance-covariance matrix for all the elements of $\underline{Y}_k$. Removing the first row and column of Σ_k will lead to the desired positive-definite variance-covariance matrix Σ_{k-1} of $\underline{Y}^*$.

- For $i = 1, 2, \cdots, k-1$, let

  $$\underline{Y}_i = (Y_1, Y_2, \cdots, Y_i),$$

  and

  $$V_i = \text{Var}(\underline{Y}_i),$$

  with V_1 as defined in (8.26). Suppose that V_{i-1} has been estimated as a positive-definite matrix. We aim now to elicit V_i and investigate its positive-definiteness.

  V_i can be partitioned as

  $$V_i = \begin{bmatrix} V_{i-1} & V_{i-1}\underline{u}_i \\ \underline{u}'_i V_{i-1} & \sigma_i^2 \end{bmatrix}, \tag{8.28}$$

  where

  $$V_{i-1}\underline{u}_i = \text{Cov}(\underline{Y}_{i-1}, Y_i),$$

  and

  $$\sigma_i^2 = \text{Var}(Y_i).$$

- It is well-known from multivariate normal distribution theory that

$$
\begin{aligned}
E(Y_i|\underline{Y}_{i-1}) - E(Y_i) &= [\underline{Y}_{i-1} - E(\underline{Y}_{i-1})]' V_{i-1}^{-1} V_{i-1} \underline{u}_i \\
&= [\underline{Y}_{i-1} - E(\underline{Y}_{i-1})]' \underline{u}_i. \qquad (8.29)
\end{aligned}
$$

Moreover, for $j \leq i-1$, taking the conditional expectation of both sides of (8.29), given that

$$\underline{y}_j = (m_{1,0} + \eta_1, m_{2,1} + \eta_2, \cdots, m_{j,j-1} + \eta_j)',$$

gives

$$E\left[E(Y_i|\underline{Y}_{i-1})|\underline{Y}_j = \underline{y}_j\right] - E(Y_i) = E\{[\underline{Y}_{i-1} - E(\underline{Y}_{i-1})]|\underline{Y}_j = \underline{y}_j\}' \ \underline{u}_i. \qquad (8.30)$$

i.e.

$$
\begin{aligned}
&E(Y_i|\underline{Y}_j = \underline{y}_j) - E(Y_i) \\
&\quad = (y_1 - E(Y_1), y_2 - E(Y_2), \cdots, y_j - E(Y_j), \\
&\qquad E(Y_{j+1}|\underline{Y}_j) - E(Y_{j+1}), \cdots, E(Y_{i-1}|\underline{Y}_j) - E(Y_{i-1})) \ \underline{u}_i. \qquad (8.31)
\end{aligned}
$$

From (8.24) and (8.31) we get

$$
\begin{aligned}
m_{i,j} - m_{i,0} = (&\eta_1, m_{2,1} - m_{2,0} + \eta_2, \cdots, m_{j,j-1} - m_{j,0} + \eta_j, \\
&m_{j+1,j} - m_{j+1,0}, \cdots, m_{i-1,j} - m_{i-1,0}) \ \underline{u}_i.
\end{aligned}
$$

This holds for $j = 1, 2, \cdots, i-1$, so we have a system of $i-1$ equations of the form

$$T_i = Q_{i-1}\underline{u}_i, \qquad (8.32)$$

where

$$T_i = \begin{bmatrix} m_{i,1} - m_{i,0} \\ m_{i,2} - m_{i,0} \\ \vdots \\ m_{i,i-1} - m_{i,0} \end{bmatrix},$$

and

$$Q_{i-1} = \begin{bmatrix} \eta_1 & m_{2,1} - m_{2,0} & m_{3,1} - m_{3,0} & \cdots & m_{i-1,1} - m_{i-1,0} \\ \eta_1 & m_{2,1} - m_{2,0} + \eta_2 & m_{3,2} - m_{3,0} & \cdots & m_{i-1,2} - m_{i-1,0} \\ \eta_1 & m_{2,1} - m_{2,0} + \eta_2 & m_{3,2} - m_{3,0} + \eta_3 & \cdots & m_{i-1,3} - m_{i-1,0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta_1 & m_{2,1} - m_{2,0} + \eta_2 & m_{3,2} - m_{3,0} + \eta_3 & \cdots & m_{i-1,i-2} - m_{i-1,0} + \eta_{i-1} \end{bmatrix}.$$

Since $m_{i,j} - m_{i,j-1} = \theta_{i,j}$, $j = 1, 2, \cdots, i-1$, multiplying both sides of (8.32) from the left by the matrix

$$M_{i-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix},$$

gives

$$\begin{bmatrix} \theta_{i,1} \\ \theta_{i,2} \\ \vdots \\ \theta_{i,i-1} \end{bmatrix} = \begin{bmatrix} \eta_1 & \theta_{2,1} & \cdots & \theta_{i-1,1} \\ 0 & \eta_2 & \cdots & \theta_{i-1,2} \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \eta_{i-1} \end{bmatrix} \underline{u}_i.$$

Provided that

$$\eta_j \neq 0, \quad j = 1, 2, i-1,$$

the upper diagonal matrix $M_{i-1}Q_{i-1}$ is non-singular and hence

$$\underline{u}_i = \begin{bmatrix} \eta_1 & \theta_{2,1} & \cdots & \theta_{i-1,1} \\ 0 & \eta_2 & \cdots & \theta_{i-1,2} \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \eta_{i-1} \end{bmatrix}^{-1} \begin{bmatrix} \theta_{i,1} \\ \theta_{i,2} \\ \vdots \\ \theta_{i,i-1} \end{bmatrix}.$$

- Since

$$\text{Var}(Y_i|\underline{Y}_{i-1}) = \text{Var}(Y_i) - \underline{u}_i' V_{i-1} \underline{u}_i,$$

we can now use the assessed conditional variance given by $V_{i,i-1}$ in (8.27) to estimate the unconditional variance σ_i^2 as follows:

$$\sigma_i^2 = V_{i,i-1} + \underline{u}_i' V_{i-1} \underline{u}_i.$$

Using the Schurr complement, the matrix V_i is positive-definite if and only if

$$\sigma_i^2 - \underline{u}_i' V_{i-1} \underline{u}_i > 0,$$

which is guaranteed from (8.27) since $V_{i,i-1} > 0$.

- Choosing the arbitrary values $\eta_j^* \neq 0$, $j = 1, 2, \cdots, i-1$, guarantees the existence of a unique solution for $\underline{u}_i$. It can be seen from the relation

$$\eta_j = \begin{cases} \log\left(\dfrac{m_{1,0}^* + \eta_1^*}{1 - (m_{1,0}^* + \eta_1^*)}\right) - \log\left(\dfrac{m_{1,0}^*}{1 - m_{1,0}^*}\right), & \text{for } j = 1, \\ \log\left(\dfrac{m_{j,j-1}^* + \eta_j^*}{m_{1,1}^*}\right) - \log\left(\dfrac{m_{j,j-1}^*}{m_{1,1}^*}\right), & \text{for } j = 2, 3, \cdots, i-1, \end{cases}$$

that $\eta_j = 0$ if and only if $\eta_j^* = 0$, $j = 1, 2, \cdots, i-1$.

- So far, the proposed method estimates V_i as a positive-definite matrix, assuming that V_{i-1} is positive-definite. Since $V_1 > 0$, the method yields a positive-definite matrix V_{k-1}, by mathematical induction.

Estimating the last row and column of Σ_k

- Let Σ_k be partitioned as follows

$$\Sigma_k = \begin{bmatrix} V_{k-1} & V_{k-1}\underline{u}_k \\ \underline{u}_k' V_{k-1} & \sigma_k^2 \end{bmatrix}, \tag{8.33}$$

where

$$V_{k-1}\underline{u}_k = \text{Cov}(\underline{Y}_{k-1}, Y_k),$$

and

$$\sigma_k^2 = \text{Var}(Y_k).$$

Note that, according to the condition that elements of $\underline{p}$ must sum to one, the conditional variance of Y_k, given any specific value for $\underline{Y}_{k-1}$, has a fixed value of zero. Hence, using the standard theory of the multivariate normal distribution, we estimate σ_k^2 as

$$\sigma_k^2 = \underline{u}_k' V_{k-1} \underline{u}_k.$$

- To estimate $\underline{u}_k$ we write, as in (8.29),

$$E(Y_k|\underline{Y}_{k-1}) - E(Y_k) = [\underline{Y}_{k-1} - E(\underline{Y}_{k-1})]' \underline{u}_k. \tag{8.34}$$

Exploiting the condition that $\sum_{i=1}^{k} p_i = 1$, we can obtain $k-1$ estimates of $E(Y_k|\underline{Y}_{k-1})$ from $k-1$ different sets of conditioning values for $\underline{Y}_{k-1}$. More preciously, let

$$\begin{aligned}
m_{k,0} &= E[Y_k|Y_1 = m_{1,0}, Y_2 = m_{2,0}, \cdots, Y_{k-1} = m_{k-1,0}],\\
m_{k,1} &= E[Y_k|Y_1 = m_{1,1}, Y_2 = m_{2,0}, \cdots, Y_{k-1} = m_{k-1,0}],\\
m_{k,i} &= E[Y_k|Y_1 = m_{1,1}, Y_2 = m_{2,2}, \cdots, Y_{i-1} = m_{i-1,i-1}, Y_i = m_{i,i-1},\\
&\qquad Y_{i+1} = m_{i+1,i}, \cdots, Y_{k-2} = m_{k-2,k-3}, Y_{k-1} = m_{k-1,k-1}],\\
&\qquad\qquad \text{for } i = 2, 3, \cdots, k-2,\\
m_{k,k-1} &= E[Y_k|Y_1 = m_{1,1}, Y_2 = m_{2,2}, \cdots, Y_{k-2} = m_{k-2,k-2}, Y_{k-1} = m_{k-1,k-1}],
\end{aligned}$$

where $m_{k-1,k-1}$ is an arbitrary value, which will be chosen such that

$$m_{k-1,k-1} \neq m_{k-1,0}.$$

We require $m_{k-1,k-1} \neq m_{k-1,0}$ in order to solve the resulting system of equations, as will be shown later.

This gives the system of $k-1$ equations,

$$T_k = Q_{k-1} \underline{u}_k, \tag{8.35}$$

where

$$T_k = \begin{bmatrix} m_{k,1} - m_{k,0} \\ m_{k,2} - m_{k,0} \\ \vdots \\ m_{k,k-1} - m_{k,0} \end{bmatrix},$$

$$Q_{k-1} = \begin{bmatrix} \eta_1 & 0 & 0 & 0 & \cdots & 0 \\ \eta_1 & m'_{2,1} & m'_{3,2} & \cdots & m'_{k-2,k-3} & m'_{k-1,k-1} \\ \eta_1 & m'_{2,2} & m'_{3,2} & \cdots & m'_{k-2,k-3} & m'_{k-1,k-1} \\ \eta_1 & m'_{2,2} & m'_{3,3} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & m'_{k-2,k-3} & m'_{k-1,k-1} \\ \eta_1 & m'_{2,2} & m'_{3,3} & \cdots & m'_{k-2,k-2} & m'_{k-1,k-1} \end{bmatrix},$$

and

$$m'_{i,j} = m_{i,j} - m_{i,0}, \quad i = 2, 3, \cdots, k-1, \quad j = i-1, i.$$

We multiply both sides of (8.35) from the left by the matrix M_{k-1}, which has a different structure from M_{i-1} $(i < k)$, taking the form

$$M_{k-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{bmatrix}.$$

The system of equations can then be written as

$$\begin{bmatrix} m_{k,1} - m_{k,0} \\ m_{k,3} - m_{k,2} \\ \vdots \\ m_{k,k-1} - m_{k,k-2} \\ m_{k,0} - m_{k,k-1} \end{bmatrix} = \begin{bmatrix} \eta_1 & & & & \\ 0 & \eta_2 & & O & \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & 0 & \eta_{k-2} & \\ -\eta_1 & -m'_{2,2} & \cdots & -m'_{k-2,k-2} & \eta_{k-1} \end{bmatrix} \underline{u}_k, \qquad (8.36)$$

where

$$m'_{i,i} = m_{i,i} - m_{i,0}, \quad i = 2, 3, \cdots, k-2,$$

$$\eta_{k-1} = m_{k-1,0} - m_{k-1,k-1}.$$

Provided that

$$\eta_j \neq 0, \quad j = 1, 2, k-1,$$

the lower triangular matrix $M_{k-1}Q_{k-1}$ is non-singular and hence

$$\underline{u}_k = \begin{bmatrix} \eta_1 & & & & \\ 0 & \eta_2 & & O & \\ \vdots & \ddots & \ddots & & \\ 0 & \cdots & 0 & \eta_{k-2} & \\ -\eta_1 & -m'_{2,2} & \cdots & -m'_{k-2,k-2} & \eta_{k-1} \end{bmatrix}^{-1} \begin{bmatrix} m_{k,1} - m_{k,0} \\ m_{k,3} - m_{k,2} \\ \vdots \\ m_{k,k-1} - m_{k,k-2} \\ m_{k,0} - m_{k,k-1} \end{bmatrix}.$$

Positive-definiteness of the variance-covariance matrix

As mentioned before, the inverse of the additive logistic transformation is applied to the k dimension random vector $\underline{p}$, transforming it into the $k-1$ dimension random vector $\underline{Y}^* = (Y_2, \ Y_3, \ \cdots, \ Y_k)$. We are interested in the hyperparameter Σ_{k-1} as this is the variance-covariance matrix of $\underline{Y}^*$. Although the whole matrix Σ_k is clearly a singular matrix, we will show that the submatrix Σ_{k-1} is sure to be a positive-definite matrix, provided that no subset of categories has a known fixed sum of probabilities.

Consider the following partition of the singular multivariate normally distributed $\underline{Y}_k$:

$$\underline{Y}_k = \begin{bmatrix} Y_1 \\ \hdashline Y_2 \\ Y_3 \\ \vdots \\ Y_{k-1} \\ \hdashline Y_k \end{bmatrix} = \begin{bmatrix} \log(p_1) - \log(1-p_1) \\ \hdashline \log(p_2) - \log(p_1) \\ \log(p_3) - \log(p_1) \\ \vdots \\ \log(p_{k-1}) - \log(p_1) \\ \hdashline \log(p_k) - \log(p_1) \end{bmatrix} = \begin{bmatrix} Y_1 \\ \hdashline \underline{Y}^{**} \\ \hdashline Y_k \end{bmatrix}.$$

Recall that, by definition,

$$\underline{Y}^* = \left[\begin{array}{c} Y_2 \\ Y_3 \\ \vdots \\ Y_{k-1} \\ \hdashline Y_k \end{array}\right] = \left[\begin{array}{c} \underline{Y}^{**} \\ \hdashline Y_k \end{array}\right].$$

Let Σ_k be conformally partitioned as

$$\Sigma_k = \left[\begin{array}{c:c:c} V_1 & \underline{a}' & b \\ \hdashline \underline{a} & V^* & \underline{c} \\ \hdashline b & \underline{c}' & \sigma_k^2 \end{array}\right],$$

where

V^* is a $(k-2) \times (k-2)$ square matrix,

$\underline{a}$ and $\underline{c}$ are $(k-2) \times 1$ vectors,

V_1, σ_k^2 and b are scalars.

The method we used to estimate $V_{k-1} = \left[\begin{array}{c:c} V_1 & \underline{a}' \\ \hdashline \underline{a} & V^* \end{array}\right]$ guarantees its positive-definiteness, hence V^* is also positive-definite.

The matrix Σ_{k-1} is then partitioned as

$$\Sigma_{k-1} = \left[\begin{array}{c:c} V^* & \underline{c} \\ \hdashline \underline{c}' & \sigma_k^2 \end{array}\right].$$

For Σ_{k-1} to be positive-definite, we must show that

$$\sigma_k^2 > \underline{c}'(V^*)^{-1}\underline{c}.$$

In fact, using the inverse of a partitioned matrix, and for $d = V_1 - \underline{a}'(V^*)^{-1}\underline{a}$, we may write

$$\begin{aligned}
\sigma_k^2 &= \left[\begin{array}{c:c} b & \underline{c}' \end{array}\right] \left[\begin{array}{c:c} V_1 & \underline{a}' \\ \hdashline \underline{a} & V^* \end{array}\right]^{-1} \left[\begin{array}{c} b \\ \hdashline \underline{c} \end{array}\right] \\
&= \left[\begin{array}{c:c} b & \underline{c}' \end{array}\right] \left[\begin{array}{c:c} d^{-1} & -d^{-1}\underline{a}'(V^*)^{-1} \\ \hdashline -(V^*)^{-1}\underline{a}d^{-1} & (V^*)^{-1} + (V^*)^{-1}\underline{a}d^{-1}\underline{a}'(V^*)^{-1} \end{array}\right] \left[\begin{array}{c} b \\ \hdashline \underline{c} \end{array}\right] \\
&= \underline{c}'(V^*)^{-1}\underline{c} + \frac{1}{d}\left\{b^2 - 2b[\underline{a}'(V^*)^{-1}\underline{c}] + [\underline{c}'(V^*)^{-1}\underline{a}][\underline{a}'(V^*)^{-1}\underline{c}]\right\} \\
&= \underline{c}'(V^*)^{-1}\underline{c} + \frac{1}{d}\left[b - \underline{a}'(V^*)^{-1}\underline{c}\right]^2.
\end{aligned}$$

So, Σ_{k-1} is positive-definite if and only if

$$b - \underline{a}'(V^*)^{-1}\underline{c} \neq 0. \tag{8.37}$$

The method used to estimate Σ_k automatically guarantees the fulfilment of such a condition.

In fact, using the following partition of $\underline{u}_k$,

$$\underline{u}_k = \begin{bmatrix} u_1 \\ \hdashline u_2 \\ u_3 \\ \vdots \\ u_{k-1} \end{bmatrix} = \begin{bmatrix} u_1 \\ \hdashline \underline{u}_2 \end{bmatrix},$$

gives

$$\begin{aligned} \begin{bmatrix} b \\ \hdashline \underline{c} \end{bmatrix} &= V_{k-1}\ \underline{u}_k \\ &= \left[\begin{array}{c:c} V_1 & \underline{a}' \\ \hdashline \underline{a} & V^* \end{array}\right] \begin{bmatrix} u_1 \\ \hdashline \underline{u}_2 \end{bmatrix} \\ &= \begin{bmatrix} V_1\ u_1 + \underline{a}'\ \underline{u}_2 \\ \hdashline \underline{a}\ u_1 + V^*\ \underline{u}_2 \end{bmatrix}. \end{aligned}$$

Condition (8.37) thus holds if and only if

$$[V_1 - \underline{a}'(V^*)^{-1}\underline{a}]\ u_1 \neq 0.$$

But $V_1 - \underline{a}'(V^*)^{-1}\underline{a} > 0$ from the positive-definiteness of V_{k-1}, and hence Σ_{k-1} is positive-definite if and only if $u_1 \neq 0$.

It can be seen from (8.36) that

$$u_1 = \frac{m_{k,1} - m_{k,0}}{\eta_1}.$$

This condition is sure to be fulfilled since

$$m_{k,0} = \log\left[\frac{1 - \sum_{j=2}^{k-1} m^*_{j,0}}{m^*_{1,0}} - 1\right]$$

and

$$m_{k,1} = \log\left[\frac{1 - \sum_{j=2}^{k-1} m^*_{j,0}}{m^*_{1,1}} - 1\right],$$

from which

$$m_{k,1} \neq m_{k,0},$$

unless

$$m^*_{1,1} = m^*_{1,0},$$

which can never occur since

$$\eta_1^* \neq 0.$$

So, the proposed method for eliciting the matrix Σ_k ensures that Σ_{k-1} is positive-definite, even though Σ_k is itself singular.

Once $\underline{\mu}_k$ and Σ_k have been estimated, equations (8.6)-(8.8) give the hyperparameters $\underline{\mu}_{k-1}$ and Σ_{k-1} of the logistic normal prior distribution of $\underline{p}$ based on the normalizing transformations given by $\underline{Y}^*$.

8.5 Feedback using marginal quartiles of the logistic normal prior

After eliciting the mean vector $\underline{\mu}_{k-1}$ and the variance-covariance matrix Σ_{k-1} of $\underline{Y}^*$, the software calculates marginal medians and quartiles of the probability of each category and displays their values as feedback to the expert. Since the initially assessed quartiles were all conditional, it is useful to inform the expert of the marginal quartiles and give her the option of changing them if she wants.

To add this feedback option to the software, we had to develop a reliable technique for estimating marginal quartiles from the elicited hyperparameters $\underline{\mu}_{k-1}$ and Σ_{k-1}. Moreover, we must correspondingly modify the elicited hyperparameters once the marginal quartiles have been changed by the expert during the feedback stage.

A simple direct method for estimating the marginal moments, or quartiles, of the logistic normal distribution in closed forms does not seem to exist in the literature. Aitchison (1986) suggested using Hermitian numerical integration methods to obtain marginal moments. However, he argued that the main practical interest is in the ratio of components, not in the component themselves. This is not the case here, as we are mainly interested in marginal probabilities, not in their ratios. Another approach, based on the Gibbs sampling technique, has been used by Forster and Skene (1994) to accurately approximate the posterior marginal densities and other summaries for a broad class of prior distributions including the Dirichlet and logistic normal distributions. However, the method approximates the marginal densities of the posterior distribution rather than the prior distribution.

Under the normality assumption of $\underline{Y}^*$ and the unit sum constraint, it has been proved in Theorem 8.1 that the marginal unconditional medians of p_j, m_j, are equal to their conditional medians, $m^*_{j,0}$, for $j = 1, 2, \cdots, k$.

Moreover, the same assumptions make it possible to estimate marginal lower and upper quartiles for each p_j, for $j = 1, 2, \cdots, k$. In the following lemma we formally state and prove the above results. Then, we propose a method of revising the estimates of $\underline{\mu}_{k-1}$ and Σ_{k-1} to reflect any change made by the expert to the marginal quartiles.

Lemma 8.4. *For any $j = 1, 2, \cdots, k$, under the assumptions of Theorem 8.1,*

$$V_j = 2\left\{\log\left[\sum_{i\neq j}^{k}\exp\left(\mu_{i,j}+\frac{1}{2}\sigma^2_{i,j}\right)\right]+E_j\right\},$$

and V_j is guaranteed to be strictly greater than zero.

Proof

Since

$$Y_{i,j} = \log\left(\frac{p_i}{p_j}\right) \sim \mathrm{N}(\mu_{i,j}, \sigma^2_{i,j}), \qquad i, j = 1, 2, \cdots, k, \quad i \neq j,$$

with known $\mu_{i,j}, \sigma^2_{i,j}$, the expected value of the lognormal distribution of (p_i/p_j) is given by

$$E\left(\frac{p_i}{p_j}\right) = \exp\left(\mu_{i,j}+\frac{1}{2}\sigma^2_{i,j}\right).$$

But

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

so

$$\begin{aligned} M_j \equiv E\left(\frac{1-p_j}{p_j}\right) &= E\left(\sum_{i\neq j}^{k}\frac{p_i}{p_j}\right) \\ &= \sum_{i\neq j}^{k} E\left(\frac{p_i}{p_j}\right) = \sum_{i\neq j}^{k}\exp\left(\mu_{i,j}+\frac{1}{2}\sigma^2_{i,j}\right). \end{aligned} \tag{8.38}$$

On the other hand, by the assumption of approximate normality for $Y_{j,j}$, we have

$$\log\left(\frac{p_j}{1-p_j}\right) \sim \mathrm{N}\left(E_j, V_j\right),$$

so

$$\log\left(\frac{1-p_j}{p_j}\right) \sim \mathrm{N}\left(-E_j, V_j\right),$$

and

$$M_j = E\left(\frac{1-p_j}{p_j}\right) = \exp\left(-E_j + \frac{1}{2}V_j\right). \tag{8.39}$$

We take M_j as in (8.38), and Theorem 8.1 gives

$$E_j = \log\left(\frac{m^*_{j,0}}{1-m^*_{j,0}}\right). \tag{8.40}$$

Equation (8.39) can be solved for V_j to give the first statement of Lemma 8.4.

Substituting $m^*_{j,0}$ for $M(p_j)$ in equation (8.16) and putting $m_{i,(j)} = \mu_{i,j}$, gives

$$E_j = -\log\left(\sum_{i\neq j}^{k} \exp(\mu_{i,j})\right). \tag{8.41}$$

This guarantees that $V_j > 0$ in (8.39), since by comparing the RHSs of (8.38) and (8.41), we can see clearly that

$$M_j > \exp\left(-E_j\right).$$

This ends the proof of Lemma 8.4.

The two unconditional quartiles of p_j can be obtained from

$$Q_1(p_j) = \frac{\exp[Q_1(Y_{j,j})]}{1+\exp[Q_1(Y_{j,j})]},$$

and

$$Q_3(p_j) = \frac{\exp[Q_3(Y_{j,j})]}{1+\exp[Q_3(Y_{j,j})]},$$

with

$$Q_1(Y_{j,j}) = E_j + \sqrt{V_j}\,\Phi^{-1}(0.25),$$

$$Q_3(Y_{j,j}) = E_j + \sqrt{V_j}\,\Phi^{-1}(0.75),$$

where Φ is the cdf of the standard normal distribution.

The unconditional quartiles $Q_1(p_j)$ and $Q_3(p_j)$ are presented to the expert as feedback with the unconditional median $M(p_j)$, for $j = 1, 2, \cdots, k$. The expert has the option of changing any of the unconditional medians and/or quartiles. The changes are reflected in estimates of the hyperparameters $\underline{\mu}_{k-1}$ and Σ_{k-1}, using the following approach.

- Let $m'(p_j)$ denote the values of $M(p_j)$ after re-assessment $(j = 1, 2, \cdots, k)$. We revise $\mu_{j,1}$ to

$$\mu^*_{j,1} = E^*(Y_{j,1}) = \begin{cases} \log(m^*(p_j)) - \log(1 - m^*(p_j)) & \text{for } j = 1, \\ \log(m^*(p_j)) - \log(m^*(p_1)) & \text{for } j = 2, \cdots, k, \end{cases}$$

  with a new normalized set of medians $m^*(p_j)$, where

$$m^*(p_j) = \frac{m'(p_j)}{\sum_{i=1}^{k} m'(p_j)}, \qquad j = 1, 2, \cdots, k,$$

- Suppose one or more of the marginal unconditional quartiles $Q_1(p_j)$ and/or $Q_3(p_j)$ are re-assessed as $Q'_1(p_j)$ and/or $Q'_3(p_j)$, respectively, for $j = 1, \cdots, k$. Then we change the variance-covariance matrix $\Sigma_{(1)}$ to

$$\Sigma^*_{(1)} = \mathrm{Var}^*(\underline{Y}_{(1)}) = D^{\frac{1}{2}}\ \Sigma_{(1)}\ D^{\frac{1}{2}}, \tag{8.42}$$

  where D is a diagonal matrix with diagonal elements

$$d_i = \frac{\sigma^{2*}_{i,1}}{\sigma^2_{i,1}}, \qquad i = 2, 3, \cdots, k,$$

  and $\sigma^{2*}_{i,1}$ is defined by

$$\begin{aligned} \sigma^{2*}_{i,1} &= \mathrm{Var}^*(\log(p_i) - \log(p_1)) \\ &= \mathrm{Var}^*(\log(p_i)) + \mathrm{Var}^*(\log(p_1)) - 2\mathrm{Cov}^*(\log(p_i), \log(p_1)). \end{aligned} \tag{8.43}$$

The modified variances and covariances, Var* and Cov*, respectively, are determined as follows.

As $Y_{j,j}$ is assumed to have an approximate normal distribution, let

$$V_j^* = \text{Var}\left[\log\left(\frac{p_j}{1-p_j}\right)\right] = \left[\frac{\log\left(\frac{Q_3'(p_j)}{1-Q_3'(p_j)}\right) - \log\left(\frac{Q_1'(p_j)}{1-Q_1'(p_j)}\right)}{1.349}\right]^2,$$

so,

$$Y_{j,j} \sim N(E_j, V_j^*).$$

Using a simple numerical integration technique on the normal pdf of $Y_{j,j}$, we can get the expectations, for $j = 1, 2, \cdots, k$, in the RHS of the following equation,

$$\text{Var}^*(\log(p_j)) = E\left\{\log\left[\frac{\exp(Y_{j,j})}{1+\exp(Y_{j,j})}\right]\right\}^2 - E^2\left\{\log\left[\frac{\exp(Y_{j,j})}{1+\exp(Y_{j,j})}\right]\right\}.$$

To attain a strictly positive value of $\sigma_{i,1}^{2*}$ as in (8.43), we modify $\text{Cov}(\log(p_i), \log(p_1))$ by putting

$$\text{Cov}^*(\log(p_i), \log(p_1)) = w_i \text{Cov}(\log(p_i), \log(p_1)) \qquad i = 2, 3, \cdots, k.$$

where

$$w_i = \sqrt{\frac{\text{Var}^*(\log(p_i))\text{Var}^*(\log(p_1))}{\text{Var}(\log(p_i))\text{Var}(\log(p_1))}} \qquad i = 2, 3, \cdots, k.$$

In (8.42) we use the diagonal matrix,

$$D = \left(\frac{\sigma_{i,1}^{2*}}{\sigma_{i,1}^2}\right),$$

so as to change the variances of $\underline{Y}_{(1)}$, while preserving correlations and also preserving the positive-definiteness for $\Sigma_{(1)}^*$.

Another feedback window is available on request for the expert, should she need to see the influence of changing one or more of the marginal quartile values. If this option is taken and further re-assessment made, then the method given in Lemma 8.4 is applied again on the modified matrix $\Sigma_{(1)}^*$, to give a new set of marginal quartiles. These can be changed again by the expert if she does not find it a satisfactory representation of her opinion.

We should mention that the new set of marginal quartiles does not necessarily have the same values as the modified quartiles. The unit sum condition of $\underline{p}$, with the normality assumption of each $Y_{j,j}$, for $j = 1, 2, \cdots, k$, always forces the marginal interquartile range for a single probability to partly depend on the other probabilities, as shown in Lemma 8.4. Hence, for mathematical coherence, the resulting set of marginal quartiles will not correspond exactly to the expert's assessments. The proposed approach that uses Lemma 8.4 and continuous feedback enables the expert to adjust the quartiles until she is happy with the feedback values.

8.6 Example: Transport preferences

In designing transport systems for the future, one ingredient is the relative importance of factors a person may consider in selecting the mode of transport for different journeys. Estimates of these preferences help in planning rail services, roads and other transport infrastructure. Such estimates are also of interest from the environmental point of view, because of the impact of transport emissions.

For a preparatory environmental study, estimates about factors affecting transport preferences in 2020 were needed. In this example, a transport expert quantified his opinion about the factors affecting the choice of transport for a hundred mile journey across UK in that year. Primary interests of the expert (Dr. James Warren, The Open University) include modelling energy and emissions to gain a better understanding of transport systems and the potential effects of transportation policy and technology on the environment. He specified five quantities as the main factors a passenger would consider in choosing the means of transport for such a journey. These factors are: cost, journey time, environmental impact, comfort, and convenience. Interest focuses on the relative frequency with which each of these quantities is the most important factor: For what proportion of people would cost be the most important factor in choosing the mode of transport for the journey? For what proportion would it be journey time? And so on. The problem can thus be described as a multinomial model with five categories, one for each factor. Our method and PEGS-Logistic software were used by the expert to quantify his opinion about a logistic normal prior for the parameters of this multinomial model.

After initializing the software and defining the model, the expert assessed his medians of the proportion of people for whom Cost/ Time/ EcoImpact/ Comfort/ Convenience would be the most important factor. These medians assessments were 0.61, 0.25, 0.04, 0.06, 0.10, respectively, and they are the blue bars in Figure 8.4. These values do not sum to 1 and the software suggests values (yellow bars) that did. Rather than accepting these suggestions, the expert revised his initial median assessments to be 0.49, 0.28, 0.04, 0.06, 0.11, respectively. As their sum is nearly equal to one, the medians suggested next were very close to his assessments and the expert accepted them as representatives of his opinions.

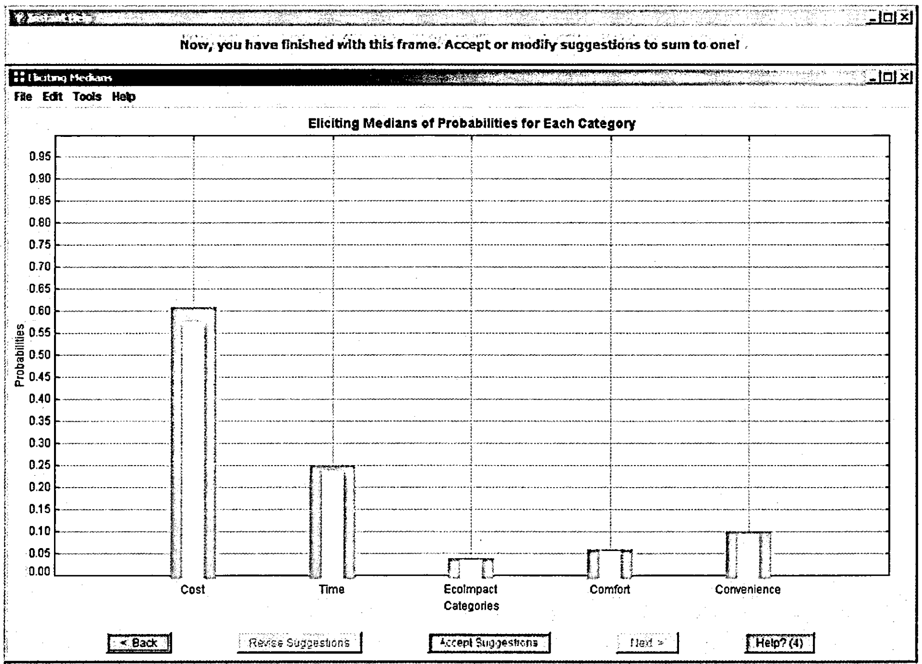


Figure 8.4: Software suggestions for initial medians

The expert then gave his assessed upper and lower quartile values for the probability of the first category; these were 0.62 and 0.43 respectively. Then conditioning on his assessed medians for previous categories, he assessed his conditional quartile values. The four conditional lower quartiles were 0.18, 0.03, 0.03, 0.10, respectively, while the four conditional upper quartiles were 0.36, 0.10, 0.08, 0.15, respectively. See Figure 8.5, in which the expert has given his two quartiles of the fourth category conditional on the probabilities of the first three categories. The quartiles of the last category follow automatically. Although the expert is not a statistician, he had no problems in assessing quartiles after a brief discussion about the method of bisection.

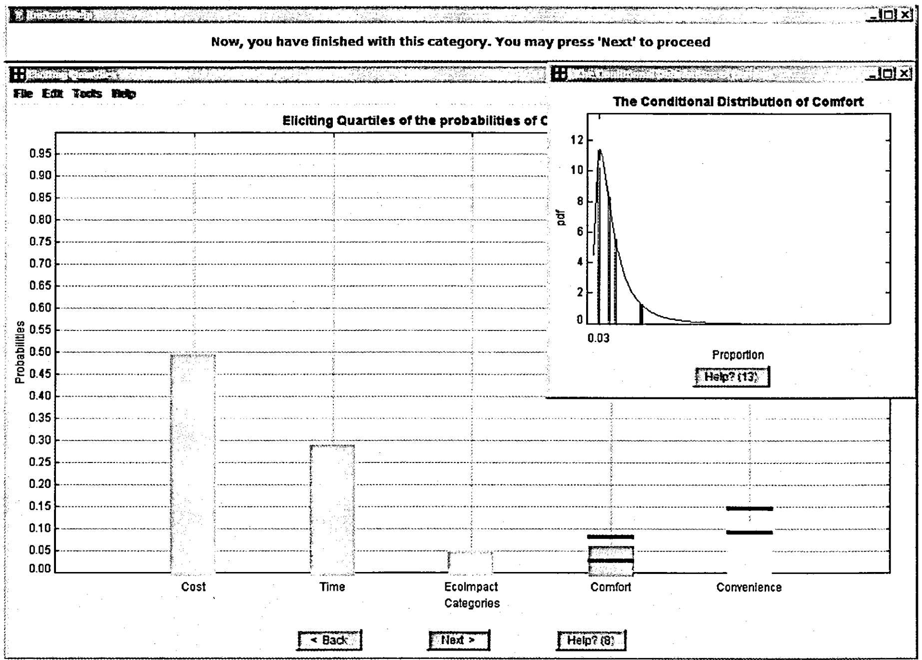


Figure 8.5: Assessing conditional quartiles

Next, the expert gave conditional median assessments of 0.41, 0.16, 0.12, 0.33 for the remaining four categories, conditional on the probability of the first category being 0.25. The number of conditions was then increased in stages. Conditional on 0.25 and 0.20 being the probabilities for the first and second categories, respectively, the expert revised his probability median assessments for the last three categories to 0.13, 0.18 and 0.25, respectively. See Figure 8.6. Finally, he gave the conditional medians of 0.19, 0.30 for the last two categories given that the probabilities of the first three categories were 0.25, 0.20 and 0.07, respectively.

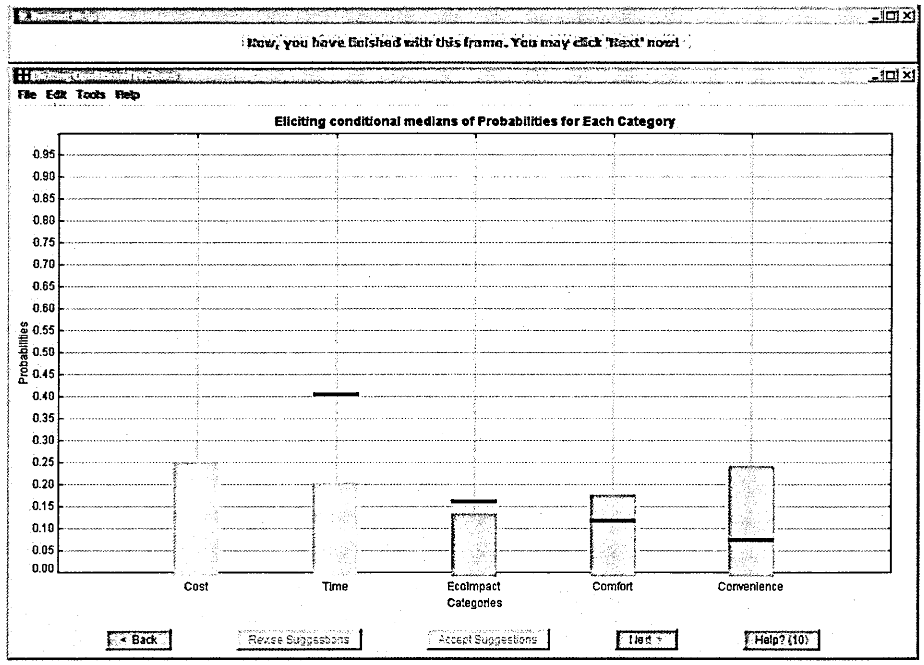


Figure 8.6: Revised conditional medians

It is worth mentioning that the suggestions given by the software played a crucial role in helping the expert choose medians that satisfy the unit sum constraint. During the elicitation process, obviously the sums of expert's assessments never equalled one exactly. When suggestions were offered by the software, he normally revised one assessment and then accepted the second round of offered suggestions. After making his conditional median assessments, the expert was then shown the unconditional medians and unconditional quartiles that were implied by all his assessments. See Figure 8.7. During this feedback stage he was invited to accept or revise these quantities. The unconditional medians that were offered were accepted by the expert as an adequate representation of his opinion. However, he decided to use the change quartiles button to revise the unconditional quartiles and then reduced the interquartile range of the last category.

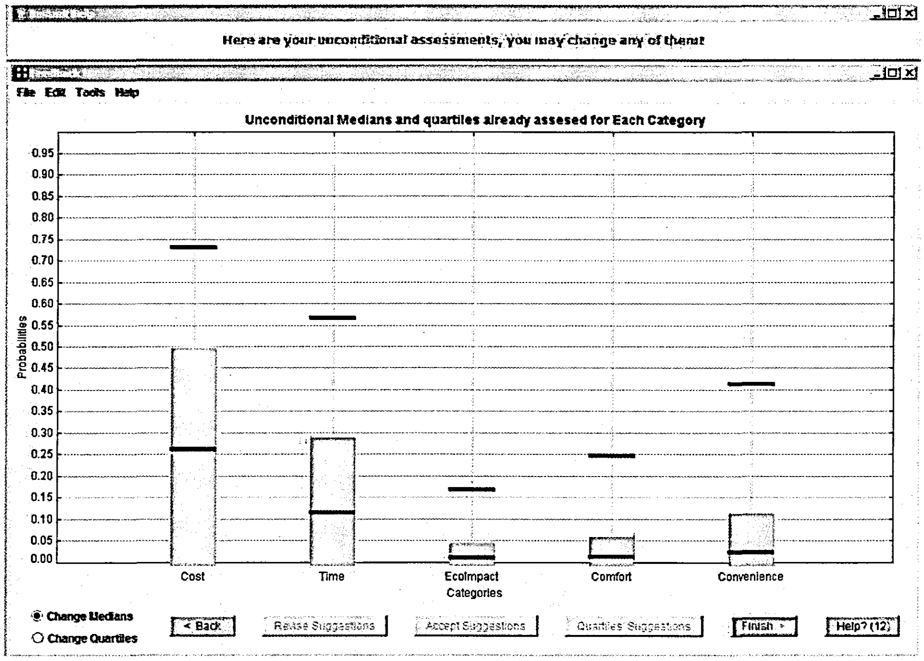


Figure 8.7: Software suggestions for marginal medians and quartiles

The elicitation process took about 20 minutes to complete. The expert commented that although the elicitation problem was quite tricky, the software gave a helpful form of visualization. He also mentioned that he had found it hard to make his median assessments sum to one, so that the software's suggestions had been very welcome. He also advised that it would be helpful if the different categories were ordered according to their importance, i.e. in a descending order according to their median probability values. He thought that this order would make it easier for him to think about conditional assessments.

The software output the following elicited hyperparameters of the logistic normal prior as in Tables 8.1 and 8.2.

Table 8.1: The elicited mean vector of a logistic normal prior

$Y_2 = \log(p_2/p_1)$	$Y_3 = \log(p_3/p_1)$	$Y_4 = \log(p_4/p_1)$	$Y_5 = \log(p_5/p_1)$
-0.5058	-2.4517	-2.0639	-1.5043

Table 8.2: The elicited variance-covariance matrix of a logistic normal prior

	$Y_2 = \log(\frac{p_2}{p_1})$	$Y_3 = \log(\frac{p_3}{p_1})$	$Y_4 = \log(\frac{p_4}{p_1})$	$Y_5 = \log(\frac{p_5}{p_1})$
$Y_2 = \log(\frac{p_2}{p_1})$	0.3414	0.1511	0.1598	-0.3035
$Y_3 = \log(\frac{p_3}{p_1})$	0.1511	0.9087	0.3677	-0.5551
$Y_4 = \log(\frac{p_4}{p_1})$	0.1598	0.3677	1.0906	-1.9076
$Y_5 = \log(\frac{p_5}{p_1})$	-0.3035	-0.5551	-1.9076	3.468

This output gives the mean vector and variance-covariance matrix of a multivariate normal distribution of degree 4 for Y_2, Y_3, Y_4, Y_5. However, the marginal moments of each p_i are not given as output. Instead, marginal medians and quartiles are presented to the expert during the feedback stage as discussed before, see Figure 8.7. The multivariate normal distribution of Y_2, Y_3, Y_4, Y_5 may be used as a prior distribution in a Bayesian analysis. Details of the additive logistic transformations are also needed:

$$p_i = \begin{cases} \dfrac{1}{1 + \sum_{j=2}^{5} \exp(Y_j)}, & \text{for } i = 1, \\[2ex] \dfrac{\exp(Y_i)}{1 + \sum_{j=2}^{5} \exp(Y_j)}, & \text{for } i = 2, 3, \cdots, 5. \end{cases}$$

Of course, the extra variable Y_1 is omitted as it is a redundant variable due to the unit sum constraint on $\underline{p}$. The software has an option to implement this prior distribution in a WinBUGS file. After the sample data are obtained, the software produces a file for a WinBUGS model that contains sample data, a multinomial likelihood and a complete specification of the logistic normal prior distribution that the expert assessed.

8.7 Concluding comments

In Chapters 6 and 7, we introduced elicitation methods for Dirichlet, generalized Dirichlet and Gaussian copula as prior distributions for the parameter vector $\underline{p}$ of the multinomial model. Hence the logistic normal distribution is our fourth suggested prior distribution for this model. Among these priors, the logistic normal prior gives the most general correlation structure. The PEGS-Multinomial software, that is freely available at http://statistics.open.ac.uk/elicitation, offers the option of eliciting any of these four prior distributions.

As noted earlier, it is tricky to elicit assessments that satisfy all the necessary requirements for multinomial models. For example, if there are only two categories, the lower probability quartile of one category and the upper quartile of the other must add up to one. As the number of categories increases the requirements that must be satisfied increases. In our proposed elicitation method, we chose assessment tasks and a structure that led to a coherent set of assessments, without the expert having to be conscious of the requirements.

Chapter 9

Eliciting multinomial models with covariates

9.1 Introduction

With multinomial models, the membership probabilities of different categories may depend on one or more continuous or categorical explanatory variables (covariates) that influence these probabilities. The simpler well-known example in this context is the logistic regression, where the probability of being in one of only two categories is related to a set of explanatory variables through the logit link function.

Suppose there are k categories, let $p_1, p_2, \cdots, p_k$ denote the membership probabilities and let $\underline{X} = (X_1, X_2, \cdots, X_m)$ be a vector of m explanatory variables. Relating $\underline{X}$ to each probability p_i using separate logit link functions is not the best choice. The inverse link functions gives

$$p_i(\underline{X}) = \frac{\exp(\alpha_i + \underline{X}'\underline{\beta}_i)}{1 + \exp(\alpha_i + \underline{X}'\underline{\beta}_i)}, \quad i = 1, 2, \cdots, k, \tag{9.1}$$

in which case, it will not be easy to investigate the conditions under which the constraint $\sum_{i=1}^{k} p_i(\underline{X}) = 1$ is fulfilled. Some other link functions are available in the literature [e.g. Aitchison (1986)]. However, the additive multinomial logistic link function is the most convenient, as it automatically accounts for the unit sum constraint. It links the classification probabilities to linear predictors in the form,

$$p_i(\underline{X}) = \begin{cases} \dfrac{1}{1 + \sum\limits_{j=2}^{k} \exp(\alpha_j + \underline{X}'\underline{\beta}_j)}, & i = 1, \\[2ex] \dfrac{\exp(\alpha_i + \underline{X}'\underline{\beta}_i)}{1 + \sum\limits_{j=2}^{k} \exp(\alpha_j + \underline{X}'\underline{\beta}_j)}, & i = 2, \cdots, k. \end{cases} \tag{9.2}$$

Expressing the model in the form of (9.2) helps to generalize results obtained in the previous chapter to the current case.

For the Bayesian analysis of the multinomial logit model, a multivariate normal prior may be assumed [e.g O'Hagan and Forster (2004)] for the parameter vector

$$\underline{\beta}^* = (\alpha_2, \ \underline{\beta}_2', \ \cdots, \ \alpha_k, \ \underline{\beta}_k')',$$

where the vectors of coefficients, $(\alpha_i, \underline{\beta}_i')'$, are category specific, for $i = 2, \cdots, k$, i.e. each category has its own vector of regression coefficients. We select the first category as the fill-up category, hence, its regression coefficients, $(\alpha_1, \underline{\beta}_1')'$, are not included in the prior distribution for identifiability.

In this chapter we propose an elicitation method for eliciting a mean vector and a positive-definite variance-covariance matrix of the normal prior distribution of $\underline{\beta}^*$. Our proposed method is based on the results obtained in the previous chapter for the logistic normal prior distribution of the multinomial model. The proposed method has been implemented in the PEGS-Multinomial with Covariates software that is freely available at http://statistics.open.ac.uk/elicitation.

In Section 9.2, we define the underlying model, namely, the base-line multinomial logit model, in terms of the additive logistic transformation. The required assumptions, notation and theoretical framework are discussed in Section 9.3. Elicitation methods and assessment tasks required for eliciting a mean vector and a positive-definite variance covariance matrix for the regression coefficients are proposed in Sections 9.4 and 9.5. Final concluding comments of this chapter are given in Section 9.6.

9.2 The base-line multinomial logit model

The model that uses the link function in (9.2) is known as the multinomial logistic (logit) model, since it has multinomial responses with a number of $k > 2$ categories. The model in (9.2) is usually given in the more general form

$$p_i(\underline{X}) = \frac{\exp(\alpha_i + \underline{X}'\underline{\beta}_i)}{\sum_{j=1}^{k} \exp(\alpha_j + \underline{X}'\underline{\beta}_j)}, \quad i = 1, 2, \cdots, k, \tag{9.3}$$

which is called the base-line multinomial logit model. See, for example, Agresti (2002) or Powers and Xie (2000). In the rest of this chapter, for ease of notation, each classification probability $p_i(\underline{X})$, as defined in (9.2), will be just denoted by p_i, for $i = 1, 2, \cdots, k$.

To attain the unit sum constraint in the base-line model, an identifiability constraint must be imposed by equating the coefficients of the "base-line" category to zeros. The selection of the base-line category is arbitrary. If we select the first category as the base-line category, then, under the identifiability constraint $(\alpha_1,\ \underline{\beta}_1')' = \underline{0}$, it can easily be shown that the model in (9.3) is equivalent to that in (9.2). Thus, the model has exactly $(k-1)(m+1)$ free parameters.

From (9.2), the linear predictor, $Y_j = \alpha_j + \underline{X}'\underline{\beta}_j$, can be written in terms of the logistic transformations in classification probabilities as

$$Y_j = \alpha_j + \underline{X}'\underline{\beta}_j = \log(p_j) - \log(p_1), \qquad \text{for } j = 2, 3, \cdots, k, \tag{9.4}$$

where the regression coefficients for the jth category are

$$\underline{\beta}_j' = (\beta_{1,j},\ \ \beta_{2,j},\ \ \cdots,\ \ \beta_{m,j}).$$

We define an extra variable, Y_1, as

$$Y_1 = \log(p_1) - \log(1 - p_1). \tag{9.5}$$

This extra variable is required to be used as a conditioning value in the elicitation process, as shown in the previous chapter. We do not assume Y_1 to be a linear predictor, since the trivial parameters, α_1 and $\underline{\beta}_1$, will not appear in the elicited prior distribution. We adopt the conventions $\alpha_1 = 0$, $\underline{\beta}_1 = \underline{0}$, for identifiability of the base-line model.

9.3 Notation and theoretical framework

We assume that the prior opinion about the linear predictors $Y_2, \cdots, Y_k$, can be adequately represented by a multivariate normal distribution of degree $k-1$. Then from equations (9.4), (9.5) and Section 8.2.1, Y_1 has an approximate normal distribution. In addition, the classification probabilities, $p_1, p_2, \cdots, p_k$, have a logistic normal distribution as defined in Section 8.2. Following O'Hagan and Forster (2004), we assume a multivariate normal prior distribution for the regression coefficients.

For tractability in the elicitation process, the expert is asked to give her assessments for the classification probabilities, $p_1, \cdots, p_k$, and consequently for $Y_2, \cdots, Y_k$, for only one covariate at a time. All other covariates are assumed to be at their reference values/levels. By doing this for each covariate in turn, the expert can concentrate on revising her assessments as a result of the change in just one explanatory covariate.

The relationship between each Y_j and each continuous covariate X_r is not necessarily linear. A piecewise-linear relationship as discussed in Chapters 3 and 4 might be a reasonable choice here that can model many types of relationships. However, in dealing with k categories and m explanatory covariates, a piecewise-linear relationship will seldom be practical as it imposes a large number of dividing points (knots) at which the expert must give assessments. This would lead to a lengthy elicitation process. So, to simplify the elicitation process, we assume that relationships are linear. Specifically, we assume a linear relationship between each continuous covariate X_r, $r = 1, 2, \cdots, m$, and each Y_j, $j = 2, \cdots, k$, of the form

$$Y_j = \alpha_j + X_r \beta_{r,j}, \qquad r = 1, \cdots, m, \quad j = 2, \cdots, k, \tag{9.6}$$

given that all other covariates are fixed at their reference values/levels. That is, equation (9.6) holds when $X_i = x_{i,0}$, for $i = 1, 2, \cdots, m$, $i \neq r$, where $x_{i,0}$ is the reference value/level of X_i. If, all covariates are at their reference values/levels, i.e. $X_i = x_{i,0}$, for $i = 1, 2, \cdots, m$, then

$$Y_j = \alpha_j, \qquad j = 2, \cdots, k. \tag{9.7}$$

To achieve this, for $r = 1, 2, \cdots, m$, if the covariate X_r is a factor (categorical variable), with a reference level $x_{r,0}$ and any number $\delta(r)$ of levels, $x_{r,1}, x_{r,2}, \cdots, x_{r,\delta(r)}$, then X_r is split into $\delta(r)$ new factors, $X_{r,i}$ defined as

$$X_{r,i} = \begin{cases} 1 & \text{if } X_r = x_{r,i} \\ 0 & \text{otherwise,} \end{cases} \tag{9.8}$$

for $i = 1, 2, \cdots, \delta(r)$.

If X_r is a continuous covariate with a reference value $x_{r,0}$, then we define a new variable X_r^* as

$$X_r^* = X_r - x_{r,0}, \qquad \text{for } r = 1, 2, \cdots, m. \tag{9.9}$$

With the new covariates defined by (9.8) and (9.9), the value of each covariate is equal to zero at its reference value.

Hence, if m consists of m_1 factors and m_2 continuous covariates, we get a new set of, say, m^* explanatory variables, where

$$m^* = \sum_{j=1}^{m_1} \delta(j) + m_2.$$

To simplify the notation, with no loss of generality, we keep the notation $X_1, X_2, \cdots, X_m$, for the set of covariates, while keeping in mind that m actually denotes m^* and that each X_r is of the form of (9.8) for a factor or (9.9) for a continuous covariate. In this sense, the models in (9.6) and (9.7) are equivalent to (9.4).

It is convenient to rearrange the regression coefficients into a matrix, say $\boldsymbol{\beta}$, of the form

$$\boldsymbol{\beta} = \left[\begin{pmatrix} \alpha_1 \\ \\ \underline{\beta}_1 \end{pmatrix}, \begin{pmatrix} \alpha_2 \\ \\ \underline{\beta}_2 \end{pmatrix}, \cdots, \begin{pmatrix} \alpha_k \\ \\ \underline{\beta}_k \end{pmatrix} \right]. \tag{9.10}$$

Then we define the new set of vectors $\underline{\alpha}$, $\underline{\beta}_{(r)}$, for $r = 1, 2, \cdots, m$, as the rows of $\boldsymbol{\beta}$, of the form

$$\underline{\alpha} = (\alpha_1, \ \ \alpha_2, \ \ \cdots, \ \ \alpha_k)', \tag{9.11}$$

$$\underline{\beta}_{(r)} = (\beta_{r,1}, \ \ \beta_{r,2}, \ \ \cdots, \ \ \beta_{r,k})', \tag{9.12}$$

and the same set with the first zero elements removed, as

$$\underline{\alpha}^1 = (\alpha_2, \ \ \alpha_3, \ \ \cdots, \ \ \alpha_k)', \tag{9.13}$$

$$\underline{\beta}^1_{(r)} = (\beta_{r,2}, \ \ \beta_{r,3}, \ \ \cdots, \ \ \beta_{r,k})'. \tag{9.14}$$

Since each column of the $\boldsymbol{\beta}$ matrix in (9.10) contains regression coefficients that correspond to one category, it is more convenient to work with the rows, which each correspond to one covariate. In this case, elements of a single row correspond to classification probabilities, and hence these elements must be inter-related in a way that reflects the unit sum constraint of the probabilities. Therefore, we assume that the elements of $\underline{\alpha}$ are correlated, and that the elements of each $\underline{\beta}_{(r)}$ are also correlated, hence statistically dependent, *a priori*, for all $r = 1, 2, \cdots, m$. While elements from different rows of $\boldsymbol{\beta}$, that corresponds to different covariates, are assumed to be independent *a priori*, so as to simplify the elicitation process and obtain a block-diagonal variance-covariance matrix.

If we let $\underline{\beta}^1 = (\underline{\beta}^{1'}_{(1)}, \ \underline{\beta}^{1'}_{(2)}, \ \cdots, \ \underline{\beta}^{1'}_{(m)})'$, then the multivariate normal prior distribution to be elicited is thus of the form,

$$\begin{pmatrix} \underline{\alpha}^1 \\ \\ \underline{\beta}^1 \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} \underline{\mu}_\alpha \\ \\ \underline{\mu}_\beta \end{pmatrix}, \begin{pmatrix} \Sigma_\alpha & \Sigma_{\alpha,\beta} \\ \\ \Sigma'_{\alpha,\beta} & \Sigma_\beta \end{pmatrix} \right). \tag{9.15}$$

9.4 Eliciting the mean vector

To elicit the mean vectors $\underline{\mu}_\alpha$ and $\underline{\mu}_\beta$, in (9.15), we proceed as follows

- The expert is asked to assume that all covariates are at their reference values/levels, i.e. $X_r = 0$, $r = 1, 2, \cdots, m$. We call this situation as the reference point. She then assesses a median value, say $m^*_{1,0,0}$, for the probability p_1 of the first category. As discussed in the previous chapter, since the choice of the first category is arbitrary, it is chosen by the expert as the most common category. Then the expert assesses median values $m^*_{j,0,0}$, $j = 2, \cdots, k$, for all the remaining categories.

- As proved in Theorem 8.1 in the previous chapter, these unconditional median assessments are equal to the conditional medians of $(p_j|p_1 = m^*_{1,0,0})$ for $j = 2, 3, \cdots, k$. For convenience, we denote both conditional and unconditional medians by $m^*_{j,0,0}$, $j = 2, \cdots, k$. Lemma 8.1 in the previous chapter states that median assessments must sum to one, so they are normalized by the PEGS-Multinomial with Covariates software to fulfill this condition.

- For each covariate in turn, the expert is asked to assume a specific value of the current covariate, say $X_r = x_r$, while all other covariates are assumed to be at their reference values/levels. Under these assumptions the expert starts by assessing a median value for p_1, say $m^*_{1,0,r}$. Then she assesses a new set of median values, say $m^*_{j,0,r}$, for $j = 2, 3, \cdots, k$, for all the remaining categories. Again, these assessments are normalized to satisfy the unit sum constraint. This process is repeated for each covariate, i.e. for $r = 1, 2, \cdots, m$.

- Figure 9.1 shows the assessed probability medians when only one of the covariates, age, has changed from its reference value to a new value (40 years). To help the expert during this stage, the software gives the previously assessed medians when all covariates were at their reference values/levels. This is presented by the upper right graph of Figure 9.1. The reference value/level of each continuous covariate/factor is also listed in the upper left table as in Figure 9.1.

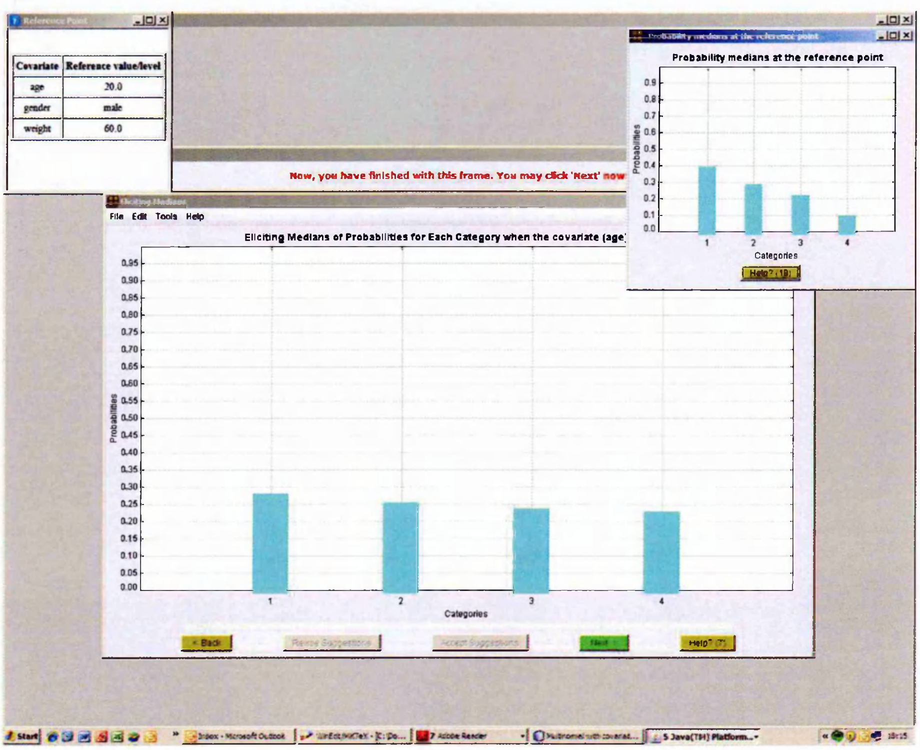


Figure 9.1: Assessing probability medians at age = 40 years

Now, let the conditional median of Y_j, given that all covariates are at their reference levels, be denoted by $m_{j,0,0}$, for $j = 1, 2, \cdots, k$. Also, let the conditional median of Y_j, given that $X_r = x_r$ and all other covariates are at their reference levels, be denoted by $m_{j,0,r}$, for $j = 1, 2, \cdots, k$, and $r = 1, 2, \cdots, m$.

As the transformations in (9.4) and (9.5) are monotonic increasing, medians and conditional medians are transformed. Hence we can write, for $r = 0, 1, 2, \cdots, m$, and $j = 2, 3, \cdots, k$,

$$m_{1,0,r} = \log(m^*_{1,0,r}) - \log(1 - m^*_{1,0,r}), \tag{9.16}$$

$$m_{j,0,r} = \log(m^*_{j,0,r}) - \log(m^*_{1,0,r}). \tag{9.17}$$

It is worth mentioning here that the validity of (9.17) is a result of defining $m^*_{j,0,r}$ as the conditional median of $(p_j|p_1 = m^*_{1,0,r})$, which implies that $m_{j,0,r}$ is a conditional median of $(Y_j|Y_1 = m_{1,0,r})$. That is why we need the redundant variable, Y_1, to be defined in (9.5).

The computed assessments from (9.16) and (9.17), together with the linearity assumptions in (9.6) and (9.7), enable us to determine $\mu_j = E(\alpha_j)$, for $j = 2, \cdots, k$, as

$$\mu_j = E(Y_j|X_i = 0, \forall i = 1, 2, \cdots, m) = m_{j,0,0}. \tag{9.18}$$

We must determine $\mu_{r,j} = E(\beta_{r,j})$ for $r = 1, 2, \cdots, m$, $j = 2, \cdots, k$. If X_r is a factor, then from (9.6) and (9.7), and utilizing the assessments in (9.16) and (9.17), we put

$$\begin{aligned}\mu_{r,j} &= E(Y_j|X_r = 1, X_i = 0, \forall i \neq r) - E(Y_j|X_i = 0, \forall i = 1, 2, \cdots, m)\\ &= m_{j,0,r} - m_{j,0,0}.\end{aligned} \tag{9.19}$$

If X_r is a continuous covariate, then $\beta_{r,j}$ is the slope of the linear relation in (9.6), so

$$\begin{aligned}\mu_{r,j} &= [E(Y_j|X_r = x_r, X_i = 0, \forall i \neq r) - E(Y_j|X_i = 0, \forall i = 1, 2, \cdots, m)]/x_r\\ &= [m_{j,0,r} - m_{j,0,0}]/x_r,\end{aligned} \tag{9.20}$$

for $r = 1, 2, \cdots, m$, and $j = 2, \cdots, k$.

Finally, we put

$$\underline{\mu}_\alpha = (\mu_2, \ \ \mu_3, \ \ \cdots, \ \ \mu_k)', \tag{9.21}$$

and

$$\underline{\mu}_\beta = (\mu_{1,2}, \quad \cdots, \quad \mu_{1,k}, \quad \mu_{2,2}, \quad \cdots, \quad \mu_{2,k}, \quad \mu_{m,2}, \quad \cdots, \quad \mu_{m,k})'. \tag{9.22}$$

9.5 Eliciting the variance matrix

To elicit a positive-definite matrix for the multivariate normal prior distribution of the regression coefficients in (9.15), we proceed as follows.

9.5.1 Eliciting the variance-covariance sub-matrices

We denote $\Sigma_\alpha = \text{Var}(\underline{\alpha}^1)$ by Σ_0, and put

$$\Sigma_{r|\alpha} = \text{Var}(\underline{\beta}^1_{(r)}|\underline{\alpha}^1) \tag{9.23}$$

and

$$\Sigma_{\beta|\alpha} = \text{Var}(\underline{\beta}^{1'}_{(1)}, \quad \underline{\beta}^{1'}_{(2)}, \quad \cdots, \quad \underline{\beta}^{1'}_{(m)}|\underline{\alpha}^1). \tag{9.24}$$

In order to develop a method for eliciting positive-definite matrices Σ_0 and $\Sigma_{r|\alpha}$ ($r = 1, \cdots, m$), we proceed as follows.

From (9.7) we put

$$\Sigma_0 = \text{Var}(\underline{Y}^1|X_i = 0, \forall i = 1, 2, \cdots, m) \equiv V_0, \tag{9.25}$$

where $\underline{Y}^1 = (Y_2, \quad Y_3, \quad \cdots, \quad Y_k)$.

For continuous covariates, if we assume that $X_r = x_r$ and $X_i = 0$, for $i = 1, 2, \cdots, m$, $i \neq r$, we have from (9.6) that

$$\text{Var}(\underline{Y}^1|X_r = x_r, \underline{\alpha}^1 = \underline{\mu}_\alpha) = x_r^2 \ \text{Var}(\underline{\beta}^1_{(r)}|\underline{\alpha}^1 = \underline{\mu}_\alpha) \equiv V_r. \tag{9.26}$$

Hence, for $r = 1, 2, \cdots, m$

$$\Sigma_{r|\alpha} = x_r^{-2} \ V_r. \tag{9.27}$$

For factors, (9.27) is reduced to

$$\Sigma_{r|\alpha} = V_r. \tag{9.28}$$

Each matrix V_r ($r = 0, 1, \cdots, m$) can be elicited as a positive-definite matrix in the way used to obtain the variance matrix of the logistic normal prior in Chapter 8.

Remark 9.1

The main difference between this chapter and Chapter 8 is that here the process of assessing the conditional medians and quartiles must be repeated $m+1$ times. In the initial step, the expert is asked to assume that $X_i = 0, \forall i = 1, 2, \cdots, m$. Then, in each successive step number r, for $r = 1, 2, \cdots, m$, the expert is asked to assume that the rth covariate has changed from 0 to x_r, i.e. $X_r = x_r$, while all other covariates are at their reference values, i.e. $X_i = 0$, for $i = 1, 2, \cdots, m$, $i \neq r$. During these remaining m steps, another key assumption is made. The expert is also conditioning on $\underline{\alpha}^1 = \underline{\mu}_\alpha$.

Under these main assumptions at step r, $r = 0, 1, \cdots, m$, the assessment tasks can be detailed as follows.

9.5.2 Assessing conditional quartiles

- Under the assumptions listed in Remark 9.1, the expert is asked to assess a lower quartile $L^*_{1,r}$ and an upper quartile $U^*_{1,r}$ for p_1. She is then asked to assume that $p_1 = m^*_{1,0,r}$ and gives a lower quartile $L^*_{2,r}$ and an upper quartile $U^*_{2,r}$ for p_2.

- For each remaining p_j, $j = 3, \cdots, k-1$, she assesses the two quartiles $L^*_{j,r}$ and $U^*_{j,r}$ given that $p_1 = m^*_{1,0,r}$, $p_2 = m^*_{2,0,r}$, ..., $p_{j-1} = m^*_{j-1,0,r}$.

- Using the interactive PEGS-Multinomial with Covariates software, and due to the unit sum constraint, the lower (upper) quartile $L^*_{k,r}$ ($U^*_{k,r}$) of p_k is automatically shown to the expert once she assesses the upper (lower) quartile $U^*_{k-1,r}$ ($L^*_{k-1,r}$) of p_{k-1}.

- With the aid of a lognormal curve produced by the software, the expert is advised to make sure that her assessed interquartile range gives an almost zero probability of p_j exceeding $1 - \sum_{i=1}^{j-1} m^*_{i,0,r}$. For more details on this, see Section 8.3.2 in the previous chapter.

9.5.3 Assessing conditional medians

- Under the assumptions listed in Remark 9.1, for $r = 0, 1, \cdots, m$, the expert is asked to assume that the median of p_1 has been changed from $m^*_{1,0,r}$ to $m^*_{1,1,r} = m^*_{1,0,r} + \eta^*_{1,r}$. Given this information, the expert is asked to change her previous medians $m^*_{j,0,r}$ of each p_j. Her new assessment, $m^*_{j,1,r}$, may be written as

$$m^*_{j,1,r} = m^*_{j,0,r} + \theta^*_{j,1,r}, \quad \text{for } j = 2, \cdots, k. \tag{9.29}$$

- In each successive step i, for $i = 2, 3, \cdots, k-2$, the expert is asked to suppose that the median values of $p_1, p_2, ..., p_i$ are $m^*_{1,1,r} = m^*_{1,0,r} + \eta^*_{1,r}, m^*_{2,2,r} = m^*_{2,1,r} + \eta^*_{2,r}, \cdots, m^*_{i,i,r} = m^*_{i,i-1,r} + \eta^*_{i,r}$, respectively. These are shown as red bars in Figure 9.2.

  Given this information, she is asked to revise the medians that she assessed at the most recent previous step $m^*_{i+1,i-1,r}, m^*_{i+2,i-1,r}, \cdots, m^*_{k,i-1,r}$, shown by black lines in Figure 9.2. Her new assessments are denoted $m^*_{i+1,i,r} = m^*_{i+1,i-1,r} + \theta^*_{i+1,i,r}$, $m^*_{i+2,i,r} = m^*_{i+2,i-1,r} + \theta^*_{i+2,i,r}$, $\cdots$, $m^*_{k,i,r} = m^*_{k,i-1,r} + \theta^*_{k,i,r}$, respectively, which are shown as the blue bars in the main graph of Figure 9.2. In other words, for $i = 1, 2, \cdots, k-2$, and $j = i+1, i+2, \cdots, k$, we can write

$$m^*_{j,i,r} = m^*_{j,i-1,r} + \theta^*_{j,i,r} \text{ is the median of } (p_j | p_1 = m^*_{1,1,r}, \cdots, p_i = m^*_{i,i,r}). \tag{9.30}$$

- For mathematical coherence, as proved in Lemma 8.3, Section 8.4.2 in the previous chapter, we have to make sure that

$$\sum_{j=1}^{i} m^*_{j,j,r} + \sum_{j=i+1}^{k} m^*_{j,i,r} = 1, \qquad i = 1, 2, \cdots, k-2.$$

  The software suggests new normalized conditional medians satisfying the above constraint.

- As mentioned in Remark 9.1, the expert assesses her conditional medians assuming that only one of the covariates, age, has changed from its reference value to 40 years, and assuming at the same time that her previously assessed medians at the reference point are correct. Probability medians at the reference point are presented to the expert in the upper right graph of Figure 9.2. The expert is asked to assume these medians are the true values while assessing her conditional medians on the main graph of Figure 9.2.

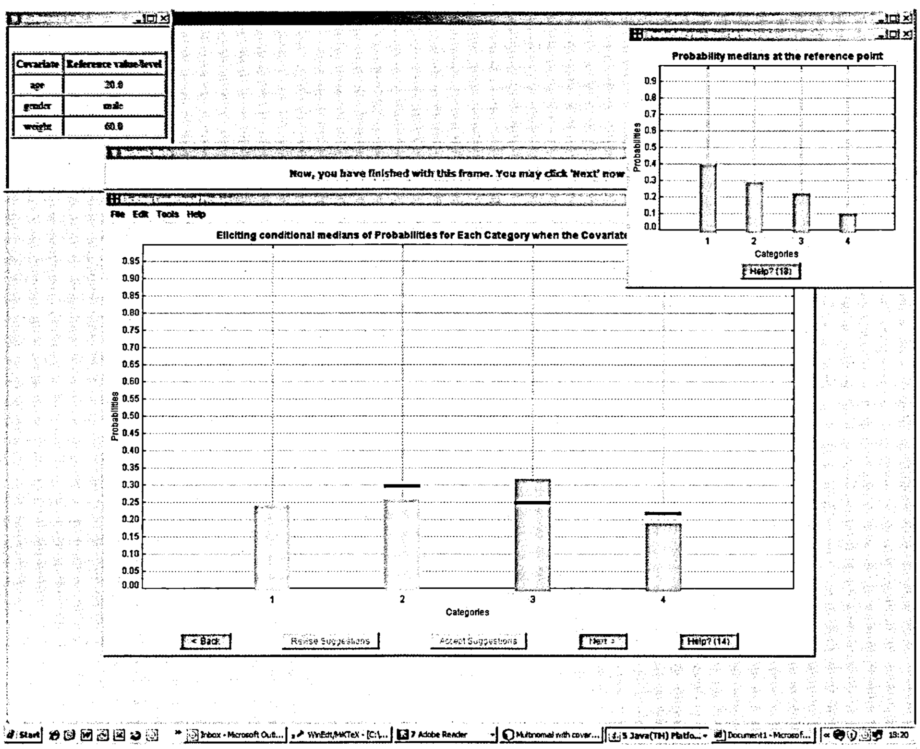


Figure 9.2: Assessing conditional medians at age = 40 years

Assessment tasks in Sections 9.5.2 and 9.5.3 will be repeated $m + 1$ times, for $r = 0, 1, \cdots, m$. Then, as detailed in the previous chapter, the normalizing one-to-one functions in (9.4) and (9.5) are used to transform the assessed conditional quartiles of $\underline{p}$ into conditional quartiles of $\underline{Y}$ and, hence, into conditional expectations, variances and covariances of the multivariate normal elements.

The method of Kadane *et al.* (1980) is modified, as in the previous chapter, to estimate a positive-definite variance-covariance matrix V_r for $\underline{Y}^1|X_r$, from the assessed conditional medians and quartiles. So, because of the unit sum constraint, each positive-definite matrix V_r is of order $(k - 1)$.

Under the assumptions leading to (9.15), and in view of (9.23) and (9.24), the diagonal blocks of the block-diagonal matrix $\Sigma_{\beta|\alpha}$ are $\Sigma_{r|\alpha}$, where each $\Sigma_{r|\alpha}$ is given by (9.27), for

$r = 1, 2, \cdots, m$. Hence, $\Sigma_{\beta|\alpha}$ is a positive-definite matrix. The unconditional variance-covariance matrix Σ_β will be obtained from $\Sigma_{\beta|\alpha}$ using the covariance matrix $\Sigma_{\alpha,\beta}$. The latter is elicited as follows.

9.5.4 Eliciting the covariance matrix $\Sigma_{\alpha,\beta}$

The covariance matrix of $\underline{\alpha}^1$ and $\underline{\beta}^1$ is the matrix $\Sigma_{\alpha,\beta}$ of order $(k-1) \times m(k-1)$. To elicit this matrix, it is convenient to conformally partition $\Sigma_{\alpha,\beta}$ as

$$\Sigma_{\alpha,\beta} = (\Sigma_{\alpha,\beta_1}, \quad \Sigma_{\alpha,\beta_2}, \quad \cdots, \quad \Sigma_{\alpha,\beta_m}), \tag{9.31}$$

where, for $r = 1, 2, \cdots, m$,

$$\Sigma_{\alpha,\beta_r} = \text{Cov}(\underline{\alpha}^1, \underline{\beta}^1_{(r)}). \tag{9.32}$$

We denote the rows of each Σ_{α,β_r} by $\underline{\sigma}'_{\alpha,\beta_r,t}$, for $t = 2, \cdots, k$, where

$$\underline{\sigma}'_{\alpha,\beta_r,t} = \text{Cov}(\alpha_t, \underline{\beta}^1_{(r)}). \tag{9.33}$$

For any specific value α_t^* satisfying $\alpha_t^* \neq \mu_t$, for $t = 2, \cdots, k$, it can be seen from (9.15), (9.32), (9.33) and the theory of multivariate normal distribution that

$$\underline{\mu}_{\beta_r|\alpha_t} = E(\underline{\beta}^1_{(r)}|\alpha_t = \alpha_t^*) = \underline{\mu}_{\beta_r} + \left[\frac{\alpha_t^* - \mu_t}{\text{Var}(\alpha_t)}\right] \underline{\sigma}'_{\alpha,\beta_r,t}. \tag{9.34}$$

From this

$$\underline{\sigma}'_{\alpha,\beta_r,t} = \left[\frac{\text{Var}(\alpha_t)}{\alpha_t^* - \mu_t}\right] (\underline{\mu}_{\beta_r|\alpha_t} - \underline{\mu}_{\beta_r}). \tag{9.35}$$

Since $\text{Var}(\alpha_t)$ is the $(t-1)$th element of the main diagonal of Σ_0 as in (9.25), then, from (9.32) and (9.33), Σ_{α,β_r} can be elicited using $(k-1)$ assessments of $\underline{\mu}_{\beta_r|\alpha_t}$, for $t = 2, \cdots, k$. Under the normality assumptions, these conditional means of the regression coefficients can be computed from the conditional median assessments of the classification probabilities. This can be detailed as follows. For each covariate X_r $(r = 1, 2, \cdots, m)$ in turn, the expert is asked to assume that each single α_t $(t = 2, \cdots, k)$ in turn has changed from μ_t to α_t^*, i.e. she is asked to assume that the true value of $(p_j|X_i = 0, \forall i = 1, 2, \cdots, m)$ has changed from $m^*_{j,0,0}$ to a new specific value, $m^*_{j,0,0,t}$. This is shown by the change from the black lines to the red bars in the upper right graph of Figure 9.3. Given this information, the expert then assesses her median of $(p_j|X_r = x_r, X_i = 0, i = 1, 2, \cdots, m, i \neq r)$, which we denote by $m^*_{j,0,r|\alpha_t}$, for $j = 2, \cdots, k$. These are assessed as the blue bars in the main graph of Figure 9.3.

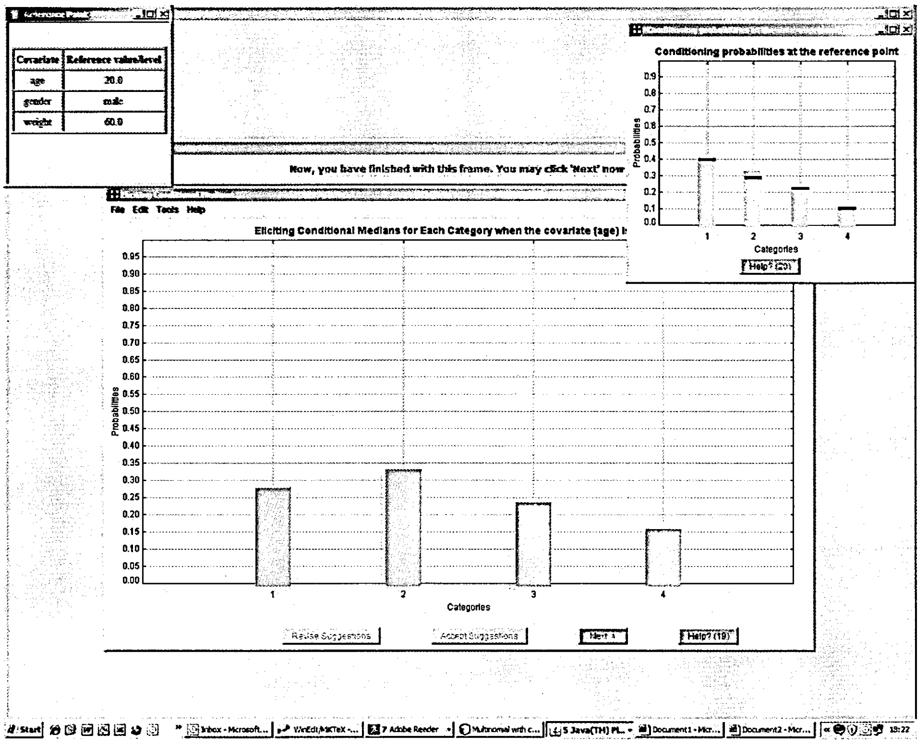


Figure 9.3: Assessing conditional medians given changes at the reference point

The choice of the specific values α^*_t is arbitrary, provided that $\alpha^*_t \neq \mu_t$. However, we select each of them to be the upper quartile of the normally distributed variable α_t, namely,

$$\alpha^*_t = \mu_t + 0.674\sqrt{\text{Var}(\alpha_t)}, \qquad \text{for } t = 2, \cdots, k. \tag{9.36}$$

This leads, from (9.4), (9.5) and (9.7), to sets of conditioning probabilities, $m^*_{j,0,0,t}$, that are given by

$$m^*_{j,0,0,t} = \frac{\exp(\alpha^\sharp_j)}{1 + \sum_{i=1}^k \exp(\alpha^\sharp_i)}, \qquad \text{for } j = 1, \cdots, k, \tag{9.37}$$

where $\alpha^\sharp_1 = 0$, $\alpha^\sharp_t = \alpha^*_t$ and $\alpha^\sharp_j = \mu_j$, for $j \neq t$.

Since, as in (9.34), we condition on changing α_t, for $t = 2, \cdots, k$, one at a time, we have to compute the resulting conditioning probabilities from this change as in (9.37). If we had chosen to first change the conditioning probabilities, the desired change for α_t would not have been guaranteed.

As in (9.17), the corresponding median assessments for Y_j can be computed, for $j = 2, \cdots, k$, $r = 0, 1, \cdots, m$, and $t = 2, \cdots, k$, as

$$m_{j,0,r|\alpha_t} = \log(m^*_{j,0,r|\alpha_t}) - \log(m^*_{1,0,r|\alpha_t}). \tag{9.38}$$

Hence, we denote $E(\beta_{r,j}|\alpha_t = \alpha_t^*)$ by $\mu_{r,j|\alpha_t}$ and compute it as follows.

If X_r is a factor, then as in (9.19), we put

$$\mu_{r,j|\alpha_t} = m_{j,0,r|\alpha_t} - m_{j,0,0|\alpha_t}. \tag{9.39}$$

If X_r is a continuous covariate, then as in (9.20), we put

$$\mu_{r,j|\alpha_t} = \frac{m_{j,0,r|\alpha_t} - m_{j,0,0|\alpha_t}}{x_r}, \tag{9.40}$$

for $r = 1, 2, \cdots, m$, $j = 2, \cdots, k$, and $t = 2, \cdots, k$.

Putting

$$\underline{\mu}_{\beta_r|\alpha_t} = (\mu_{r,2|\alpha_t}, \quad \mu_{r,3|\alpha_t}, \quad \cdots, \quad \mu_{r,k|\alpha_t})', \tag{9.41}$$

all the components of $\underline{\sigma}'_{\alpha,\beta_r,t}$ as in (9.35), and hence of Σ_{α,β_r} as in (9.32), are elicited. Then $\Sigma_{\alpha,\beta}$ as in (9.31) is fully determined.

After obtaining the covariance matrix $\Sigma_{\alpha,\beta}$, and utilizing the elicited matrix $\Sigma_{\beta|\alpha}$, we get Σ_β from the conditional variance

$$\Sigma_{\beta|\alpha} = \Sigma_\beta - \Sigma'_{\alpha,\beta}\Sigma_\alpha^{-1}\Sigma_{\alpha,\beta}, \tag{9.42}$$

which gives

$$\Sigma_\beta = \Sigma_{\beta|\alpha} + \Sigma'_{\alpha,\beta}\Sigma_\alpha^{-1}\Sigma_{\alpha,\beta}. \tag{9.43}$$

Since $\Sigma_{\beta|\alpha}$ and Σ_α are positive-definite, so is Σ_β. Also, from (9.43) and using the Schurr complement, the full variance-covariance matrix of the multivariate normal prior distribution in (9.15) is positive-definite. It is of order $(k-1)(m+1)$ and does not contain variances or covariances of α_1, nor the elements of $\underline{\beta}_1$. This is equivalent to the usual identifiability assumption of the base-line multinomial logit models, where the regression coefficients of the base-line category are set equal to zeros.

9.6 Concluding comments

A novel method has been introduced for eliciting a multivariate normal prior distribution for the regression coefficients in a multinomial logit model with explanatory covariates. The method is an extension of our proposed method in Chapter 8 for eliciting a logistic normal prior for classification probabilities in a multinomial model. Specifically, under a base-line multinomial logit model containing k categories and m explanatory covariates, assessment tasks of a standard multinomial model are repeated $m+1$ times. The expert assesses conditional medians and quartiles for the multinomial probabilities at specific values of each explanatory covariate. This determines a mean vector and a positive-definite variance-covariance matrix of a multivariate normal prior distribution for $(k-1)(m+1)$ regression coefficients.

Chapter 10

Concluding comments

This chapter summarizes the main results and conclusions of the thesis. We give a brief review of the elicitation methods proposed throughout this thesis, commenting on the main assumptions, strength and weakness points of each proposed method. In addition, the inter-relationships between related methods are mentioned and clarified. The proposed methods divide naturally in two groups: methods of quantifying expert opinion for GLMs and methods of prior elicitation for multinomial models. The proposed methods in each group are briefly discussed in order. Some extensions for further future research are given.

The method proposed by Garthwaite and Al-Awadhi (2006) and its extension in Garthwaite and Al-Awadhi (2011) can be considered a general tool for eliciting a multivariate normal prior for the regression coefficients in any GLM. In their method, opinion about the relationship between each continuous predictor variable and the response variable is modeled by a piecewise-linear function. This gives a flexible model that can represent a wide variety of opinion. Expert opinion about each categorical predictor variable (factor) is elicited through a bar-chart. Each slope of the piecewise-linear relationships and each level of the factors has a corresponding regression coefficient. The expert assesses conditional medians and quartiles of the response variable at different selected design points. In this sense, the method applies the idea of conditional means prior proposed by Bedrick *et al.* (1996). Conditional assessments are transformed, under the normality assumption of regression coefficients, to estimate a mean vector and a variance-covariance matrix for the multivariate normal prior distribution. Conditional quartiles are assessed in a structural way that ensures that the resulting matrix is positive-definite.

The method proposed by Garthwaite and Al-Awadhi (2011) has been implemented in interactive graphical user-friendly software, in which the expert draws piecewise-linear curves and bar-charts by clicking on interactive graphs on a computer screen to give her assessments. The software computes and offers suggestions to the expert to help reduce the burden of making assessments. A prototype of this software was written in Java by Jenkinson (2007) and has been modified and extended in the current thesis to be more flexible and to include more options. A detailed description of the method and the current modifications to the software has been given in Chapter 3. Previously the software could only handle logistic regression but now it handles a wide range of GLMs. As noted earlier, an important feedback option has been added to the software. As each covariate is assessed separately, this feedback option is very useful for helping the expert see the joint impact of all explanatory covariates that her assessments imply.

A simplifying assumption in the method of Garthwaite and Al-Awadhi (2011), that has been relaxed in this thesis, is that regression coefficients had been assumed to be independent, *a priori*, if attached to different explanatory variables. This yielded a block-diagonal variance-covariance matrix and reduced the number of required assessments for its elicitation. However, this independence assumption can be unrealistic in many practical situations. We proposed three elicitation methods for a multivariate normal prior distribution that do not impose this simplifying assumption. The proposed methods elicit full variance-covariance matrices, but additional assessments are needed in order to estimate the off-diagonal elements.

As noted earlier, the three proposed methods differ in their flexibility and in the number of additional assessments that they require. The first method is a direct extension to the method of Garthwaite and Al-Awadhi (2011). It is the most flexible method among the three and permits different correlations between regression coefficients attached to the same pair of covariates. Consequently, it requires a large number of conditional assessments, but it should prove useful when there are only a few pairs of variables that, *a priori*, have highly correlated regression coefficients.

The second proposed method uses only one assessment to model the correlation between all regression coefficients attached to any specific pair of explanatory covariates. This assumption, of fixed correlations for all elements belonging to the same pair of vectors of coefficients, is useful as it reduces the assessment tasks to just one task. The expert is asked to use a slider to determine the correlation between two vectors of regression coefficients. This can be attractive as an easy and quick method for eliciting correlations if only two vectors of regression coefficients are thought to be correlated. Moreover, for the case where more than two vectors have correlated regression coefficients, we extended the method and showed it will yield a full variance-covariance matrix that is positive-definite.

The third method we proposed is suitable for GLMs that contain a large number of correlated vectors. It uses a few assessments that directly reflect the pattern of correlations between all pairs of vectors. In a dialogue box, the expert assesses the relative magnitudes and signs of the average correlations between each pair of vectors. Hence, for n vectors of coefficients, $n(n-1)/2$ assessments are needed. These relative magnitudes should reflect the strength of the average correlation of each pair relative to other pairs. It is a comparatively easy task for the expert as these assessments need not be coherent correlation coefficients; they are scaled later to attain statistical coherence. The method avoids incremented conditioning and assesses all covariances simultaneously.

After assessing the relative magnitudes, using the PEGS-GLM (Correlated Coefficients), the third method can be used alone or together with one of the other two proposed methods, to obtain correlations. The default option, that implements this method alone, is to use one slider to determine correlation coefficients based on simultaneous interactive graphs that show the changes of different variables according to their assessed relative magnitudes. The other two alternate options need an assessment of the correlation of only one pair of vectors, then all other correlation coefficients are computed from this assessment using the relative magnitudes. The correlation assessment for one of the highly correlated pairs may be obtained using one of the other two proposed methods. The first of them needs more assessments, while the second method assumes a fixed correlation structure for the elements of the highly correlated pair of vectors. Figure 10.1 shows the different options available to the expert for choosing which method to use when she is assessing correlations between regression coefficients in GLMs. These are the different options offered by our PEGS-GLM (Correlated Coefficients) software that is freely available at http://statistics.open.ac.uk/elicitation.

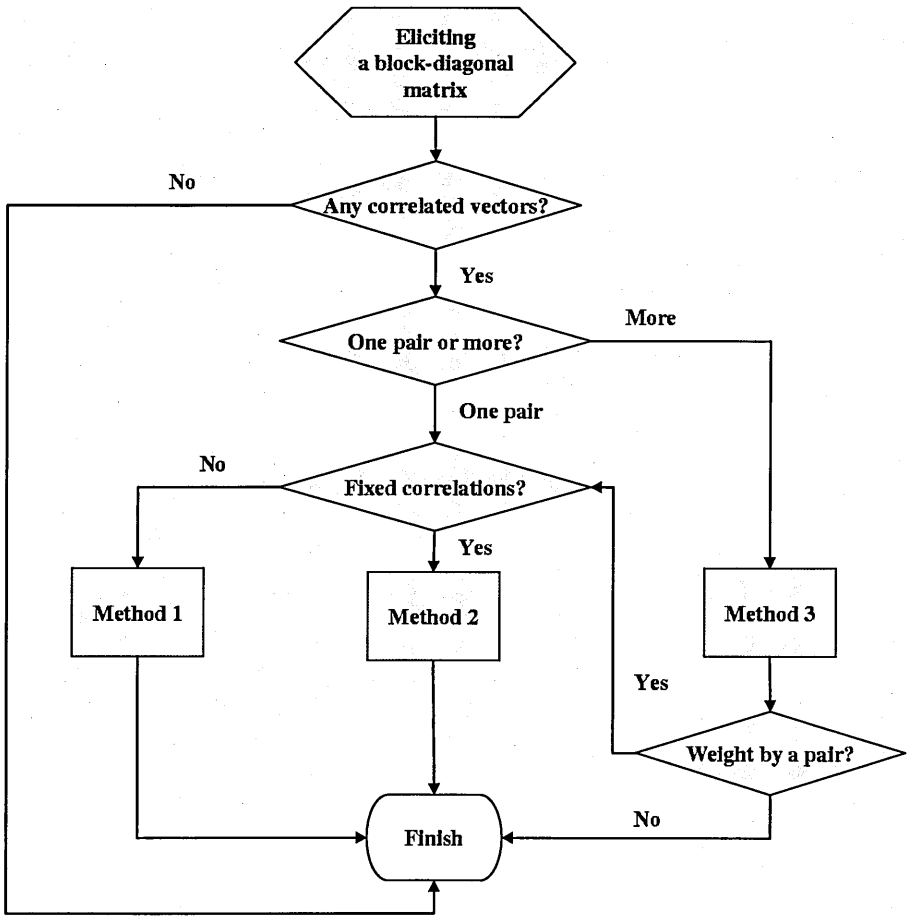


Figure 10.1: Options for assessing correlations between regression coefficients

To complete the prior structure of GLMs with normal and gamma response variables, we proposed two methods of eliciting prior distributions for the extra parameters in these models. One of these methods elicits a conjugate chi-squared prior distribution for the random error variance in normal linear models. The expert is asked to revise her assessments conditional on various sets of hypothetical future samples. A number of sets of hypothetical data are used in order to obtain several estimates of the hyperparameter that is most difficult to assess, namely, the degrees of freedom parameter of the chi-squared distribution. Reconciliation of these estimates, using the geometric mean, yields an overall estimate of the number of degrees of freedom. The second hyperparameter of the chi-squared prior distribution is also determined from the same assessments. The use of interactive graphical software greatly facilitates the tasks that the expert must perform.

For a gamma response variable, the additional parameter that must be assessed is the scale parameter. We assumed that prior opinion about this positive-valued parameter can be reasonably quantified as a lognormal distribution. To determine the hyperparameters of the lognormal prior distribution, the expert is asked to give a point estimate and an interquartile range for the lower quartile of the gamma response variable. We proved that the lower quartile is a monotonic increasing function of the scale parameter. The expert's assessments are thus transformed to quartiles of the lognormal distribution, and hence to the mean and variance of the lognormal distribution. An example of the questions that can be asked in order to obtain the expert's assessments has been given. As noted earlier, no other reasonable elicitation methods for the scale parameters of gamma GLMs seems to be available in the literature.

Eliciting flexible prior distributions for the classification probabilities in multinomial models has been another important interest of this thesis. In this context, we started by proposing two elicitation methods for the natural conjugate Dirichlet prior. The first method is based on marginal quartile assessments of the classification probabilities. These assessments were used to elicit separate marginal beta distributions of the Dirichlet prior distribution. A normal approximation and least-squares techniques have been used to obtain beta parameters from the quartile assessments. From three reconciliations of beta distributions into a Dirichlet prior distribution, the expert is asked to select the reconciliation that best describes her opinions, based on graphical feedback. The second method elicits conditional quartile assessments for the classification probabilities. These conditional assessments are used to determine conditional beta distributions that are averaged to obtain a Dirichlet prior distribution.

The same marginal and conditional quartile assessments for classification probabilities have been used to elicit two other flexible prior distributions for multinomial models. Conditional quartile assessments were used to elicit conditional beta distributions of a generalized Dirichlet prior distribution. As noted earlier, this distribution is more flexible than the standard Dirichlet distribution for quantifying expert opinion. It has the same number of hyperparameters as the total number of parameters in the conditional beta distributions that determine it. Hence no reconciliation is needed. The generalized Dirichlet distribution has a more general dependence structure than the standard Dirichlet. For example, its correlation structure allows positive correlations between classification probabilities.

Marginal assessments were used to elicit marginal beta distributions for multinomial probabilities. Then, instead of assuming a Dirichlet prior, the beta marginals were used in a Gaussian copula function to model the joint prior distribution of multinomial probabilities. This required further conditional quartile assessments to describe the correlation structure between these probabilities. The monotonicity of the Gaussian copula transformation allowed conditional quartiles of the multinomial probabilities to be transformed into normal quartiles. The latter were 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 assessment of the conditional quartiles has been used to ensure that the elicited variance-covariance matrix is positive-definite.

The conditional quartile assessments that were used to elicit correlations for a Gaussian copula prior were also used in a new method for eliciting a logistic normal prior distribution for multinomial probabilities. Quantifying expert opinion as a logistic normal prior raised some interesting points that do not seem to have arisen in elicitation contexts before. We made use of the natural approximation of the lognormal sum by another lognormally distributed random variable. In addition, our proposed method has extensively used the notion of singular multivariate normal distribution; available literature shows that conditional properties of the singular normal distribution is nearly identical to their corresponding properties in the standard normal distribution. These results were used to prove that the medians, not only the means, of multinomial probabilities must sum to one, assuming they follow a logistic normal distribution. This was critical in building the elicitation method as it enables assistance to be given to the expert that leads to statistically coherent assessments.

The four proposed prior distributions are interrelated regarding the assessments that they use. Each type of assessments can be used to elicit more than one prior distribution. The Prior Elicitation Graphical Software package for Multinomial models, PEGS-Multinomial, that is freely available at http://statistics.open.ac.uk/elicitation, arranges the assessment tasks that are required for the four proposed prior distributions. Software is also available that elicits each of the prior distributions separately. The flowchart in Figure 10.2 shows the options for prior distributions that are available in PEGS-Multinomial and the corresponding assessments that they require. For example, it shows that a Gaussian copula prior is elicited using two types of assessments, and that a standard Dirichlet prior is elicited using either marginal or conditional assessments, as discussed before. Since conditional beta assessments can be used to elicit both the standard and generalized Dirichlet distributions, the software gives the option of eliciting both of them using the same conditional quartiles.

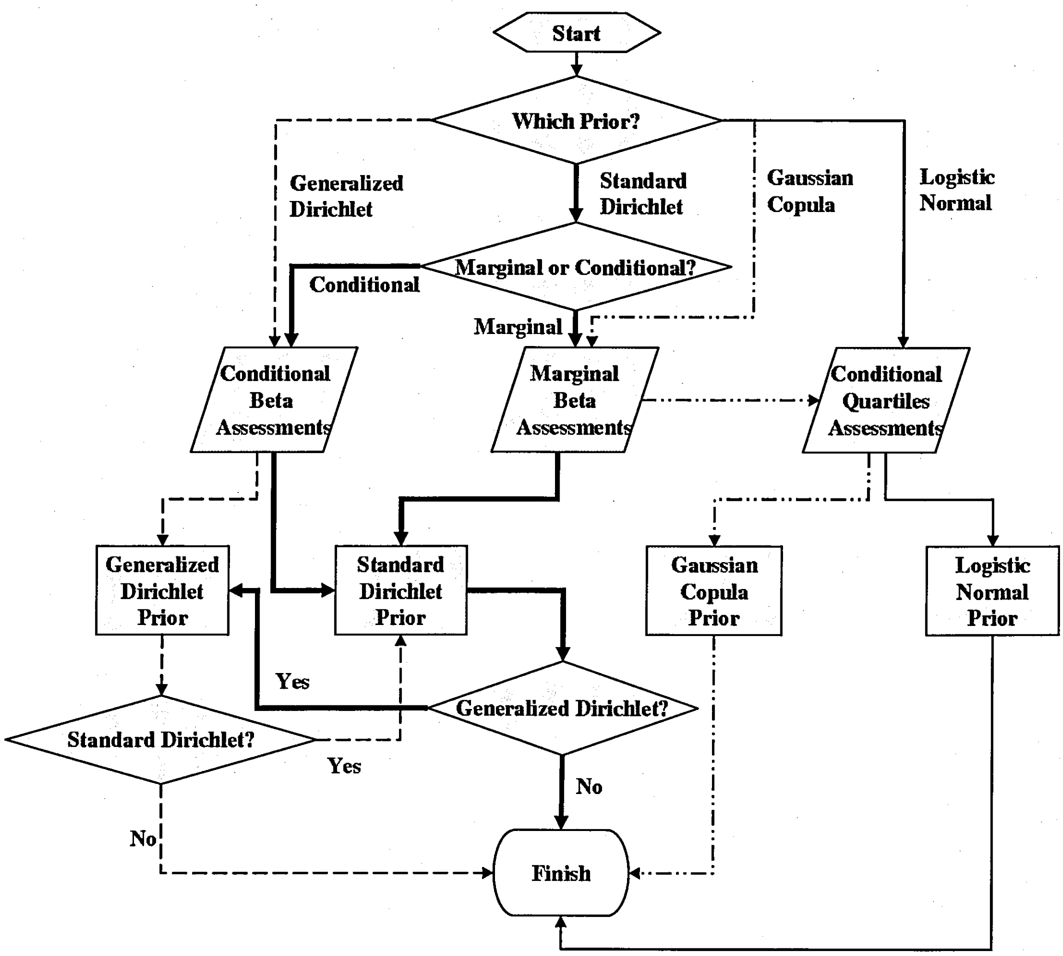


Figure 10.2: A flowchart of the prior elicitation software for multinomial models

All the proposed prior elicitation methods for multinomial models and their implementing software have been used in examples by real experts. In all examples, the experts suggested the problem according to their fields of expertise. They understood the multinomial formulation and were keen to participate in the elicitation process. After a brief discussion about the ideas of the bisection method and conditional assessments they had no problem in assessing quartiles and conditional quartiles. All the experts expressed the view that visualization of the problem had helped them a lot in quantifying their opinions. They also made use of the coherent suggestions given by the software and used the feedback options to revise some of their assessments. Thus the software proved important in providing visualization, coherent suggestions and feedback. It also helped the experts review and revise their assessments, and reduced the time taken by the elicitation processes.

Future research in assessment methods for GLMs may include eliciting prior distributions for the overdispersion parameters in binomial and Poisson GLMs. In these important GLMs, it is common that the data show a greater variability than the theoretical variability assumed by the model. However, no elicitation method have been proposed in the literature for quantifying opinion about overdispersion parameters. A reasonable approach might be to assume a generalized binomial distribution or a generalized Poisson distribution for the response variable, instead of the standard binomial or Poisson distributions. These generalized distributions have extra parameters that allow for overdispersion. Methods of assessing suitable prior distributions for these extra parameters need to be developed.

Another extension to the proposed method for GLMs elicitation concerns the proportional hazard model. This model, also known as the Cox regression model, is often used to model survival data in medical research. See, for example, Collett (1994). Due to its wide practical importance, a huge bulk of research has been devoted to investigating both theoretical and applied aspects of Bayesian analysis of a proportional hazard model. See, Ibrahim and Chen (1998) and Zuashkiani *et al.* (2008), among others. Quantifying opinion about these models has also attracted some attention. See, for example, Chaloner *et al.* (1993) and Henschel *et al.* (2009). Adaptation is needed for the current GLM elicitation methods to handle a proportional hazard model.

The method of eliciting logistic normal prior distributions for multinomial models has already been extended further in Chapter 9. The extended method treats the case of multinomial models in which classification probabilities are influenced by explanatory covariates. Specifically, we proposed a method that quantifies opinion about the parameters of a baseline multinomial logit model as a multivariate normal prior distribution. The method uses conditional median and quartile assessments for the classification probabilities at different combinations of the explanatory variables. These assessments have been obtained in a structured way that yields the mean vector and positive-definite variance-covariance matrix of the prior multivariate normal distribution. Another desirable extension would be to elicit a logistic normal prior distribution for the cell probabilities of contingency tables. The logistic normal distribution is considered a reasonable prior for contingency tables, see for example Goutis (1993). Hence, our proposed elicitation method for a logistic normal prior promises to be useful in further contexts.

Other models for which elicitation methods still need to be developed include time series analysis, extreme values analysis and modelling the spread of infectious diseases. These models sometimes investigate cases for which data are scarce, the events are rare, or situations are new and uncontrollable. Expert opinion is highly important in such situations, so the need for appropriate elicitation methods is clear.

Bibliography

Abadir, K. M. and Magnus, J. R. (2005). *Matrix Algebra*. Cambridge University Press, New York.

Agresti, A. (2002). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New Jersey, second edition.

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society, Series B*, **44**, 139–177.

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

Al-Awadhi, S. A. (1997). *Elicitation of Prior Distribution for a Multivariate Normal Distribution*. Ph.D. thesis, University of Aberdeen, UK.

Al-Awadhi, S. A. and Garthwaite, P. H. (1998). An elicitation method for multivariate normal distributions. *Communications in Statistics-Theory and Methods*, **27**, 1123–1142.

Al-Awadhi, S. A. and Garthwaite, P. H. (2001). Prior distribution assessment for a multivariate normal distribution: An experimental study. *Journal of Applied Statistics*, **28**, 5–23.

Al-Awadhi, S. A. and Garthwaite, P. H. (2006). Quantifying expert opinion for modelling fauna habitat distributions. *Computational Statistics*, **21**, 121–140.

Albajar, R. A. and Fidalgo, J. F. L. (1997). Characterizing the general multivariate normal distribution through the conditional distributions. *Extracta Mathematica*, **12**, 15–18.

Albert, J. H. and Gubta, A. K. (1982). Mixtures of Dirichlet distributions and estimation in contingency tables. *The Annals of Statistics*, **10**, 1261–1268.

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

Bedrick, E. J., Christensen, R., and Johnson, W. (1996). A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, **91**, 1450–1460.

Bland, R. P. and Owen, D. B. (1966). A note on singular normal distributions. *Annals of the Institute of Statistical Mathematics*, **18**, 113–116.

Bunn, D. W. (1978). Estimation of a Dirichlet prior distribution. *Omega*, **6**, 371–373.

Bunn, D. W. (1979). Estimation of subjective probability distributions in forecasting and decision making. *Technological Forecasting and Social Change*, **14**, 205–216.

Chaloner, K. and Duncan, G. T. (1983). Assessment of a beta prior distribution: PM elicitation. *The Statistician*, **32**, 174–180.

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.

Chaloner, K., Church, T., Louis, T. A., and Matts, J. P. (1993). Graphical elicitation of a prior distribution for a clinical trial. *The Statistician*, **42**, 341–353.

Chen, M.-H. and Dey, D. K. (2003). Variable selection for multivariate logistic regression models. *Journal of Statistical Planning and Inference*, **111**, 37–55.

Chen, M.-H. and Ibrahim, J. G. (2003). Conjugate priors for generalized linear models. *Statistica Sinica*, **13**, 461–476.

Chen, M.-H., Ibrahim, J. G., and Yiannoutsos, C. (1999). Prior elicitation, variable selection and Bayesian computation for logistic regression models. *Journal of the Royal Statistical Society, Series B*, **61**, 223–242.

Chen, M.-H., Ibrahim, J. G., and Shao, Q.-M. (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*, **84**, 121–137.

Chen, M.-H., Ibrahim, J. G., Shao, Q.-M., and Weiss, R. E. (2003). Prior elicitation for model selection and estimation in generalized linear mixed models. *Journal of Statistical Planning and Inference*, **111**, 57–76.

Chen, M.-H., Huamg, L., Ibrahim, J. G., and Kim, S. (2008). Bayesian variable selection and computation for generalized linear models with conjugate priors. *Bayesian Analysis*, **3**, 585–614.

Clemen, R. C. and Reilly, T. (1999). Correlations and copulas for decision and risk analysis. *Management Science*, **45**, 208–224.

Clemen, R. T., Fischer, G. W., and Winkler, R. L. (2000). Assessing dependence: Some experimental results. *Management Science*, **46**, 1100–1115.

Collett, D. (1994). *Modelling Survival Data in Medical Research*. Chapman and Hall, London.

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.

Daneshkhah, A. and Oakley, J. (2010). Eliciting multivariate probability distributions. In K. Böcker, editor, *Rethinking Risk Measurement and Reporting: Volume I*. Risk Books, London.

Demarta, S. and McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, **73**, 111129.

Denham, R. and Mengersen, K. (2007). Geographically assisted elicitation of expert opinion for regression models. B*ayesian Analysis*, **2**, 99–136.

Dickey, J. M. (1968). Three multidimensional-integral identities with Bayesian applications. *The Annals of Mathematical Statistics*, **39**, 1615–1627.

Dickey, J. M. (1983). Multiple hypergeometric functions: Probabilistic interpretations of statistical uses. *Journal of the American Statistical Association*, **78**, 628–637.

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.

Dickey, J. M., Dawid, A. P., and Kadane, J. B. (1986). Subjective probability assessment methods for multivariate-t and matrix-t models. In P. Goel and A. Zellner, editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 177–195. North-Holland, Amestrdam.

Fan, D. Y. (1991). The distribution of the product of independent beta variables. *Communications in Statistics-Theory and Methods*, **20**, 4043–4052.

Fenton, L. F. (1960). The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems*, **CS-8**, 57–67.

Fischer, M. H. (2001). Cognition in the bisection task. *TRENDS in Cognitive Sciences*, **5**, 460–462.

Flanagan, M. T. (2011). Michael Thomas Flanagan's Java scientific library. http://www.ee.ucl.ac.uk/~mflanaga/java/. [Accessed 9 March 2011].

Forster, J. J. and Skene, A. M. (1994). Calculation of marginal densities for parameters of multinomial distributions. *Statistics and Computing*, **4**, 279–286.

Frees, E. W. and Valdez, E. A. (1998). Understanding relations using copulas. *North American Actuarial Journal*, **2**, 125.

Garthwaite, P. H. (1994). Assessment of prior distributions for regression models: An experimental study. *Communications in Statistics-Simulation and Computation*, **23**, 871–895.

Garthwaite, P. H. (1998). Quantifying expert opinion for modelling habitat distributions. Technical report, Sustainable Forest Management 1998/02, Department of Natural Resources Queensland.

Garthwaite, P. H. and Al-Awadhi, S. A. (2001). Non-conjugate prior distribution assessment for multivariate normal sampling. *Journal of the Royal Statistical Society, Series B*, **63**, 95–110.

Garthwaite, P. H. and Al-Awadhi, S. A. (2006). Quantifying opinion about a logistic regression using interactive graphics. Technical Report 06/07, Statistics Group, The Open University, UK.

Garthwaite, P. H. and Al-Awadhi, S. A. (2011). Quantifying subjective opinion about generalized linear and piecewise-linear models. In preparation.

Garthwaite, P. H. and Dickey, J. M. (1985). Double- and single-bisection methods for subjective probability assessment in a location-scale family. *Journal of Econometrics*, **29**, 149–163.

Garthwaite, P. H. and Dickey, J. M. (1988). Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society, Series B*, **50**, 462–474.

Garthwaite, P. H. and Dickey, J. M. (1992). Elicitation of prior distributions for variable selection problems in regression. *Annals of Statistics*, **20**, 1697–1719.

Garthwaite, P. H. and O'Hagan, A. (2000). Quantifying expert opinion in UK water industry: An experimental study. *The Statistician*, **49**, 455–477.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, **100**, 680–701.

Garthwaite, P. H., Chilcott, J. B., Jenkinson, D. J., and Tappenden, P. (2008). Use of expert knowledge in evaluating costs and benefits of alternative service provisions: A case study. *International Journal of Technology Assessment in Health Care*, **24**, 350–357.

Gautschi, W. (1998). The incomplete gamma functions since tricomi. In *In Tricomi's Ideas and Contemporary Applied Mathematics, Atti dei Convegni Lincei, n. 147, Accademia Nazionale dei Lincei, Roma*, pages 203–237.

Genz, A. and Kwong, K. (1999). Numerical evaluation of singular multivariate normal distributions. *Journal of Statistical Computation and Simulation*, **68**, 1–21.

Good, I. J. (1976). On the application of symmetric Dirichlet distributions and their mixtures to contingency tables. *The Annals of Statistics*, **4**, 1159–1189.

Goutis, C. (1993). Bayesian estimation methods for contingency tables. *Journal of the Italian Statistical Society*, **2**, 35–54.

Grunwald, G. K., Raftery, A. E., and Guttorp, P. (1993). Time series of continuous proportions. *Journal of the Royal Statistical Society, Series B*, **55**, 103–116.

Gupta, A. K. and Nadarajah, S. (2004). Products and linear combinations. In A. K. Gupta and S. Nadarajah, editors, *Handbook of Beta Distribution and Its Applications*. Marcel Dekker, Inc., New York.

Hankin, R. K. S. (2010). A generalization of the Dirichlet distribution. *Journal of Statistical Software*, **33**, 1–18.

Henschel, V., J., E., Hölzel, D., and Mansmann, U. (2009). A semiparametric Bayesian proportional hazards model for interval censored data with frailty effects. *BMC Medical Research Methodology*, **9:9**. Available at http://www.biomedcentral.com/1471−2288/9/9 [Accessed 27th February 2009].

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, **70**, 271–289.

Hora, S. C., Hora, J. A., and Dodd, N. G. (1992). Assessment of probability distributions for continuous random variables: A comparison of the bisection and fixed value methods. *Organizational Behavior and Human Decision Processes*, **51**, 133–155.

Hughes, G. and Madden, L. V. (2002). Some methods for eliciting expert knowledge of plant disease epidemics and their application in cluster sampling for disease incidence. *Crop Protection*, **21**, 203215.

Ibrahim, J. G. and Chen, M.-H. (1998). Prior distributions and Bayesian computation for proportional hazards models. *Sankhyā: The Indian Journal of Statistics*, **Spl Series**, 48–64.

Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *Journal of the American Statistical Association*, **86**, 981–986.

Ibrahim, J. G. and Laud, P. W. (1994). A predictive approach to the analysis of designed experiments. *Journal of the American Statistical Association*, **89**, 309–319.

James, A., Low Choy, S., and Mengersen, K. L. (2010). Elicitator: an expert elicitation tool for regression in ecology. *Environmental Modelling & Software*, **25**, 129–145.

Jenkinson, D. J. (2007). *Quantifying Expert Opinion as a Probability Distribution*. Ph.D. thesis, The Open University, UK.

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

Johnson, N., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1. Wiley, New York, second edition.

Jouini, M. N. and Clemen, R. T. (1996). Copula models for aggregating expert opinions. *Operations Research*, **44**, 444–457.

Kadane, J. B. and Wolfson, L. J. (1998). Experiences in elicitation. *The Statistician*, **47**, 3–19.

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.

Kadane, J. B., Shmueli, G., Minka, T. P., Borle, S., and Boatwright, P. (2006). Conjugate analysis of the Conway-Maxwell-Poisson distribution. *Bayesian Analysis*, **1**, 363–374.

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

Koornwinder, T. H. (2008). On a monotonticity property of the normalized incomplete gamma function. http://staff.science.uva.nl/~thk/art/comment/. [Accessed 5 December 2011].

Krzysztofowicz, R. and Reese, S. (1993). Stochastic bifurcation processes and distributions of fractions. *Journal of the American Statistical Association*, **88**, 345–354.

Kurowicka, D. and Cooke, R. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd, Chichester.

Kwong, K. and Iglewicz, B. (1996). On singular multivariate normal distribution and its applications. *Computational Statistics and Data Analysis*, **22**, 271–285.

Kynn, M. (2005). *Eliciting Expert Knowledge for Bayesian Logistic Regression in Species Hapitat Modelling in Natural Resources*. Ph.D. thesis, Queensland University of Technology, Australia.

Kynn, M. (2006). Designing elicitor: Software to graphically elicit priors for logistic regression models in ecology. Available at http://www.winbugs-development.org.uk/elicitor/files/designing.elicitor.pdf [Accessed 10th October 2008].

Kynn, M. (2008). The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society, Series A*, **171**, 239–264.

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

Lewandowski, D. (2008). *High Dimensional Dependence: Copulae, Sensitivity, Sampling*. Ph.D. thesis, Delft University of Technology, Netherlands.

Lindley, D. V., Tversky, A., and Brown, R. V. (1979). On the reconciliation of probability assessments. *Journal of the Royal Statistical Society, Series A*, **142**, 146–180.

Lochner, R. H. (1975). A generalized Dirichlet distribution in Bayesian life testing. *Journal of the Royal Statistical Society, Series B*, **37**, 103–113.

Low Choy, S., O'Leary, R., and Mengersen, K. (2009). Elicitation by design in ecology: Using expert opinion to inform priors for Bayesian statistical models. *Ecology*, **90**, 265–277.

Low Choy, S., James, A., Murray, J., and Mengersen, K. (2010). Indirect elicitation from ecological experts: From methods and software to habitat modelling and rock-wallabies. In A. O'Hagan and M. West, editors, *The Oxford Handbook of Applied Bayesian Analysis*. Oxford University Press, Inc., New York.

Mahmoud, A. S. H. (2010). New quadrature-based approximations for the characteristic function and the distribution function of sums of lognormal random variables. *IEEE Transactions on Vehicular Technology*, **59**, 3364–3372.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London, second edition.

Meyer, M. C. and Laud, P. W. (2002). Predictive variable selection in generalized linear models. *Journal of the American Statistical Association*, **97**, 859–871.

Miller, R. B. (1980). Bayesian analysis of the two-parameter gamma distribution. *Technometrics*, **22**, 65–69.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308–313.

Nelsen, R. B. (1999). *An Introduction to Copulas.* Lecture Notes in Statistics, 139. Springer-Verlag, New York.

Oakley, J. (2010). Eliciting univariate probability distributions. In K. Böcker, editor, *Rethinking Risk Measurement and Reporting: Volume I.* Risk Books, London.

Oakley, J. E. and O'Hagan, A. (2010). SHELF: the Sheffield elicitation framework (version 2.0). School of Mathematics and Statistics, University of Sheffield, UK. http://tonyohagan.co.uk/shelf. [Accessed 9 March 2011].

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *The Statistician*, **47**, 21–35.

O'Hagan, A. and Forster, J. (2004). *Bayesian 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.

O'Leary, R. A., Low Choy, S., Murray, J. V., Kynn, M., Denham, R., Martin, T. G., and Mengersen, K. (2009). Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby *Petrogalae peicillata*. *Environmmetrics*, **20**, 379–398.

Oman, S. D. (1985). Specifying a prior distribution in structured regression problems. *Journal of the American Statistical Association*, **80**, 190–195.

Palomo, J., Insua, D. R., and Ruggert, F. (2007). Modeling external risks in project management. *Risk Analysis*, **27**, 961–978.

Patel, J. K. and Read, C. B. (1982). *Handbook of the Normal Distribution*. Marcel Dekker, Inc., New York.

Peterson, C. R. and Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, **68**, 29–46.

Powers, D. A. and Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. Academic Press, San Diego, CA.

Pratt, J. W., Raiffa, H., and Schalifer, R. (1995). *Introduction to Statistical Decision Theory*. The MIT Press, London.

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

Rayens, W. S. and Srinivasan, C. (1994). Dependence properties of generalized Liouville distributions on the simplex. *Journal of the American Statistical Association*, **89**, 1465–1470.

Safak, A. (1993). Statistical analysis of the power sum of multiple correlated log-normal components. *IEEE Transactions on Vehicular Technology*, **42**, 58–61.

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.

Shields, M., Gorber, S. C., and Tremblay, M. S. (2008). Effects of measurement on obesity and morbidity. *Health Reports, Statistics Canada, Catalogue 82-003*, **19**, 1–8.

Staël von Holstein, C. A. S. (1971). The effect of learning on the assessment of subjective probability distributions. *Organizational Behavior and Human Performance*, **6**, 304–315.

Styan, G. P. H. (1970). Notes on the distribution of quadratic forms in singular normal variables. *Biometrika*, **57**, 567–572.

Sweeting, T. (1981). Scale parameters: a Bayesian treatment. *Journal of the Royal Statistical Society, Series B*, **43**, 333–338.

Tellambura, C. and Senaratne, D. (2010). Accurate computation of the MGF of the lognormal distribution and its application to sum of lognormals. *IEEE Transactions on Communications*, **58**, 1568–1577.

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.

Tricomi, F. G. (1952). Sulla funzione gamma incompleta. *Annali di Matematica Pura ed Applicata*, **31**, 263–279.

van Dorp, J. R. and Kotz, S. (2002). A novel extension of the triangular distribution and its parameter estimation. *The Statistician*, **51**, 93–79.

van Dorp, J. R. and Mazzuchi, T. A. (2000). Solving for the parameters of a beta distribution under two quantile constraints. *Journal of Statistical Computation and Simulation*, **67**, 189–201.

van Dorp, J. R. and Mazzuchi, T. A. (2003). Parameter specification of the beta distribution and its Dirichlet extensions utilizing quantiles. *Beta Distributions and Its Applications*, **29**, 1–37.

van Dorp, J. R. and Mazzuchi, T. A. (2004). Parameter specification of the beta distribution and its Dirichlet extensions utilizing quantiles. In A. K. Gupta and S. Nadarajah, editors, *Handbook of Beta Distribution and Its Applications*. Marcel Dekker, Inc., New York.

Wallsten, T. S. and Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, **29**, 151–173.

West, M. (1985). Generalized linear models: Scale parameters, outlier accommodation and prior distributions. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics 2*, pages 531–558. Elsevier, North-Holland.

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

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, **80**, 73–83.

Wilks, S. S. (1962). *Mathematical Statistics*. John Wiley & Sons, Inc., New York.

Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, **62**, 776–800.

Wong, T. T. (1998). Generalized Dirichlet distribution in Bayesian analysis. *Applied Mathematics and Computation*, **97**, 165–181.

Wong, T. T. (2005). A Bayesian approach employing generalized Dirichlet priors in predicting microchip yields. *Journal of the Chinese Institute of Industrial Engineers*, **22**, 210–217.

Wong, T. T. (2007). Perfect aggregation of Bayesian analysis on compositional data. *Statistical Papers*, **48**, 265–282.

Wong, T. T. (2010). Parameter estimation of generalized Dirichlet distributions from the sample estimates of the first and the second moments of random variables. *Computational Statistics and Data Analysis*, **54**, 1756–1765.

Yi, W. and Bier, V. M. (1998). An application of copulas to accident precursor analysis. *Management Science*, **44**, S257–S270.

Zuashkiani, A., Banjevic, D., and Jardine, A. (2008). Estimating parameters of proportional hazards model based on expert knowledge and statistical data. *Journal of the Operational Research Society*, pages 1–16. Available at http://www.palgrave-journals.com/jors/journal/vaop/ncurrent/full/jors2008119a.html [Accessed 17th June 2009].