Belief representation for counts in Bayesian inference and experimental design

Kevin James Wilson

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics & Statistics
Newcastle University
Newcastle upon Tyne
United Kingdom

April 2011

Acknowledgements

Firstly I would like to thank my supervisor Dr Malcolm Farrow. I don't believe a more patient person could possibly exist. His quiet reassurance, guidance and advice, as well as numerous anecdotes and stories on topics such as cycling, trains and Scandinavian place names, made the last three and a half years far more pleasurable than they had any right to be.

I would also like to thank my parents. The knowledge that I have the unconditional support of my closest family members has given me the confidence to try things I never thought I would be able to do. This thesis is proof of that.

I have been very fortunate with the friends I have made during my PhD. From lunch in the penthouse to haggis night and the pub quiz the atmosphere has always been enjoyable and supportive. I am also very grateful to my friends in South Shields, especially David and Paul who somehow managed to live with my eccentricities for the duration of my PhD.

Finally I would like to acknowledge the financial support I have recieved from the Engineering and Physical Sciences Research Council.

Abstract

Bayesian inference for such things as collections of related binomial or Poisson distributions typically involves rather indirect prior specifications and intensive numerical methods (usually Markov chain Monte Carlo) for posterior evaluations. As well as requiring some rather unnatural prior judgements this creates practical difficulties in problems such as experimental design. This thesis investigates some possible alternative approaches to this problem with the aims of making prior specification more feasible and making the calculations necessary for updating beliefs or for designing experiments less demanding, while maintaining coherence.

Both fully Bayesian and Bayes linear approaches are considered initially. The most promising utilises Bayes linear kinematics in which simple conjugate specifications for individual counts are linked through a Bayes linear belief structure. Intensive numerical methods are not required. The use of transformations of the binomial and Poisson parameters is proposed.

The approach is illustrated in two examples from reliability analysis, one involving Poisson counts of failures, the other involving binomial counts in an analysis of failure times. A survival example based on a piecewise constant hazards model is also investigated.

Applying this approach to the design of experiments greatly reduces the computational burden when compared to standard fully Bayesian approaches and the problem can be solved without the need for intensive numerical methods. The method is illustrated using two examples, one based on usability testing and the other on bioassay.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

This thesis is concerned with the analysis of collections of quantities with conditional binomial or Poisson distributions. Such situations occur frequently. General classes of problem include dose response studies for such things as quantal bioassays, time series analyses utilising log-linear and logistic-linear models and survival modelling in the analysis of, for example, grouped life tables. A specific example in Fa-Si-Oen & Pievatolo (2000) involved the numbers of ruptures in pipelines over six years. The pipelines were categorised into eight systems by depth, diameter and site and were further categorised by year. Each combination of system and year in these data corresponds to a Poisson random variable with a mean specific to that system and that year.

In Martz *et al.* (1996) the number of successful starts of each of a collection of emergency diesel generators (EDGs) for nuclear power stations followed a binomial distribution. The number of trials was the number of demands for that EDG and the unknown parameter of interest was the probability that the EDG started successfully.

We are concerned with a subjective analysis from the point of view of an interested party, termed "the expert", who has prior beliefs about the collection of unknowns in the analysis. The most widely used subjective approach to statistical inference is the Bayesian paradigm. This gives a full joint probability distribution to all unknowns in the analysis. Beliefs are then updated by conditioning on the observations and using Bayes theorem to form a posterior distribution.

The Bayesian approach is not the only one to take a subjective view of probability, however. Another such approach is found in Bayes linear statistics (Goldstein & Wooff, 2007). A Bayes linear analysis takes expectation, rather than probability, as its primitive. A partial prior specification is made for unknowns and beliefs are updated using

a process of linear fitting. A Bayes linear analysis can be viewed either as a useful approximation to a fully Bayesian analysis or as an alternative view of inference in which artificial distributional assumptions are not necessary.

Typically the individual binomial or Poisson parameters are not independent in the expert's prior beliefs. For example, in the case of the pipelines data, if a larger than expected number of ruptures were observed in one of the systems in the first year, this may very well lead to a revision upwards of the expected numbers of ruptures in the same system in subsequent years.

Typically such data are analysed using a generalised linear model with the linear predictors related via a linear model to a set of coefficients which are given a multivariate normal prior distribution. See, for example, Dellaportas & Smith (1993); Clayton (1996). Marginal predictive distributions are thus of rather complicated form, making prior elicitation difficult. Computation of posterior distributions requires numerical methods, usually Markov chain Monte Carlo (MCMC).

This can seem rather heavy handed for apparently simple problems and can become a major obstacle in areas such as experimental design (Müller, 1999). In such situations a method for analysing related Poisson and binomial distributions without the necessity for intensive numerical methods is desirable.

Initially we review the standard Bayesian generalised linear modelling approach. In the remainder of the chapter we consider a model for Poisson and binomial random variables from the area of Bayesian time series which is tractable and so does not require numerical methods, the power steady approach of Smith (1979, 1981). We then give an introduction to the elicitation of prior beliefs in Bayesian statistics and the chapter ends with an outline of the rest of the thesis.

1.2 Bayesian generalised linear models

Generalised linear models (GLMs) are used when standard linear regression techniques are inappropriate. More specifically, the approach is suitable when the response variable is discrete or has a range which is restricted. Two important cases of a discrete response variable are when such a variable follows a Poisson or binomial distribution. For information on generalised linear models see Nelder & Wedderburn (1972); McCullagh & Nelder (1989); Congdon (2001).

Suppose, more generally, we consider a distribution from the exponential family. That

is, a distribution whose density can be parameterised as

$$f(y \mid \lambda, \phi) = \exp\left\{\frac{y\lambda - b(\lambda)}{a(\phi)} + c(y, \phi)\right\},$$

for response variable y, where λ is the canonical or natural parameter, ϕ is the scale parameter and $a(), b(), c()$ are functions. To avoid repetition we will use the word "density" for either a probability density function for continuous random variables or a probability mass function for discrete random variables. The mean and variance of y (McCullagh & Nelder, 1989) are

$$\mu = \mathrm{E}(y \mid \lambda, \phi) = b^{'}(\lambda), \quad \mathrm{Var}(y \mid \lambda, \phi) = b^{''}(\mu).$$

We include covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^{'}$ for individuals $i = 1, \ldots, n$ by introducing a linear predictor $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^{'}$ such that

$$\eta_i = g(\mu_i),$$

where $g()$ is known as the link function. The linear predictor takes the form of a function of the covariates. The canonical or natural link function for an exponential family distribution (McCullagh & Nelder, 1989) is that which satisfies

$$\eta_i = b^{'-1}(\mu_i) = \lambda_i, \tag{1.1}$$

for each i.

Suppose now that counts $Y_1, \ldots, Y_n$ are observed and that our beliefs are such that, conditional on the values of unknown parameters $\theta_1, \ldots, \theta_n$, either

$$Y_i \sim \mathrm{Poisson}(\theta_i) \quad \text{or} \quad Y_i \sim \mathrm{bin}(N_i,\ \theta_i),$$

where $\mathrm{bin}(N_i,\ \theta_i)$ denotes a binomial distribution with known number of trials N_i, for $i = 1, \ldots, n$. Further suppose that $Y_i,\ Y_j$ are conditionally independent given $\theta_i,\ \theta_j$ for $i \neq j$.

Taking the Poisson distribution, its density is

$$f(y) = \frac{\theta^y e^{-\theta}}{y!} = \exp\left\{\frac{y \log \theta - \theta}{1} - \log y!\right\},$$

and so $\eta = \log \theta$, $b(\eta) = e^{\eta}$, $a(\phi) = 1$ and $c(y, \phi) = \log y!$. This gives $\mu = \theta = e^{\eta}$ and $b^{'}(\eta) = e^{\eta}$ and so the natural link function for the Poisson distribution is

$$\eta_i = \log \mu_i = \log \theta_i. \tag{1.2}$$

We can apply a similar process to the binomial distribution. The density of a bin(N, θ) is

$$f(y) = \begin{pmatrix} N \\ y \end{pmatrix} \theta^y (1-\theta)^{N-y} = \exp\left\{ \frac{y \log(\theta/1-\theta) - [-N \log(1-\theta)]}{1} - \log \begin{pmatrix} N \\ y \end{pmatrix} \right\},$$

and so $\eta = \log(\theta/1-\theta)$, $b(\eta) = N \log(e^\eta + 1)$, $a(\phi) = 1$ and $c(y, \phi) = \log \begin{pmatrix} N \\ y \end{pmatrix}$. This gives

$$\mu = b^{'}(\eta) = \frac{N e^\eta}{1 + e^\eta},$$

and so the natural link function for the binomial distribution is

$$\eta_i = \log\left(\frac{\mu_i}{N_i - \mu_i}\right) = \log\left(\frac{\theta_i}{1-\theta_i}\right), \tag{1.3}$$

from Equation 1.1 since $\mu_i = N_i \theta_i$, which is known as the logit link function.

Thus the link function $g()$ is used to transform the unknown binomial or Poisson parameters θ_i to the linear predictor

$$\eta_i = g(\theta_i),$$

with $-\infty < \eta_i < \infty$. Using the natural logarithm in the Poisson case as in Equation 1.2 leads to log-linear models (McCullagh & Nelder, 1989; Cameron & Trivedi, 1998).

In a binomial model as well as the logit link given in Equation 1.3, probit, $\eta_i = \Phi^{-1}(\theta_i)$, where $\Phi^{-1}()$ is the inverse of the standard normal distribution function, and complementary log-log, $\eta_i = \log(-\log[1-\theta_i])$, link functions are commonly used (McCullagh & Nelder (1989); Congdon (2001)).

The linear predictors $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)'$ are typically related via a linear model

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\gamma},$$

to a vector of unknown coefficients $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^{'}$ which are given a multivariate normal prior distribution. That is

$$\boldsymbol{\gamma} \sim \mathrm{N}(\boldsymbol{M}, \Sigma),$$

for prior mean vector $\boldsymbol{M}$ and variance matrix Σ.

This induces a multivariate normal prior distribution over $\boldsymbol{\eta}$. This non-conjugate structure makes prior elicitation awkward and requires intensive numerical methods for posterior computation. The design matrix $\mathbf{X}$ contains the values of the covariates.

In this standard approach the likelihood and prior distributions are not conjugate. Therefore numerical methods, usually MCMC, are necessary to evaluate posterior distributions. A methodology for such situations in which posterior distributions are tractable and hence numerical or simulation methods are not needed would greatly reduce the computational burden. This is particularly desirable in areas where the analysis of real problems can quickly become computationally infeasible, such as Bayesian experimental design.

Tractable methods for the analysis of Poisson and binomial distributions have been proposed previously in the area of Bayesian time series. One is the power steady model of Smith (1979, 1981). We review this approach in the following section.

West *et al.* (1985) introduced an approach to the analysis of non-Normal time series, based on dynamic linear models, called dynamic generalised linear models (DGLMs), which combine fully Bayesian conjugate updating with Bayes linear updating to provide a fully analytic analysis. Updating is not commutative in DGLMs however.

Settimi & Smith (2000) compared the DGLM approach of West *et al.* (1985) to a Bayes linear approach and another analytic approach based on the DGLM from Gargoum & Smith (1994, 1997). The comparison was with a full MCMC approach and they concluded that the DGLM is a good approximation to the "exact" MCMC solution.

1.3 Power steady models

In time series analysis, tractable models, named power steady models, have been developed for binomial and Poisson random variables amongst others (Smith, 1979, 1981). This is a generalisation of the steady model for Normal random variables given in Harrison & Stevens (1976). For a full definition of the power steady model see Smith (1979).

A steady model is one in which expectations for the parameters of interest remain constant over future time periods but variances increase to reflect increased uncertainty associated with moving further into the future.

The power steady model works in the following way. Suppose we have the time dependent parameter θ_t, $t = 1, 2, \ldots$ and, for observation sets $D_t = y_1, y_2, \ldots, y_t$, relevant densities of $\theta_t \mid D_t$ and $\theta_{t+1} \mid D_t$ are $f_{t|t}(\theta)$ and $f_{t+1|t}(\theta)$ respectively. Then a power steady model satisfies

$$f_{t+1|t}(\theta) \propto [f_{t|t}(\theta)]^k, \tag{1.4}$$

where $k \in [0, 1)$ is called the rate of the steady model. If k is not time dependent then this is a simple power steady model. We see from Equation 1.4 that the prior for θ_{t+1}

is given by the posterior for θ_t raised to some power less than one. The effect of this is that the variance of θ_t, and so the uncertainty associated with θ_t, increases with time.

If we take the beta-binomial case so that

$$\theta_t \mid D_t \sim \text{beta}(a_t, b_t),$$

then $f_{t|t}(\theta) \propto \theta^{a_t}(1-\theta)^{b_t}$. Application of the power steady model (Equation 1.4) yields

$$\begin{aligned} f_{t+1|t}(\theta) &\propto [\theta^{a_t}(1-\theta)^{b_t}]^k \\ &= \theta^{ka_t}(1-\theta)^{kb_t}, \end{aligned}$$

and so $\theta_{t+1} \mid D_t \sim$beta$(ka_t, kb_t)$. Plots of beta densities for different values of k are given in Figure 1.1 for some arbitrary parameter values $a = 15, b = 20$. We see the effect of the rate parameter; the smaller k becomes the more diffuse the density and so the higher the uncertainty associated with inferences and forecasts for later time periods.

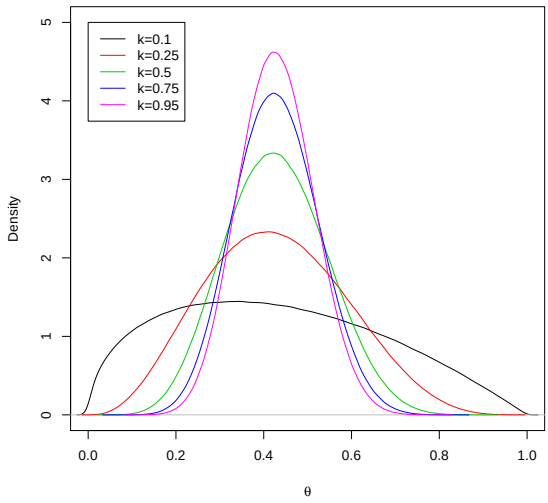

Figure 1.1: A plot of beta densities for beta$(15k, 20k)$ for different values of k

Observing y_{t+1} from

$$Y_{t+1} \mid \theta_{t+1} \sim \text{bin}(n_{t+1}, \theta_{t+1}),$$

leads to a tractable fully Bayesian update as the beta and binomial distributions are conjugate. Thus the posterior distribution is $\theta_{t+1} \mid D_{t+1} \sim$beta$(a_{t+1}, b_{t+1})$, where $a_{t+1} = ka_t + y_{t+1}$ and $b_{t+1} = kb_t + n_{t+1} - y_{t+1}$.

Smith (1979) also considers a gamma-Poisson model. In this case $\theta_t \mid D_t \sim$gamma(a_t, b_t) with density $f_{t|t}(\theta) \propto \theta^{a_t} e^{-b_t \theta}$. If we apply the power steady model (Equation 1.4) then the prior density for $\theta_{t+1} \mid D_t$ is

$$\begin{aligned} f_{t+1|t}(\theta) &\propto [\theta^{a_t} e^{-b_t \theta}]^k \\ &= \theta^{k a_t} e^{-k b_t \theta}, \end{aligned}$$

and so $\theta_{t+1} \mid D_t \sim$gamma$(ka_t, kb_t)$. Thus the parameters are updated by the power steady model in the same form as the beta-binomial case. A plot of the gamma distribution density is given for different values of k in Figure 1.2. The parameter values used in the plot are $a = 5, b = 1$. Once again as the rate parameter k decreases the gamma density becomes more diffuse. Thus the smaller k is made the more uncertainty is associated with the parameter θ for the next step.

However, in this case, there is also an effect on the location of the density. The mode decreases with decreasing k although the mean remains the same. This is in contrast to the beta densities of Figure 1.1 where, although the mode decreases, this decrease is far less pronounced.

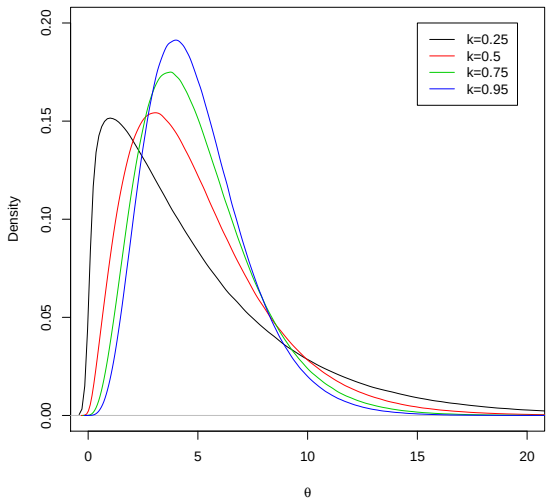

Figure 1.2: A plot of gamma densities for gamma$(5k, k)$ for different values of k

Observation of

$$Y_{t+1} \mid \theta_{t+1} \sim \text{Poisson}(\theta_{t+1}),$$

means we can apply Bayes theorem and perform a conjugate update to obtain $\theta_{t+1} \mid$

D_{t+1}. The resulting distribution is gamma(a_{t+1}, b_{t+1}), where

$$a_{t+1} = ka_t + y_{t+1} \quad b_{t+1} = kb_t + 1.$$

In this formulation a joint distribution for $\theta_1, \ldots, \theta_t$ has not been specified. Thus, without such a joint distribution, updates and forecasts more than one step ahead cannot be made.

The idea of steady state models is extended to the multivariate case, $\boldsymbol{\theta}_t = (\theta_{1t}, \ldots, \theta_{pt})'$ in Smith (1981). Initially he considered the extension given by

$$f_{t+1|t}(\boldsymbol{\theta}) \propto [f_{t|t}(\boldsymbol{\theta})]^k,$$

with $k \in (0, 1]$. He found that, in general, this leads to an update which is overly restrictive as it implies that information is lost about all of the parameters at the same rate.

Instead, the stacked power steady model was developed (Smith, 1981). Rather than a single rate constant k, a vector $\boldsymbol{k} = (k_1, \ldots, k_p)'$ is used so that each parameter evolves at a separate rate. If the parameters are ordered so that information is lost about θ_{1t} slowest and θ_{pt} most quickly the resulting stacked power steady model is:

(i) If k_i is associated with θ_{it} then $0 < k_i < k_{i+1} \leq 1$ for $i = 1, \ldots, p-1$.

(ii) Each $\theta_i \mid \theta_{i+1}, \ldots, \theta_p$ evolves as

$$f_{t+1|t}(\theta_i) = [f_{t|t}(\theta_i)]^{k_i},$$

i.e., as a simple power steady model.

Power steady models are very useful when considering a process with a natural ordering such as a time series. We wish to also model situations in which this is not the case and so we shalll not consider power steady models further in this thesis.

1.4 Elicitation of prior information

What is elicitation?

In order to perform a Bayesian, or Bayes linear, analysis it is necessary first to specify prior information for the unknowns in the analysis. In a full Bayesian analysis this takes the form of a joint prior distribution over all unknowns such as parameters. In a Bayes linear analysis this corresponds to a full second-order specification. This information

is found as the subjective beliefs of an 'expert'. The term 'expert' simply refers to the person or people from whom information is being elicited, usually somebody within the field in which the investigation is to take place. Elicitation is the process used to transform the expert's beliefs into prior distributions or moments.

Why is elicitation necessary?

As Kadane (1998) and Farrow (2003) note, expertise in a specific subject is not the same as expertise in statistics and probability. Therefore, though the expert has a great deal of knowledge of the subject in question, they invariably and understandably find it difficult to transfer that knowledge directly to specifying prior distributions for parameters. Garthwaite *et al.* (2005) describe the elicitation process as that of a facilitator helping the expert to express his current knowledge in probabilistic form.

Questions should, when possible, only be asked in terms of observable quantities (Kadane, 1998). It is up to the facilitator to choose which, and how many, quantities to elicit from the expert. It is important to educate the expert in the type of questions to be asked before the procedure begins in order that they aquire a feel for the process.

Which quantities should be elicited?

A joint prior distribution is chosen for the parameters of interest in a Bayesian analysis. The elicitation procedure is then often concerned with finding the values of the hyperparameters in these distributions. Unless these parameters have a direct meaning (for example, the mean and variance in a normal distribution) this is not possible to do directly. One way to overcome this is to elicit moments (mean, variance, etc.) directly and then convert these into parameter values. For symmetric distributions people tend to be able to estimate the mean well. This is not the case in skewed distributions, however, where estimates of modes and medians tend to be more accurate (Garthwaite *et al.*, 2005).

Non-statisticians (and, indeed, statisticians) are also very poor at determining variances accurately in general. As a result of this, the spread of a distribution is usually elicited via other quantities, typically credible intervals (Garthwaite *et al.*, 2005), most often in the form of quantiles (Garthwaite & O'Hagan, 2000). Commonly upper and lower quartiles are chosen but these tend to lead to overconfidence in the spread of the distribution as estimates of the interquartile range are generally too small. Instead of quartiles, Garthwaite & O'Hagan (2000) recommend using tertiles (33% and 67% points) as there is evidence this reduces overconfidence somewhat. This was also found by Peterson *et al.* (1972).

There are important psychological considerations to take into account when performing

elicitations. Issues such as how a question is asked and even in what order questions are asked can affect the answers given (Payne, 1951).

There are also certain inherent biases, often called heuristics, in how people assess probabilities which must be overcome in any elicitation. These are:

- judgement by representativeness,
- judgement by availability,
- anchoring.

Example 1: The following example of judgement by representativeness is given in Slovic (1972), taken from Kahneman & Tversky (1972). Participants were given the following description of a student.

'Tom W. is of high intelligence although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place. His writing is rather dull and mechanical, occasionally enlivened by somewhat corny puns and by flashes of imagination of the sci-fi type. He has a strong drive for competence. He seems to have little feel and little sympathy for other people, and does not enjoy interacting with others. Self-centered, he nonetheless has a deep moral sense.'

The participants (graduate psychology students) were then asked to rank the following subjects in order, with one being the most likely subject Tom is a graduate student in; business administration, computer sciences, engineering, humanities and education, law, library sciences, medicine, physical and life sciences and social science and social work.

Most people chose computer science and engineering as the most likely subjects. They did this by fitting the description to their stereotypes of the subjects. They ignored the base rates, i.e., the fact that there are relatively few engineering and computer science graduate students.

Example 2: Suppose you were given the following description of a place.

'It is so clean here, you could eat off the streets. Snow capped mountains frame picturesque log cabins, with geraniums in every window.'

Now suppose you are given a list of places which includes Europe, Brazil and Switzerland. When asked which of the places the letter was most likely written from, Bar-Hillel & Neter (1993) found that Switzerland was rated ahead of Europe. This is a second example of judgement by representativeness. People judge the likelihood of each of the places based on how well the description fits their image of that place. Once again base rates are ignored, specifically that Switzerland is part of Europe.

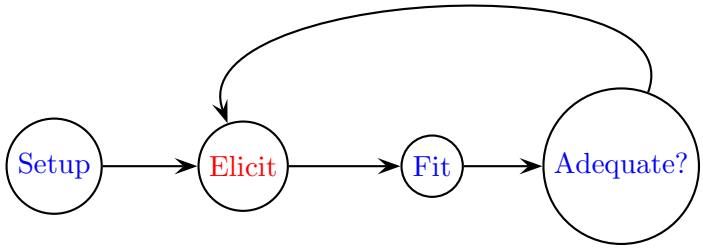

Figure 1.3: The four stages of the elicitation process

This example exibits a more serious failing than the last in that the answers violate a rule of probability, namely that if event A is a subset of event B then $\Pr(A) \leq \Pr(B)$.

Example 3: When asked to judge the probability of an event occuring people often base their estimates on how readily they can recall instances of that event taking place (Slovic, 1972). When asked whether the letter k is more likely to be the first or third letter in a word most subjects choose the first letter (Tversky & Kahneman, 1971). This is because it is easier to think of words which start with the letter k rather than have k as their third letter.

Actually k is far more likely to be the third letter in a word than the first (as in likely!). This is an example of judgement by availability.

Example 4: Anchoring occurs when subjects are given a value for something (their anchor) and then asked to adjust that value in the light of new information. The anchor could be explicit or implicit. An example of an explicit anchor would be:

Q1. Do you think Sunderland's football team will end up with more or fewer than 40 points this season?

Q2. How many points do you think Sunderland will end up with?

The first question provides the anchor value of 40 points. In answer to the second question the subject will adjust this value up or down depending on how they think Sunderland will do. It has been found, however, that this adjustment will not be large enough even though there is no implication that 40 is a sensible figure (Slovic, 1972).

Garthwaite *et al.* (2005) describe the elicitation process as consisting of four stages; setup, elicit, fit and adequacy. A representation of the process, also in the paper, is given in Figure 1.3.

The setup stage contains everything which needs to be done before conducting the

elicitation. It involves, amongst other things,

- identifying and recruiting the expert,
- identifying the unknowns in the analysis,
- formulating the questions,
- training the expert in probability.

In the elicit stage questions are put to the expert to obtain the required summaries. The fit stage then takes these summaries and fits a probability distribution to them.

The adequacy stage is concerned with whether the fitted distribution is an accurate reflection of the expert's beliefs. If they feel that it is then the elicitation is complete. If not the process returns to the elicit stage for re-evaluation of the elicited summaries.

We investigate elicitation for the beta binomial model in Section 4.2.1. The information in this section shall be used later in the thesis when we consider which quantities to elicit for specific models.

1.5 Thesis outline

In this thesis we consider subjectivist approaches to the problem of related binomial and Poisson distributions. The aim of the thesis is to find such an approach in which

- intensive numerical or simulation based methods are not required in the calculation of posterior quantities,
- a careful assessment of genuine prior beliefs for the unknowns in the analysis can be made, and
- realistically complex problems can be solved within a reasonable time frame in the area of Bayesian experimental design.

The remainder of the thesis is structured as follows. In Chapter 2 we consider fully Bayesian approaches to the problem in two dimensions. Most of the models in this chapter are based on the introduction of density multipliers, a method for constructing joint densities in which correlations are induced between parameters whilst the conjugacy of the Bayesian updates is preserved. Two specific types of density multiplier are considered, copula functions and mixtures. An example is given in each case.

Chapter 3 investigates Bayes linear approaches to the two parameter problem. Initially Bayes linear methods (Goldstein & Wooff, 2007) are introduced. We consider a

model for count data which takes advantage of ideas of exchangeability between variables. Bayes linear kinematics (Goldstein & Shaw, 2004), a Bayes linear equivalent of probability kinematics, is then utilised. We apply Bayes linear kinematic updating to the binomial and Poisson parameters and suggest suitable transformations to these parameters in order to increase the effectiveness and suitability of the updating.

In Chapter 4 we extend the problem of related binomial and Poisson distributions into more than two parameters and then apply the most promising approach from the two preceding chapters, that of Bayes linear kinematics performed on the transformed parameters, in the case of the binomial distribution. Special attention is paid to ensuring a commutative solution exists. An example involving the effects of smoking on health is considered. An approach to the specification of a coherent covariance structure for the transformed parameters based upon ideas in Farrow (2003) is used.

Chapter 5 is concerned with the application of the generalised Bayes linear kinematic approach to reliability and survival analysis. Two applications in the area of reliability are investigated. The first involves Poisson counts of failures and the second binomial counts in the analysis of failure times. The methodology is applied to an example in both cases. The remaining part of the chapter considers a piecewise constant hazards model in which hazards for different individuals are considered proportional. In particular we present a commutative solution in contrast to the non-commutative solution of Gamerman (1991).

In Chapter 6 we apply the Bayes linear kinematic methodology to Bayesian experimental design. Two problems are considered; usability testing and bioassay. In the usability testing application we give a Bayes linear solution to the problem explored by Valks (2005) using fully Bayesian methods. Within the context of bioassay we provide solutions to both the sample size and design point problems simultaneously. To do this we introduce the Bayes linear kinematic benefit utility.

In Chapter 7 we give some conclusions and areas for further work.

Chapter 2

Fully Bayesian approaches

2.1 Introduction

In this chapter we consider the problem of two correlated binomial probabilities or Poisson parameters. We consider models based on fully Bayesian approaches. Most of the approaches take a joint density for parameters in the form of a density multiplier, a function which, when multiplied by the marginal densities, induces a correlation between the parameters. Within this context two classes of density multipliers are considered; copula functions and mixtures. The copula family investigated is the Farlie-Gumbel-Morgenstern family and some extensions of this. In terms of mixtures for the binomial distribution we consider mixture distributions of beta densities and in the Poisson case the result is mixtures of gamma densities. We see that copula functions are a special case of multipliers which obey a marginality property. We also consider a Dirichlet model in the binomial case. We illustrate the methodology developed using a binomial example involving patients who have had heart attacks and a Poisson example involving the numbers of failures of piston rings in compressors.

2.2 Density Multipliers

Let us suppose that counts X_1, X_2 are observed and that our beliefs are such that, conditional on the values of the unknown parameters θ_1, θ_2, either

$$X_i \sim \text{bin}(n_i, \theta_i) \quad \text{or} \quad X_i \sim \text{Po}(\theta_i),$$

for $i = 1, 2$ where X_1, X_2 are conditionally independent given θ_1, θ_2.

We can give θ_1, θ_2 conjugate prior distributions. So for X_i binomially distributed we use

a beta prior distribution $\theta_i \sim$beta(a_i, b_i) and if X_i is Poisson we have a gamma prior distribution $\theta_i \sim$gamma(a_i, b_i). Then, if θ_1, θ_2 were independent, their joint density would take the form

$$f_0(\theta_1, \theta_2) = f_{01}(\theta_1) f_{02}(\theta_2),$$

where $f_{0i}(\theta_i)$ is the appropriate beta or gamma density.

However, if we believe θ_1, θ_2 to be dependent, then we must find some other way to represent their joint density. One possibility would be to use a function $g(\theta_1, \theta_2)$ so that

$$f_0(\theta_1, \theta_2) \propto f_{01}(\theta_1) f_{02}(\theta_2) g(\theta_1, \theta_2).$$

We call the function $g(\theta_1, \theta_2)$ a *density multiplier*. It is used to induce a correlation between θ_1 and θ_2.

Now let us suppose that we observe $X_1 = x_1$ and $X_2 = x_2$. In our beta-binomial setup the prior marginal densities are given by

$$f_{0i}(\theta_i) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} \theta_i^{a_i - 1} (1 - \theta_i)^{b_i - 1}, \tag{2.1}$$

and the likelihood is

$$L(x_1, x_2 \mid \theta_1, \theta_2) = \begin{pmatrix} n_1 \\ x_1 \end{pmatrix} \theta_1^{x_1} (1 - \theta_1)^{n_1 - x_1} \begin{pmatrix} n_2 \\ x_2 \end{pmatrix} \theta_2^{x_2} (1 - \theta_2)^{n_2 - x_2}. \tag{2.2}$$

Using Bayes theorem we obtain the posterior joint density for θ_1, θ_2,

$$\begin{aligned} f_1(\theta_1, \theta_2) &= f_0(\theta_1, \theta_2) L(x_1, x_2 \mid \theta_1, \theta_2) \\ &\propto \theta_1^{a_1 + x_1 - 1} (1 - \theta_1)^{b_1 + n_1 - x_1 - 1} \theta_2^{a_2 + x_2 - 1} (1 - \theta_2)^{b_2 + n_2 - x_2 - 1} \times g(\theta_1, \theta_2) \\ &\propto \theta_1^{A_1 - 1} (1 - \theta_1)^{B_1 - 1} \theta_2^{A_2 - 1} (1 - \theta_2)^{B_2 - 1} \times g(\theta_1, \theta_2), \end{aligned}$$

where $A_1 = a_1 + x_1$, $B_1 = b_1 + n_1 - x_1$, $A_2 = a_2 + x_2$ and $B_2 = b_2 + n_2 - x_2$.

Similarly, if we take the Poisson-gamma setup the prior density of θ_i is

$$f_{0i}(\theta_i) = \frac{b_i^{a_i} \theta_i^{a_i - 1} e^{-b_i \theta_i}}{\Gamma(a_i)},$$

and the likelihood is

$$L(x_1, x_2 \mid \theta_1, \theta_2) = \frac{\theta_1^{x_1} e^{-\theta_1}}{x_1!} \times \frac{\theta_2^{x_2} e^{-\theta_2}}{x_2!}. \tag{2.3}$$

Thus the posterior density of θ_1, θ_2, found by applying Bayes theorem, is

$$\begin{aligned} f_1(\theta_1, \theta_2) &\propto \theta_1^{a_1-1} e^{-b_1\theta_1} \theta_2^{a_2-1} e^{-b_2\theta_2} g(\theta_1, \theta_2) \times \theta_1^{x_1} e^{-\theta_1} \theta_2^{x_2} e^{-\theta_2} \\ &\propto \theta_1^{A_1-1} e^{-B_1\theta_1} \theta_2^{A_2-1} e^{-B_2\theta_2} \times g(\theta_1, \theta_2), \end{aligned}$$

where $A_1 = a_1 + x_1$, $B_1 = b_1 + 1$, $A_2 = a_2 + x_2$ and $B_2 = b_2 + 1$. Thus we see that, in both cases, the posterior density is given by

$$f_1(\theta_1, \theta_2) \propto f_{11}(\theta_1) f_{12}(\theta_2) g(\theta_1, \theta_2),$$

where $f_{1i}(\theta_i)$ follow the same distributions as their prior counterparts but with a_i and b_i updated to A_i and B_i. We also see that the density multiplier, $g(\theta_1, \theta_2)$, is unaffected when data are observed and so multiple updates are possible without the loss of conjugacy.

We must now find suitable functional forms for the density multiplier $g(\theta_1, \theta_2)$ to take. There is also the question of how to measure association for variables on $[0, 1]$. We review methods of association in the following section.

2.3 Measures of Association

Initially let us consider the bivariate Normal distribution. If $\boldsymbol{X} = (X_1, X_2)'$ then

$$\boldsymbol{X} \sim \text{BVN}(\boldsymbol{\mu}, \Sigma),$$

where

$$\boldsymbol{\mu} = (\mu_1, \mu_2)' \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_2^2 \end{pmatrix}.$$

Thus the bivariate Normal distribution is defined by 5 parameters; 2 means μ_1, μ_2, 2 variances σ_1^2, σ_2^2 and a product-moment correlation (or covariance σ_{12}^2). The product-moment correlation is found from the covariance and variances as

$$\rho_{12} = \frac{\sigma_{12}^2}{\sqrt{\sigma_1^2 \sigma_2^2}}.$$

For variables on $(-\infty, \infty)$, the product-moment correlation is a widely used measure of association. However, this simple approach may not be appropriate for situations where the variables of interest are on a restricted domain such as $0 < \theta_i < 1$ for $i = 1, 2$.

We shall consider four alternative measures of association; Kendall's τ, Spearman's ρ,

the Pearson (product-moment) correlation applied to transformations of the parameters and a specification made directly in terms of observables.

2.3.1 Kendall's τ

Suppose $(\theta_1, \theta_2)'$ have some bivariate distribution. Consider observing a sequence of independent draws $(T_{1j}, T_{2j})'$, $j = 1, 2, 3, \ldots$ from this distribution.

Then Kendall's τ (Kendall, 1938; Kruskal, 1958) is defined as

$$\tau_{12} = \Pr\{(T_{11} - T_{12})(T_{21} - T_{22}) > 0\} - \Pr\{(T_{11} - T_{12})(T_{21} - T_{22}) < 0\}.$$

That is, Kendall's τ is given by the probability of concordance minus the probability of discordance, where the pairs of observations are concordant if $T_{11} > T_{12}$ and $T_{21} > T_{22}$ or $T_{11} < T_{12}$ and $T_{21} < T_{22}$ and discordant if the second inequality in each condition is reversed.

Equivalently,

$$\tau_{12} = 2\Pr\{(T_{11} - T_{12})(T_{21} - T_{22}) > 0\} - 1.$$

2.3.2 Spearman's ρ

If a third draw is made and T_{12} and T_{23} are independent then Spearman's ρ (Spearman, 1904; Kruskal, 1958) can be defined as

$$\rho_{12} = 3\left[\Pr\{(T_{11} - T_{12})(T_{21} - T_{23}) > 0\} - \Pr\{(T_{11} - T_{12})(T_{21} - T_{23}) < 0\}\right].$$

The interpretation is not as straightforward as with Kendall's τ. However, Spearman's ρ is proportional to the probability of concordance minus the probability of discordance for (T_{11}, T_{12}) and (T_{21}, T_{23}).

2.3.3 Transformations

Let us define the transformed quantities $(\eta_1, \eta_2)'$ as

$$\eta_i = g(\theta_i),$$

where $g()$ is a suitable transformation such as logit, probit or complementary log-log in the binomial case or natural logarithm in the Poisson model. We could then simply calculate the Pearson product-moment correlation of the transformed quantities.

The choice of transformation is a little arbitrary, at least in the binomial case, however.

We would also need to consider how an elicitation would be carried out in this case.

2.3.4 Directly from observables

In the case of related binomial parameters we could consider a method based on observables. Consider individual Bernoulli trials where for trial j

$$X_{ij} = \begin{cases} 1, & \text{with probability } \theta_i, \\ 0, & \text{with probability } 1 - \theta_i, \end{cases}$$

for $i = 1, 2$. Expectations in terms of the Bernoulli variables are given by

$$\begin{aligned} \mathrm{E}(X_{ij}) &= \mathrm{E}(\theta_i), \\ \mathrm{E}(X_{i1}, X_{i2}) &= \mathrm{E}(\theta_i^2), \\ \mathrm{E}(X_{11}X_{21}) &= E(\theta_1\theta_2). \end{aligned}$$

The required variances and covariance are then

$$\begin{aligned} \mathrm{Var}(\theta_i) &= \mathrm{E}(X_{i1}X_{i2}) - \mathrm{E}(X_{i1})^2, \\ \mathrm{Cov}(\theta_1, \theta_2) &= E(X_{11}X_{21}) - \mathrm{E}(X_{11})E(X_{21}), \end{aligned}$$

and so the product-moment correlation can be found.

This method would appear to argue in favour of the simple product-moment correlation for $(\theta_1, \theta_2)'$. However, would such an elicitation work in practice bearing in mind the mean-variance relationship for Bernoulli and binomial variables? Also, can we relate these moments to parameters of tractable joint distributions?

2.4 Copulas

2.4.1 Introduction

A copula (Nelson, 1999, 2006) is a joint distribution function $G(u_1, \ldots, u_n)$ for a number of random variables $U_1, \ldots, U_n$ where each marginal distribution is $U_i \sim \mathrm{U}(0, 1)$. The corresponding joint probability density function is

$$g(u_1, \ldots, u_n) = \frac{\partial}{\partial u_1} \cdots \frac{\partial}{\partial u_n} G(u_1, \ldots, u_n).$$

Copulas also obey certain properties. It follows from the definitions above that

1. $G(0,\ldots,0)=0$ and $G(1,\ldots,1)=1$,
2. $G(u_1,\ldots,u_n)=0$ if at least one of $u_1,\ldots,u_n$ is 0.

The requirement that each marginal is uniform means that $G(1,\ldots,1,u_i,1,\ldots,1)=u_i$. That is, the value of the distribution function at a point where every variable takes the value 1 except for u_i is u_i. To see this, consider, without loss of generality, the case of U_n. The marginal distribution function of U_n is

$$\begin{aligned} H_n(u_n) &= \int_0^1 \cdots \int_0^1 \int_0^{u_n} g(u_1,\ldots,u_{n-1},u_n^*)\, du_1 \cdots du_{n-1} du_n^*, \\ &= G(1,\ldots,u_n). \end{aligned}$$

If this is to be the distribution function of a $\mathrm{U}(0,1)$ distribution then it must be the case that $H_n(u_n)=u_n$.

A copula must also be n-increasing. This is the n-dimensional analogue of a one-dimensional non-decreasing function. In the 2-dimensional case this requirement is satisfied as long as

$$G(u_{12},u_{22})-G(u_{12},u_{21})-G(u_{11},u_{22})+G(u_{11},u_{21})\geq 0$$

for $u_{11}\leq u_{12}$ and $u_{21}\leq u_{22}$ (Nelson, 1999, 2006).

It is possible to choose a copula function in such a way that $U_1,\ldots,U_n$ are not independent even though all of the marginals are uniform.

If $X_1,\ldots,X_n$ are some continuous random variables with distribution functions $F_1(x_1),\ldots,F_n(x_n)$, then

$$U_i=F_i(X_i)$$

where $U_i\sim \mathrm{U}(0,1)$. Then $U_1,\ldots,U_n$ can be linked through a suitable copula. If $X_1,\ldots,X_n$ are dependent then $U_1,\ldots,U_n$ are also dependent. The distribution function of $U_1,\ldots,U_n$ is then $G(u_1,\ldots,u_n)$.

Clearly U_i is a strictly non-decreasing function of X_i so $U_i\leq u_i \Leftrightarrow X_i\leq x_i$. Thus we can also make G the distribution function of $F_1(X_1),\ldots,F_n(X_n)$. That is, the distribution function of $X_1,\ldots,X_n$ is $G[F_1(x_1),\ldots,F_n(x_n)]$. If this is differentiated with respect to $x_1,\ldots,x_n$ the joint probability density function of $X_1,\ldots,X_n$ is obtained. This is

$$f_{\underline{X}}(x_1,\ldots,x_n)=f_1(x_1)\cdots f(x_n)g[F_1(x_1),\ldots,F_n(x_n)] \tag{2.4}$$

where $f_i(x_i)$ is the marginal probability density function of X_i. We can show that the marginal distribution of X_i, under this structure, is $F_i(x_i)$. We know from above that

$X_i < x_i \Leftrightarrow U_i < u_i$. Then,

$$\Pr(X_i < x_i) = \Pr(U_i < u_i) = u_i = F_i(x_i). \tag{2.5}$$

2.4.2 Families of copula functions

There are many different families of copula functions. Two of the most widely used bivariate families are the Farlie-Gumbel-Morgenstern (FGM) family (Farlie, 1960; Gumbel, 1958; Morgenstern, 1956) and the Archimedian family. We shall now consider these families in two dimensions.

Farlie-Gumbel-Morgenstern family

This is a polynomial copula family in which members take the form of quadratic functions of the two random variables. They have been widely used in many different areas (Balakrishnan & Lai, 2009) due to their simple analytic form. The copula family is given by

$$G(u_1, u_2) = u_1 u_2 + \lambda u_1 u_2 (1 - u_1)(1 - u_2),$$

for some parameter λ which controls the strength of the dependence between U_1 and U_2. The parameter λ is constrained to $[-1, 1]$ so that G is a valid distribution function.

Archimedian family

Copulas in this family take the form

$$G(u_1, u_2) = \phi^{-1}[\phi(u_1) + \phi(u_2)],$$

where ϕ is called the generator function and must satisfy $\phi(1) = 0$, $\lim_{x \to 0} \phi(x) = \infty$, $\phi^{'}(x) < 0$ and $\phi^{''}(x) > 0$. Some commonly used generators and their copulas are;

(i) The Clayton copula (Clayton, 1978), also known as the Cook-Johnson copula (Cook & Johnson, 1981): This takes the generator $\phi(x) = \frac{1}{\theta}(x^{-\theta} - 1)$ resulting in the copula

$$G(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}},$$

where $\theta \in [-1, \infty) \backslash \{0\}$ is the parameter which controls the dependence between u_1 and u_2.

(ii) The Gumbel copula (Gumbel, 1960): This copula has a logarithmic generator

$\phi(x) = (-\log(x))^\theta$ which leads to the copula

$$G(u_1, u_2) = \exp\{-[(-\log u_1)^\theta + (-\log u_2)^\theta]^{\frac{1}{\theta}}\},$$

for $\theta \in [1, \infty)$.

(iii) The Frank copula (Frank, 1979): Once again a logarithmic generator function is used. On this occasion $\phi(x) = \log\left(\frac{e^{-\theta x}-1}{e^{-\theta}-1}\right)$ and so

$$G(u_1, u_2) = \frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right\}$$

with support $(-\infty, \infty)\backslash\{0\}$ for θ.

2.5 Copulas for counts

Let us return to counts X_1 and X_2 and parameters θ_1 and θ_2 which follow either the beta-binomial model or gamma-Poisson model of Section 2.2. Then associated with θ_1 and θ_2 are their distribution functions, $U = F_{01}(\theta_1)$ and $V = F_{02}(\theta_2)$, which we can link together using a suitable copula $C(u, v)$.

The joint prior density of θ_1 and θ_2 is then

$$f_0(\theta_1, \theta_2) = f_{01}(\theta_1) f_{02}(\theta_2) c(u, v), \tag{2.6}$$

where $c(u, v)$ is the copula density associated with $C(u, v)$. Thus we see that copulas are a special case of density multiplier where the normalising constant is equal to one.

This means, by the property given in Equation 2.5, that the prior expectations and variances of θ_1 and θ_2 are simply those of the marginal beta and gamma distributions. In the beta-binomial model they are

$$\mathrm{E}_0(\theta_i) = \frac{a_i}{a_i + b_i}, \quad \mathrm{Var}_0(\theta_i) = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)},$$

for $i = 1, 2$, and in the gamma-Poisson setup

$$\mathrm{E}_0(\theta_i) = \frac{a_i}{b_i}, \quad \mathrm{Var}_0(\theta_i) = \frac{a_i}{b_i^2}.$$

Thus, using copulas, prior marginality is preserved. If we observe $X_1 = x_1$ and $X_2 = x_2$, i.e., one of the likelihoods in Equations 2.2 and 2.3, then in either case we obtain a

joint posterior density of

$$\begin{aligned} f_1(\theta_1, \theta_2) &\propto f_0(\theta_1, \theta_2) L(x_1, x_2 \mid \theta_1, \theta_2) \\ &= f_{01}(\theta_1) f_{02}(\theta_2) c(u, v) L(x_1, x_2 \mid \theta_1, \theta_2) \\ &= f_{11}(\theta_1) f_{12}(\theta_2) c(u, v). \end{aligned} \tag{2.7}$$

Notice here that U and V are still the prior distribution functions of θ_1 and θ_2 and so $c(u, v)$ is a copula density for the prior not the posterior. Thus the distribution is still conjugate. However, we no longer have the marginality property so summaries and predictive distributions are no longer straightforward and the normalising constant is no longer equal to one.

In both cases the likelihood is the same as in Section 2.2 and so $f_{11}(\theta_1)$ and $f_{12}(\theta_2)$ are of the same form as $f_{01}(\theta_1)$ and $f_{02}(\theta_2)$ but with a_i and b_i updated to A_i and B_i.

2.6 The Farlie-Gumbel-Morgenstern Copula for counts

Suppose we have either the beta-binomial or Poisson-gamma setup explored in the previous section. Then, under the Farlie-Gumbel-Morgenstern (FGM) copula, the joint distribution function for θ_1 and θ_2 is

$$C(u, v) = uv[1 + \lambda(1 - u)(1 - v)],$$

for $-1 < \lambda < 1$. The copula density is found by differentiating this quantity to give

$$c(u, v) = 1 + \lambda(1 - 2u)(1 - 2v).$$

The joint prior density for θ_1 and θ_2 is given in Equation 2.6. It is therefore

$$f_0(\theta_1, \theta_2) = f_{01}(\theta_1) f_{02}(\theta_2)[1 + \lambda(1 - 2u)(1 - 2v)], \tag{2.8}$$

for the relevent prior densities $f_{01}(\theta_1)$ and $f_{02}(\theta_2)$.

2.6.1 Specification of prior parameters

Let us first consider specifying λ. We consider two of the measures of association for constrained variables considered in Section 2.3 (Nelson, 1999, 2006), Kendall's τ and Spearman's ρ. Kendall's τ can be expressed for copula functions as

$$\tau_{\theta_1, \theta_2} = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

The equation for Spearman's ρ can be similarly expressed and is given by

$$\rho_{\theta_1,\theta_2} = 12 \int_0^1 \int_0^1 C(u,v)dudv - 3.$$

For further explanation of this see Schweizer & Wolff (1981). We choose to use Spearman's ρ in order to specify the association parameter λ. Substituting the FGM copula into this integral we see that

$$\begin{aligned} \int_0^1 \int_0^1 C(u,v)dudv &= \int_0^1 \int_0^1 [uv + \lambda(u-u^2)(v-v^2)]dudv \\ &= \int_0^1 [\frac{1}{2}v + \frac{1}{6}\lambda(v-v^2)]dv \\ &= \frac{1}{4} + \frac{1}{36}\lambda \end{aligned}$$

Thus $\rho_{\theta_1,\theta_2} = \frac{1}{3}\lambda$ and so, since $\lambda \in [-1,1] \Rightarrow \rho_{\theta_1,\theta_2} \in [-\frac{1}{3}, \frac{1}{3}]$. Hence for the FGM copula it is not possible to specify a prior correlation of greater than $1/3$ or less than $-1/3$. This is clearly not desirable. This will also have an effect on the possible values for the posterior correlation and it would seem reasonable to assume that these would be even more restricted than the prior correlation. We shall discuss this further in Section 2.8.

2.6.2 The Posterior Density

Application of Bayes theorem on observation of data leading to one of the likelihoods in Equation 2.2 or Equation 2.3 leads to a posterior density of the form of Equation 2.7, namely,

$$f_1(\theta_1,\theta_2) \propto f_{11}(\theta_1)f_{12}(\theta_2)[1+\lambda(1-u)(1-v)], \tag{2.9}$$

where u and v are the prior distribution functions of θ_1 and θ_2. Thus the marginality property is lost and the integrating constant must be found before posterior moments can be calculated.

2.6.3 Example: A Clinical Trial

The Anturane Reinfarction Trial Research Group (1980) reported a clinical trial on the use of the drug sulfinpyrazone in patients who had suffered myocardial infarctions (heart attacks). The data were reproduced in Hand *et al.* (1994). The idea was to see whether the drug had an effect on the number dying within a certain time. Patients in one group were given the drug while patients in another group were given a placebo (inactive substitute).

Table 2.1 gives the number of all analysable deaths up to 24 months after the myocardial infarction and the total number of eligible patients who were not withdrawn and did not suffer a non-analysable death during the study.

	Deaths	Total
Sulfinpyrazone (group 1)	44	560
Placebo (group 2)	62	540

Table 2.1: Post heart attack deaths

This situation can be represented by saying that there are two groups, containing n_1 and n_2 patients, and two parameters, θ_1 and θ_2, such that, given these parameters, the number of deaths X_i in group i is distributed as

$$X_i \mid \theta_i \sim \text{bin}(n_i, \theta_i). \tag{2.10}$$

Natural selections of prior distributions for the parameters θ_i are

$$\theta_i \sim \text{beta}(a_i, b_i), \tag{2.11}$$

but it seems reasonable that the prior beliefs would be such that θ_1 and θ_2 would not be independent. That is, if we observe a number of deaths in group i then this will cause us to revise our beliefs about the probability of death in the other group as well as group i.

Let us suppose that our prior beliefs are such that we wish to specify a prior Spearman's Rho of $1/3$ between θ_1 and θ_2 which is achieved by setting $\lambda = 1$. Thus

$$C(u, v) = uv(1 + (1 - u)(1 - v)).$$

We now turn to the marginal specification, that of prior values for a_1, b_1, a_2, and b_2. In order to do this suppose that past evidence suggests that a suitable symmetric probability interval for θ_2 is

$$\Pr(\theta_2 < 0.06) = \Pr(\theta_2 > 0.19) = 0.025,$$

and so $\Pr(\theta_2 < 0.19) = 0.975$. Thus we have two equations with two unknowns, namely

$$\int_0^{0.06} f_{02}(\theta_2) d\theta_2 = 0.025,$$

and

$$\int_0^{0.19} f_{02}(\theta_2)d\theta_2 = 0.975,$$

where $f_{02}(\theta_2)$ is the beta density given in Equation 2.1. We can use these two equations in two unknowns to find values for a_2 and b_2. Iterative methods are used leading to the parameter values of

$$a_2 = 10.72, \quad b_2 = 80.84.$$

Suppose that *a priori* our beliefs about θ_1 and θ_2 are that θ_1 has the same mean and a standard deviation double that of θ_2 to reflect the greater uncertainty that is felt about the new drug.

As we have seen, one useful property of copulas is that they preserve the marginality of prior means and variances for θ_1 and θ_2. Thus it is known that

$$\mathrm{E}_0(\theta_2) = \frac{a_2}{a_2 + b_2} = 0.117,$$

and

$$\mathrm{Var}_0(\theta_2) = \frac{a_2 b_2}{(a_2 + b_2)^2(a_2 + b_2 + 1)} = 0.0011.$$

This gives a standard deviation for θ_2 of 0.0334. Doubling this, a standard deviation for θ_1 is aquired and squaring this enables a variance to be found for θ_1 which leads to two equations;

$$\mathrm{E}_0(\theta_1) = \frac{a_1}{a_1 + b_1} = 0.117,$$

and

$$\mathrm{Var}_0(\theta_1) = \frac{a_1 b_1}{(a_1 + b_1)^2(a_1 + b_1 + 1)} = 0.0045,$$

for a_1 and b_1 which can be solved resulting in

$$a_1 = 2.59, \quad b_1 = 19.55.$$

It is concluded that the prior marginal distributions are

$$\theta_1 \sim \mathrm{beta}(2.59, 19.55), \quad \theta_2 \sim \mathrm{beta}(10.72, 80.84).$$

We have now specified the prior joint density given in Equation 2.8 exactly. We can plot it as in Figure 2.1. All figures and calculations in this thesis are performed in R (R Development Core Team, 2011) unless otherwise stated. From the plot we can see the dependence between θ_1 and θ_2 has produced a skewed joint density. The values of θ_1 and θ_2 at the mode of this joint density are both much lower than their prior expectations of 0.125. This skewness is also evident if the marginal densities are plotted.

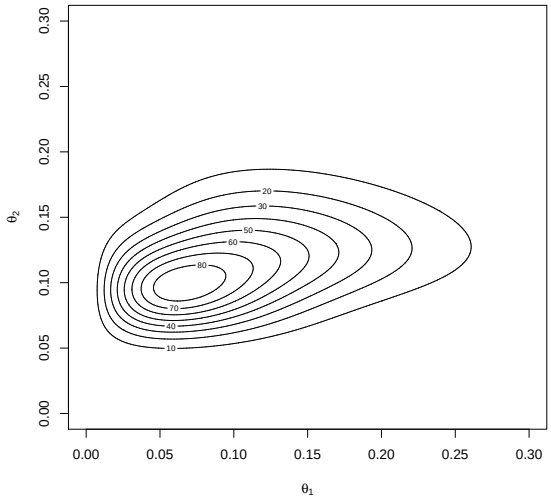

Figure 2.1: A plot of the joint prior density for the FGM copula model

The posterior joint density takes the form of Equation 2.9 up to proportionality with $A_1 = 46.59, B_1 = 535.55, A_2 = 72.72, B_2 = 558.84$. Having found the posterior joint density a posterior contour plot can be produced, as in Figure 2.2.

It would appear from the contour plot that the probability of death is higher in group 2 than group 1 and so sulfinpyrazone is of some benefit for patients who have had heart attacks. If we compare this plot to that of the joint prior density given in Figure 2.1 we can see that the posterior is far less skewed. There also appears to have been a significant reduction in correlation between θ_1 and θ_2 from prior to posterior. The reduction in uncertainty can be seen if we plot the prior and posterior joint densities on the same axes as in Figure 2.3.

Posterior means and variances for θ_1 and θ_2 are given in Table 2.2.

i	$\mathrm{E}_1(\theta_i)$	$\mathrm{Var}_1(\theta_i)$
1	0.081	1.27×10^{-4}
2	0.114	1.56×10^{-4}

Table 2.2: Posterior means and variance for θ_1 and θ_2

We see from the table that the posterior mean probability of death for patients taking sulfinpyrazone is lower than that of patients taking the placebo. Both variances have decreased significantly from prior to posterior.

To calculate these quantities we used Maple$^{\mathrm{TM}}$(Monagan *et al.*, 2005) to compute the

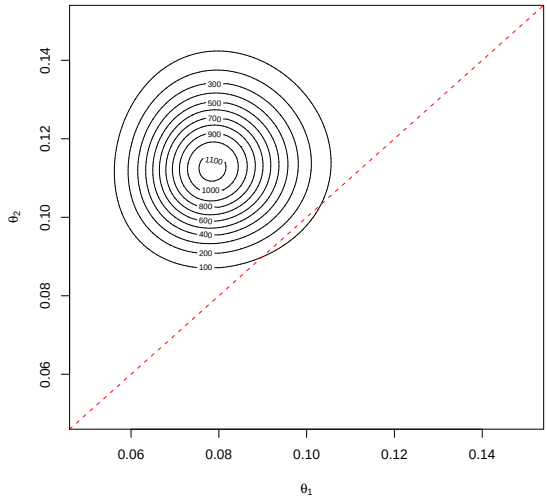

Figure 2.2: A posterior contour plot for the FGM copula

integrals. First A_1, A_2, B_1, B_2 were rounded to the nearest integer. Then the required densities are defined in terms of polynomials in θ_1, θ_2 and u and v. Now, u and v are proportional to partial beta functions, i.e.,

$$\int_0^{\theta} y^{a-1}(1-y)^{b-1}dy,$$

which are finite polynomials. Hence the integrating constant is a finite sum and posterior moments are then ratios of finite sums and can be found approximately in Maple$^{\text{TM}}$. The same method can be used to calculate the prior product moment correlation. For a Spearman's Rho of 1/3 the prior product moment correlation is 0.298.

2.6.4 Example: piston ring failures

We consider a subset of the data given in Davies & Goldsmith (1972), reproduced in Hand *et al.* (1994), on the numbers of failures of piston rings in steam driven compressors. The number of failures in two of the compressors over some time period are given in Table 2.3.

Compressor	Failures
1	46
2	33

Table 2.3: Piston ring failures in compressors 1 and 2

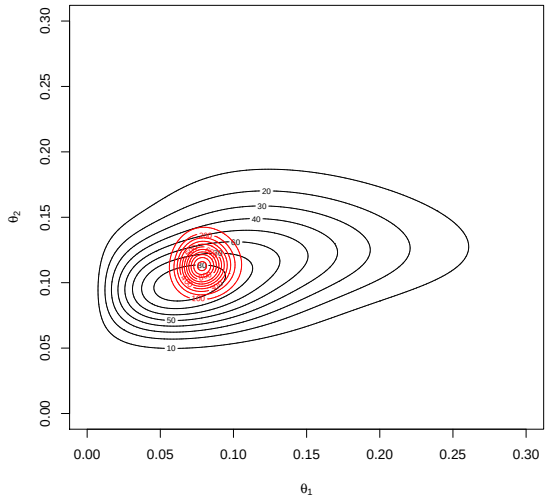

Figure 2.3: Prior and posterior contour plots for the FGM copula, clinical trial example

The number of failures in compressor $i = 1, 2$, X_i, follows a Poisson distribution

$$X_i \mid \theta_i \sim \text{Po}(\theta_i),$$

with expected number of failures θ_i. In order to perform an analysis using the FGM copula we give each θ_i the conjugate prior distribution,

$$\theta_i \sim \text{gamma}(a_i, b_i).$$

Suppose we wish to specify a prior Spearman's ρ between θ_1 and θ_2 of $\frac{1}{4}$. The FGM copula is then

$$C(u, v) = uv\left[1 + \frac{3}{4}(1-u)(1-v)\right].$$

In terms of the gamma distribution parameters, a_i and b_i, suppose we have no reason to believe either compressor would be more prone to failures than the other. Further suppose that our prior beliefs are such that

$$\text{E}_0(\theta_i) = 30, \quad \text{Var}_0(\theta_i) = 30, \tag{2.12}$$

for $i = 1, 2$. For a gamma distribution with mean m_i and variance v_i the parameter values are

$$a_i = \frac{m_i^2}{v_i}, \quad b_i = \frac{m_i}{v_i}.$$

Thus, for the specifications above,

$$a_i = 30, \quad b_i = 1.$$

The prior density takes the form of Equation 2.8. A contour plot of this density is given in the left hand side of Figure 2.4.

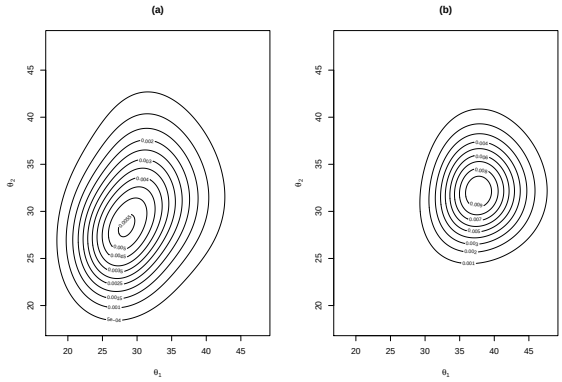

Figure 2.4: Prior and posterior contour plots using the FGM copula for the Poisson example

We observe $x_1 = 46$ failures in group 1 and $x_2 = 33$ failures in group 2. The posterior density takes the form of Equation 2.7 with $A_1 = 76, B_1 = 2, A_2 = 63$ and $B_2 = 2$. A contour plot of the posterior density is given in the right hand side of Figure 2.4.

There has been a clear reduction in uncertainty from prior to posterior. The location of the density has also changed, with the mode of the joint density higher in both θ_1 and θ_2 in the posterior.

The posterior expectations and variances of θ_1 and θ_2 are given in Table 2.4.

θ_1	$\mathrm{E}_1(\theta_i)$	$\mathrm{Var}_1(\theta_i)$
1	37.50	19.83
2	30.89	15.38

Table 2.4: Posterior expectations and variances using the FGM copula for piston ring failures

Clearly the domain of the prior Spearman's Rho being $[-1/3, 1/3]$ significantly restricts the FGM copula's usefulness in modelling to situations of only weak dependence. Thus extensions to the copula have been proposed with the aim of increasing the range of correlation which can be represented. We shall now consider two such extensions.

2.7 Extensions to the FGM copula

Johnson & Kotz (1977) proposed an extension which replaces the standard FGM copula with a finite sum. This was built upon in Huang & Kotz (1984). They call the result an "iterated" copula. Taking $m-1$ iterations gives a copula of the form

$$C_{J,m}(u,v) = uv + \sum_{i=1}^{m} \lambda_i (uv)^{[\frac{1}{2}i]+1} \{(1-u)(1-v)\}^{[\frac{1}{2}i+\frac{1}{2}]},$$

where $[x]$ means take the largest integer less than or equal to x. Taking zero iterations (setting $m=1$) gives

$$C_{J,1}(u,v) = uv + \lambda_1 (uv)(1-u)(1-v),$$

the usual FGM copula. Taking a first iteration we see that

$$C_{J,2}(u,v) = uv + \lambda_1 (uv)(1-u)(1-v) + \lambda_2 u^2 v^2 (1-u)(1-v),$$

for $-1 \leq \lambda_1 \leq 1$ and $-\lambda_1 - 1 \leq \lambda_2 \leq \frac{1}{2}\{3 - \lambda_1 + \sqrt{9 - 6\lambda_1 - 3\lambda_1^2}\}$. Differentiating leads to a density of $c(u,v) = 1 + \lambda_1(1-2u)(1-2v) + \lambda_2(2u-3u^2)(2v-3v^2)$ and Spearman's Rho can be found via

$$\int_0^1 \int_0^1 C_{J,2}(u,v) du dv = \frac{1}{4} + \frac{1}{36}\lambda_1 + \frac{1}{144}\lambda_2.$$

Thus $\rho_{\theta_1,\theta_2} = \frac{1}{3}\lambda_1 + \frac{1}{12}\lambda_2$. This increases the potential positive correlation which can be specified using the copula to 0.434 (Huang & Kotz, 1999).

Lin (1987) considered the Johnson and Kotz extension and then proposed a similar iterated FGM copula built upon this. If we reverse the powers in the summation in the Johnson and Kotz extension we arrive at the Lin extension:

$$C_{L,m}(u,v) = uv + \sum_{i=1}^{m} \gamma_i (uv)^{[\frac{1}{2}i+\frac{1}{2}]} \{(1-u)(1-v)\}^{[\frac{1}{2}i]+1}.$$

Setting $m=2$ as before gives

$$C_{l,2}(u,v) = uv + \gamma_1 uv(1-u)(1-v) + \gamma_2 uv(1-u)^2(1-v)^2,$$

where γ_1 and γ_2 have the same restrictions as λ_1 and λ_2. Differentiating gives the copula density as

$$c_{L,2}(u,v) = 1 + \gamma_1(1-2u)(1-2v) + \gamma_2(1-4u+3u^2)(1-4v+3v^2)$$

In terms of the correlation given by Spearman's Rho

$$\rho_{\theta_1,\theta_2} = 12\int_0^1\int_0^1 C_{L,2}(u,v)dudv = \frac{1}{3}\gamma_1 + \frac{1}{12}\gamma_2,$$

as in the Johnson and Kotz extension. The admissable correlations for the Lin extension are generally the same as in the Johnson and Kotz extension (Lin, 1987).

2.7.1 Heart attack example

We can view the analysis carried out using the FGM copula for the heart attack example as a special case of the Johnson and Kotz extension with $\lambda_1 = 1$ and $\lambda_2 = 0$. Clearly when we take the extension there is no longer a unique choice of λ_1 and λ_2 so that our prior specifications are those given in Section 2.6.1. We can investigate the effect of choosing different values of λ_1 and λ_2. Prior and posterior contour plots for some alternative values of λ_1 and λ_2 which maintain our prior specifications are given in Figure 2.5.

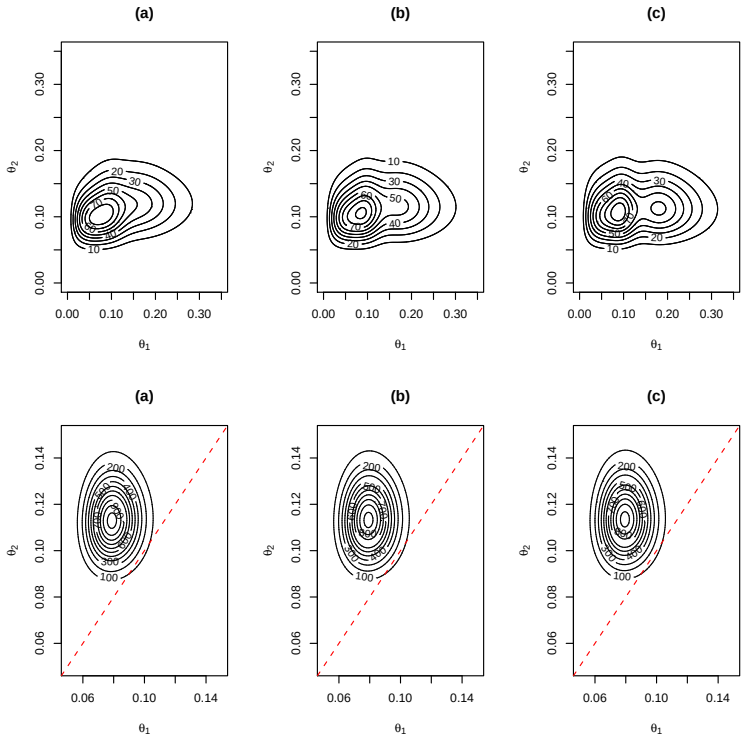

Figure 2.5: Prior (top panel) and posterior (bottom panel) contour plots for the Johnson and Kotz extension for different values of λ_1 and λ_2, heart attack example

In (a) $\lambda_1 = \lambda_2 = 0.8$. With λ_1 being fairly close to its previous value of 1 the prior

contour plot is similar to that of the standard FGM copula in Figure 2.1. For (b) values of $\lambda_1 = 0.6$ and $\lambda_2 = 1.6$ were used and in (c) $\lambda_1 = 0.4$ and $\lambda_2 = 2.4$. Clearly as we decrease λ_1 and increase λ_2 the joint prior density is becoming more spread and in (c) appears to be bimodal. All of the posterior densities are very similar and adhere extremely closely to that from the FGM copula given in Figure 2.2.

If we now consider the Lin extension we again do not have unique parameter values γ_1 and γ_2 to give a prior Spearman's Rho of $\frac{1}{3}$. Setting γ_1 and γ_2 to the values used previously for λ_1 and λ_2 gives prior and posterior contour plots as in Figure 2.6. We

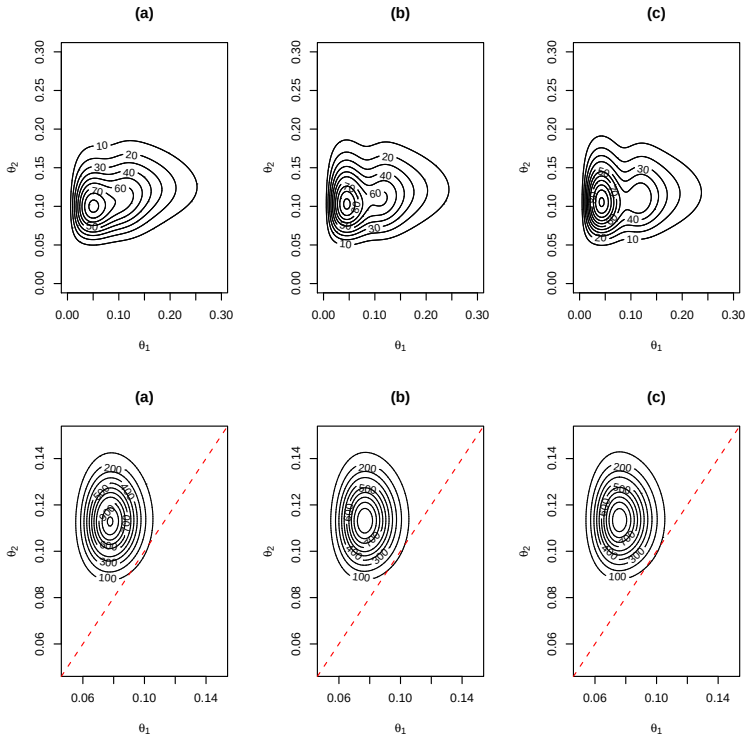

Figure 2.6: Prior (top panel) and posterior (bottom panel) contour plots for Lin extension for different values of γ_1 and γ_2

see a similar pattern as with the Johnson and Kotz extension. The prior joint density tends towards bimodality as γ_1 decreases and γ_2 increases. There appears to be more density concentrated close to the zero boundaries with the Lin extension however. Once again all of the posterior densities look very similar as a consequence of the volume of observations.

2.8 Discussion

The FGM copula has produced a solution to the problem of two correlated binomial or Poisson parameters in which such quantites as posterior moments can be found analytically. However, the approach is not straightforward as posterior quantities are fairly complicated finite sums and, in the example considered, were calculated in Maple (TM).

In practice this method is only of very limited use due to the serious restrictions on the prior specification of the correlation between the two parameters, i.e. between $-1/3$ and $1/3$. In order to overcome this problem two extensions to the FGM copula have been examined, the Johnson and Kotz extension and the Lin extension. Both extend the range of possible positive correlation when a single iteration is taken to an absolute maximum of 0.434.

Clearly this is an improvement, although not a large one. If a greater prior correlation is required then further iterations could be taken as these could extend the range of possible correlation further still. This would, however, make the calculation of posterior quantities more complicated.

An alternative to extending the FGM copula would be to use a different family of copula functions. Unfortunately other commonly used families of copulas do not have the simple polynomial form of the FGM copula, although they do allow for stronger correlations to be specified. Thus it would require numerical integration to find quantities such as posterior moments.

2.9 Mixtures

Consider the beta-binomial and gamma-Poisson setups explored in Section 2.2. In both situations the joint prior density between θ_1 and θ_2 is

$$f_0(\theta_1, \theta_2) \propto f_{01}(\theta_1) f_{02}(\theta_2) g(\theta_1, \theta_2).$$

One way of specifying $g(\theta_1, \theta_2)$ would be such that the resulting density is of the form

$$f_0(\theta_1, \theta_2) = \sum_{j=1}^{m} p_{0j} f_{0j1}(\theta_1) f_{0j2}(\theta_2),$$

where p_{0j} are weights for which $\sum_{j=1}^{m} p_{0j} = 1$. That is, the resulting joint density of θ_1 and θ_2 is a mixture of products of beta or gamma densities of the form

$$f_{0ji}(\theta_i) = \frac{\Gamma(a_{ji} + b_{ji})}{\Gamma(a_{ji})\Gamma(b_{ji})}\theta_i^{a_{ji}-1}(1-\theta_i)^{b_{ji}-1},$$

or

$$f_{0ji}(\theta_i) = \frac{b_{ji}^{a_{ji}}\theta_i^{a_{ji}-1}e^{b_{ji}\theta_i}}{\Gamma(a_{ji})},$$

respectively for some parameter values a_{ji} and b_{ji}.

For the use of finite mixtures in Bayesian statistics see McLachlan & Peel (2000); Lavine & West (1992). First Bayes is a teaching package which allows, amongst other things, simple statistical analyses of univariate mixtures to be performed. It is free to download (http://tonyohagan.co.uk/1b/).

The number of components in such a mixture, m, could be finite or infinite. The marginal distributions of θ_1 and θ_2 are no longer simply

$$\theta_i \sim \text{beta}(a_i, b_i) \quad \text{and} \quad \theta_i \sim \text{gamma}(a_i, b_i),$$

as in the case of copulas. Observation of x_1 successes in group 1 and x_2 successes in group 2 gives a posterior joint density of

$$f_1(\theta_1, \theta_2) = \sum_{j=1}^{m} p_{1j} f_{1j1}(\theta_1) f_{1j2}(\theta_2).$$

The marginal density terms in the above summation are of the form of their prior counterparts but with each of the sets of parameter values updated to

$$A_{ji} = a_{ji} + x_i, \quad B_{ji} = b_{ji} + n_i - x_i,$$

in the beta-binomial case and

$$A_{ji} = a_{ji} + x_i, \quad B_{ji} = b_{ji} + 1,$$

for the gamma-Poisson setup. The mixture weights also change.

We now investigate specification of $g(\theta_1, \theta_2)$ to produce such mixtures for both binomial and Poission distributions. We restrict ourselves to finite mixtures.

2.10 A two component beta mixture

Clearly there are many possibilities for the density multiplier $g(\theta_1, \theta_2)$ which lead to mixtures. In this section one shall be introduced which was first discussed by Valks (2005) for the beta-binomial case. The function is given by

$$g(\theta_1, \theta_2) = 1 + k\theta_1^{m_1}\theta_2^{m_2} \tag{2.13}$$

and satisfies the relationships explored in Section 2.9. There are three additional parameters, k, m_1 and m_2, as well as the four associated with the beta prior distributions for θ_1 and θ_2. The values of these parameters are found in order to satisfy our prior beliefs. The joint prior density is

$$\begin{aligned} f_0(\theta_1, \theta_2) &= \frac{1}{C_0} f_{01}(\theta_1) f_{02}(\theta_2)[1 + k\theta_1^{m_1}\theta_2^{m_2}] \\ &= p_{01} f_{011}(\theta_1) f_{012}(\theta_2) + p_{02} f_{021}(\theta_1) f_{022}(\theta_2), \end{aligned}$$

where $p_{02} = 1 - p_{01}$ and, for $i = 1, 2$, the relevent beta densities are

$$f_{01i}(\theta_i) = \frac{\Gamma(a_{i1} + b_{i1})}{\Gamma(a_{i1})\Gamma(b_{i1})} \theta_i^{a_{i1}-1}(1 - \theta_i)^{b_{i1}-1},$$

and

$$f_{02i}(\theta_i) = \frac{\Gamma(a_{i2} + b_{i2})}{\Gamma(a_{i2})\Gamma(b_{i2})} \theta_i^{a_{i2}-1}(1 - \theta_i)^{b_{i2}-1},$$

for $a_{i1} = a_i, b_{i1} = b_i, a_{i2} = a_i + m_i$ and $b_{i2} = b_i$. We see that this is a two component mixture distribution. Each component is a product of two beta densities. The weight term for the first component, p_{01}, is given by

$$\begin{aligned} p_{01}^{-1} &= \int_0^1 \int_0^1 f_{01}(\theta_1) f_{02}(\theta_2)[1 + k\theta_1^{m_1}\theta_2^{m_2}] d\theta_1 d\theta_2 \\ &= 1 + k \frac{\Gamma(a_{11} + b_{11})}{\Gamma(a_{12} + b_{12})} \frac{\Gamma(a_{12})}{\Gamma(a_{11})} \frac{\Gamma(a_{21} + b_{21})}{\Gamma(a_{22} + b_{22})} \frac{\Gamma(a_{22})}{\Gamma(a_{21})}. \end{aligned}$$

Prior expectations and variances for θ_1 and θ_2, which are necessary in the process of prior specification, can be found. The prior expectations of θ_1, θ_2 are

$$\begin{aligned} \mathrm{E}_0(\theta_i) &= \int_0^1 \int_0^1 \theta_i f_0(\theta_1, \theta_2) d\theta_1 d\theta_2 \\ &= p_{01} \left(\frac{a_{i1}}{a_{i1} + b_{i1}} \right) + p_{02} \left(\frac{a_{i2}}{a_{i2} + b_{i2}} \right) \\ &= p_{01}\mathrm{E}_{01}(\theta_i) + p_{02}\mathrm{E}_{02}(\theta_i). \end{aligned}$$

Thus the expectations are weighted sums of the expectations of the components of the mixture. In order to find prior variances the second moments of each of the parameters are required. These are

$$\begin{aligned}
\mathrm{E}_0(\theta_i^2) &= \int_0^1 \int_0^1 \theta_i^2 f_0(\theta_1, \theta_2) d\theta_1 d\theta_2 \\
&= p_{01} \frac{a_{i1}(a_{i1}+1)}{(a_{i1}+b_{i1})(a_{i1}+b_{i1}+1)} + p_{02} \frac{(a_{i2})(a_{i2}+1)}{(a_{i2}+b_{i2})(a_{i2}+b_{i2}+1)} \\
&= p_{01}\mathrm{E}_{01}(\theta_i^2) + p_{02}\mathrm{E}_{02}(\theta_i^2).
\end{aligned}$$

Finally, the prior covariance between θ_1 and θ_2 is given by $\mathrm{Cov}_0(\theta_1, \theta_2) = \mathrm{E}_0(\theta_1\theta_2) - \mathrm{E}_0(\theta_1)\mathrm{E}_0(\theta_2)$, where

$$\begin{aligned}
\mathrm{E}_0(\theta_1\theta_2) &= p_{01} \left(\frac{a_{11}}{a_{11}+b_{11}}\right)\left(\frac{a_{21}}{a_{21}+b_{21}}\right) + p_{02} \left(\frac{a_{12}}{a_{12}+b_{12}}\right)\left(\frac{a_{22}}{a_{22}+b_{22}}\right) \\
&= p_{01}\mathrm{E}_{01}(\theta_1)\mathrm{E}_{01}(\theta_2) + p_{02}\mathrm{E}_{02}(\theta_1)\mathrm{E}_{02}(\theta_2).
\end{aligned}$$

From all of these quantities prior specifications can be made.

Suppose that we update our beliefs by observing x_1 successes out of n_1 trials in group 1 and x_2 successes out of n_2 trials in group 2. In this case the posterior joint density shall be

$$f_1(\theta_1, \theta_2) = p_{11} f_{111}(\theta_1) f_{112}(\theta_2) + p_{12} f_{121}(\theta_1) f_{122}(\theta_2). \tag{2.14}$$

The posterior mixing probabilities are then

$$\begin{aligned}
p_{1j} \propto p_{0j} \frac{\Gamma(a_{1j}+b_{1j})}{\Gamma(a_{1j})\Gamma(b_{1j})} \frac{\Gamma(a_{1j}+x_1)\Gamma(b_{1j}+n_1-x_1)}{\Gamma(a_{1j}+b_{1j}+n_1)} \\
\times \frac{\Gamma(a_{2j}+b_{2j})}{\Gamma(a_{2j})\Gamma(b_{2j})} \frac{\Gamma(a_{2j}+x_2)\Gamma(b_{2j}+n_2-x_2)}{\Gamma(a_{2j}+b_{2j}+n_2)},
\end{aligned}$$

for $j = 1, 2$. They are then normalised by dividing by the total.

The posterior marginal densities $f_{11i}(\theta_i)$ and $f_{12i}(\theta_i)$ and moments $\mathrm{E}_1(\theta_1)$, $\mathrm{E}_1(\theta_2)$, $\mathrm{E}_1(\theta_1^2)$, $\mathrm{E}_1(\theta_2^2)$ and $\mathrm{E}_1(\theta_1\theta_2)$ are the same as their prior counterparts but using the posterior mixture probabilities and with a_1, b_1, a_2 and b_2 replaced by

$$\begin{aligned}
A_1 = a_1 + x_1, \quad B_1 = b_1 + n_1 - x_1, \\
A_2 = a_2 + x_2, \quad B_2 = b_2 + n_2 - x_2.
\end{aligned}$$

2.10.1 Heart attack example

The model now has 7 hyperparameters; $a_1, b_1, a_2, b_2, k, m_1$ and m_2. We shall reduce this to 6 by setting $m_1 = m_2$ as we wish to make 5 prior specifications. It is assumed that the same prior specification is to be made as in the copulas case, that is

$$\mathrm{E}_0(\theta_1) = \mathrm{E}_0(\theta_2) = 0.117, \tag{2.15}$$

$$\mathrm{Var}_0(\theta_1) = 0.0045, \quad \mathrm{Var}_0(\theta_2) = 0.0011. \tag{2.16}$$

A prior correlation of approximately 0.3 will also be specified.

Iteration methods are used to find suitable values for the parameters to represent our prior beliefs. First approximations for the beta parameters are taken as the specifications from the copulas model. That is, $a_1 = 2.59$, $b_1 = 19.55$, $a_2 = 10.72$, and $b_2 = 80.84$. In her thesis Valks (2005) noted that for each $m = m_1 = m_2$ increasing k increases the correlation until some maximum value after which the correlation decreases with increasing k. The correlation also increases as m gets larger.

She then set out some general advice for the sizes of k and m and following this advice values of $k = 20,000,000$ and $m = 4$ are chosen. It is necessary to readjust the values for the beta parameters and these are found to be $a_1 = 3.01$, $b_1 = 33.8$, $a_2 = 11.6$ and $b_2 = 98.6$ and the prior weighting parameter for the components of the mixture is then $p_{01} = 0.613$. This leads to the values of the means and variances of θ_1 and θ_2 of

$$\mathrm{E}_0(\theta_1) = \mathrm{E}_0(\theta_2) = 0.117,$$

and

$$\mathrm{Var}_0(\theta_1) = 0.0045, \quad \mathrm{Var}_0(\theta_2) = 0.0011.$$

The prior-moment correlation is 0.30 which is almost identical to that specified in the copula model.

Figure 2.7 illustrates that the prior density appears to be fairly similar to that using the FGM copula. The correlation between θ_1 and θ_2 can be seen clearly in the joint density.

The posterior density is of the form given in Equation 2.14. The only change from the prior specifications occurs in the values of the parameters a_1, b_1, a_2 and b_2 and the wieghting parameter p_{01} which are updated via Bayes Theorem to

$$A_1 = 47.01, \ A_2 = 73.6, \ B_1 = 549.8, \ B_2 = 576.6, \ p_{11} = 0.838.$$

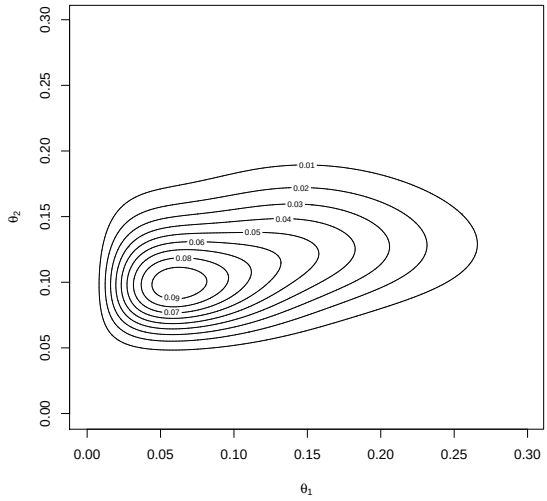

Figure 2.7: A prior contour plot for $g(\theta_1, \theta_2) = 1 + k\theta_1^n\theta_2^n$

Figure 2.8 gives a contour plot of the joint posterior density. The dashed line is at $\theta_1 = \theta_2$ and the fact that almost the entire density is above this line indicates that there is fairly strong evidence that the new treatment is affecting the survival rate of patients.

The posterior density indicates that the correlation between the two parameters has reduced. We can also find the posterior moments of θ_1 and θ_2. These are given in Table 2.5.

Group	$\mathrm{E}_1(\theta_j)$	$\mathrm{Var}_1(\theta_j)$
1	0.080	1.28×10^{-4}
2	0.114	1.59×10^{-4}

Table 2.5: Posterior moments for the 2 component mixture

Table 2.5 agrees with the posterior contour plot. The posterior expectation for the probability of death in the group of patients taking sulfinpyrazone is much lower than for the placebo group. Posterior variances are far lower than prior variances indicating that most of the uncertainty has been explained upon observation of the data. The posterior correlation is 0.032, far lower than the prior correlation.

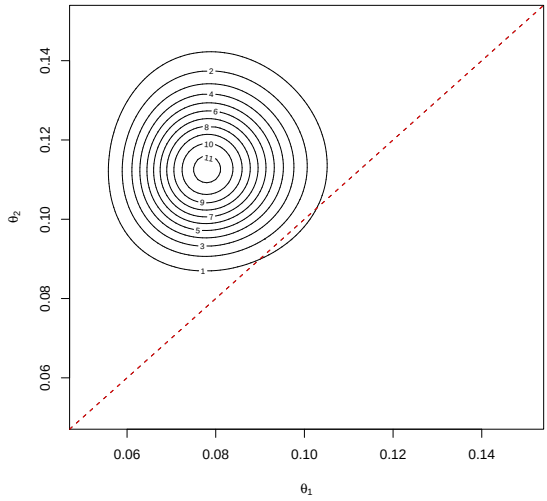

Figure 2.8: A posterior contour plot for $g(\theta_1, \theta_2) = 1 + k\theta_1^n\theta_2^n$

2.11 A two component gamma mixture

We can use the density multiplier in Equation 2.13 to define a two component mixture in the gamma-Poisson model. Doing so will lead to a joint density of the form

$$f_0(\theta_1, \theta_2) = p_{01} f_{011}(\theta_1) f_{012}(\theta_2) + p_{02} f_{021}(\theta_1) f_{022}(\theta_2),$$

where $p_{02} = 1 - p_{01}$, as in the beta-binomial model. The prior densities f_{01i} and f_{02i} for $i = 1, 2$ are

$$f_{011}(\theta_i) = \frac{b_{i1}^{a_{i1}} \theta_i^{a_{i1}-1} e^{-b_{i1}\theta_i}}{\Gamma(a_{i1})},$$

and

$$f_{02i}(\theta_i) = \frac{b_{i2}^{a_{i2}} \theta_i^{a_{i2}-1} e^{b_{i2}\theta_i}}{\Gamma(a_{i2})},$$

for $a_{i1} = a_i, b_{i1} = b_i, a_{i2} = a_i + m_i$ and $b_{i2} = b_i$. The weight component in this case is

$$\begin{aligned} p_{01}^{-1} &= \int_0^\infty \int_0^\infty f_{01}(\theta_1) f_{02}(\theta_2)[1 + k\theta_1^{m_1}\theta_2^{m_2}] d\theta_1 \theta_2 \\ &= 1 + k b_1^{-m_1} b_2^{-m_2} \frac{\Gamma(a_{12})}{\Gamma(a_{11})} \frac{\Gamma(a_{22})}{\Gamma(a_{21})}. \end{aligned}$$

From this we can find expressions for the prior moments of θ_1 and θ_2 in the same way as the beta-binomial case. The prior expectations, for $i = 1, 2$, are

$$\begin{aligned} \mathrm{E}_0(\theta_i) &= p_{01}\mathrm{E}_{01}(\theta_i) + p_{02}\mathrm{E}_{02}(\theta_i) \\ &= p_{01}\left(\frac{a_{i1}}{b_{i1}}\right) + p_{02}\left(\frac{a_{i2}}{b_{i2}}\right). \end{aligned}$$

In order to find the variance we calculate the second moment of each θ_i.

$$\begin{aligned} \mathrm{E}_0(\theta_i^2) &= p_{01}\mathrm{E}_{01}(\theta_i^2) + p_{02}\mathrm{E}_{02}(\theta_i^2) \\ &= p_{01}\left(\frac{a_{i1}(a_{i1}+1)}{b_{i1}^2}\right) + p_{02}\left(\frac{a_{i2}(a_{i2}+1)}{b_{i2}^2}\right). \end{aligned}$$

In order to find the prior covariance we need the first mixed moment between θ_1 and θ_2. This can be found in the same way so that

$$\begin{aligned} \mathrm{E}_0(\theta_1\theta_2) &= p_{01}\mathrm{E}_{01}(\theta_1)\mathrm{E}_{01}(\theta_2) + p_{02}\mathrm{E}_{02}(\theta_1)\mathrm{E}_{02}(\theta_2) \\ &= p_{01}\left(\frac{a_{11}a_{21}}{b_{11}b_{21}}\right) + p_{02}\left(\frac{a_{12}a_{22}}{b_{12}b_{22}}\right). \end{aligned}$$

From these 5 equations the prior expectations, variances and covariances can be specified using the parameters in the model.

If we observe x_1 successes in group 1 and x_2 successes in group 2, over a single time period in each, then we obtain the joint posterior density by application of Bayes Theorem. It is

$$f_1(\theta_1, \theta_2) = p_{11}f_{111}(\theta_1)f_{112}(\theta_2) + p_{12}f_{121}(\theta_1)f_{122}(\theta_2), \tag{2.17}$$

where the weight components are

$$p_{1j} \propto p_{0j}\frac{b_{1j}^{a_{1j}}}{(b_{1j}+1)^{a_{1j}+x_1}}\frac{b_{2j}^{a_{2j}}}{(b_{2j}+1)^{a_{2j}+x_2}}\frac{\Gamma(a_{1j}+x_1)}{\Gamma(a_{1j})}\frac{\Gamma(a_{2j}+x_2)}{\Gamma(a_{2j})}.$$

The posterior densities $f_{1ji}(\theta_i)$ are gamma as with their prior counterparts. The parameter values have changed, however, and are $A_i = a_i + x_i$ and $B_i = b_i + 1$.

2.11.1 Example: piston ring failures

We wish to make the same prior specifications as in Section 2.6.4. This is done iteratively using the above equations for moments. The resulting gamma parameter values are $a_1 = a_2 = 37.75$, $b_1 = b_2 = 1.34$, $m_1 = m_2 = 8$ and $p_{01} = 0.692$. We choose to specify p_{01} rather than k as setting one fixes the other. This leads to prior expectations

and variances for θ_1 and θ_2 of

$$\mathrm{E}_0(\theta_i) = 30.0 \quad \mathrm{Var}_0(\theta_i) = 30.0,$$

and a prior correlation between θ_1 and θ_2 of 0.25. The joint density for θ_1 and θ_2 is given in the left hand side of Figure 2.9.

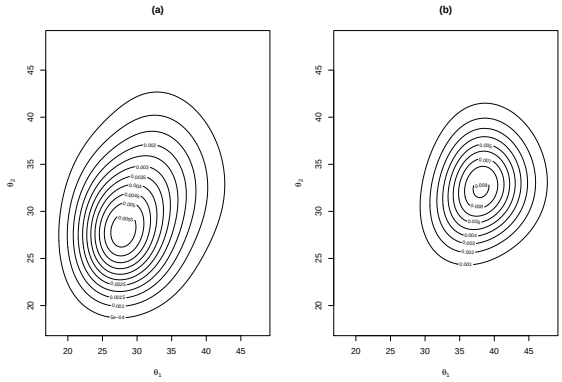

Figure 2.9: Prior and posterior contour plots for the 2 component gamma mixture

We observe 46 piston-ring failures in compressor 1 and 33 piston-ring failures in compressor 2. This leads to a posterior joint density of the form given in Equation 2.17 with $p_{11} = 0.272$, $A_1 = 83.75, B_1 = 2.34, A_2 = 70.75$ and $B_2 = 2.34$. We can plot the posterior density as in the right hand side of Figure 2.9.

The densities are similar to those found for the copula model. There has been a reduction in the spread of the density from prior to posterior indicating a reduction in uncertainty on observing the data. The mode of the joint density has also increased in both θ_1 and θ_2, particularly θ_1 in which compressor more failures were observed.

The posterior means and variances of θ_1 and θ_2 are given in Table 2.6. The posterior

Compressor	$\mathrm{E}_1(\theta_i)$	$\mathrm{Var}_1(\theta_i)$
1	38.28	18.67
2	32.72	16.30

Table 2.6: Posterior moments for the 2 component mixture in the piston-ring failures example

correlation between θ_1 and θ_2 is 0.133. We see that both expectations have increased on observing higher than expected numbers of failures in the two groups. Both posterior variances are lower than both prior variances indicating the reduction in uncertainty we observed in the contour plots. This is also seen in the reduction in the correlation.

2.12 A finite component beta mixture

One way to set up a mixture distribution would be to think of beta priors as though they represent 'prior observations'. So using a beta(a, b) prior for a single θ might be thought of as representing a prior successes and b prior failures.

Thus we can think of a beta mixture prior as representing a distribution of possible 'prior observations'. For example if $f_0(\theta; a, b)$ represents the beta(a, b) density for θ then the mixture density,

$$f_0(\theta) = p_0 f_0(\theta; a_{(1)}, b_{(1)}) + (1 - p_0) f_0(\theta; a_{(2)}, b_{(2)})$$

represents probability p_0 of having $a_{(1)}$ prior successes and $b_{(1)}$ prior failures and probability $1 - p_0$ of having $a_{(2)}$ prior successes and $b_{(2)}$ prior failures.

Now suppose we extend this to the to the bivariate θ_1, θ_2 case. If we wish to make θ_1 and θ_2 positively correlated then we have a distribution for the numbers of 'prior successes and failures' we have had, a_1, b_1, a_2, b_2, with a_1 and a_2 positively associated and b_1 and b_2 positively associated.

We could then fix

$$a_1 + b_1 = N_1, a_2 + b_2 = N_2.$$

so that a_1, a_2 are negatively associated with b_1, b_2. Having done this our mixture could take the form

$$f_0(\theta_1, \theta_2) = \sum_{j=0}^{m} p_{0j} f_{01}(\theta_1; a_1 + c_j, b_1 + N_1 - c_j) f_{02}(\theta_2; a_2 + c_j, b_2 + N_2 - c_j),$$

where $\sum_{j=0}^{m} p_{0j} = 1$. One way to define p_{0j} and c_j would be to give c_j a binomial distribution so that

$$c_j \sim \text{bin}(m, p^*),$$

with p_{0j} calculated from this as $p_{0j} = \Pr(c_j = j)$. Similarly, for negative correlation the mixture could take the form

$$f_0(\theta_1, \theta_2) = \sum_{j=0}^{m} p_{0j} f_{01}(\theta_1; a_1 + c_j, b_1 + N_1 - c_j) f_{02}(\theta_2; a_2 + N_2 - c_j, b_2 + c_j).$$

We can calculate prior moments for this mixture. The prior expectations of θ_1 and θ_2

are

$$\begin{aligned} \mathrm{E}_0(\theta_i) &= \sum_{j=0}^{m} p_{0j} \int_0^1 \theta_i f_{0i}(\theta_i; a_i + c_j, b_i + N_i - c_j) d\theta_i \\ &= \sum_{j=0}^{m} p_{0j} \frac{a_i + c_j}{a_i + b_i + N_i}, \end{aligned}$$

for $i = 1, 2$. The prior variances are given by $\mathrm{Var}_0(\theta_i) = \mathrm{E}_0(\theta_i^2) - \mathrm{E}_0(\theta_i)^2$, where the second moment of θ_i is

$$\begin{aligned} \mathrm{E}_0(\theta_i^2) &= \sum_{j=0}^{m} p_{0j} \int_0^1 \theta_i^2 f_{0i}(\theta_i; a_i + c_j, b_i + N_i - c_j) d\theta_i \\ &= \sum_{j=0}^{m} p_{0j} \frac{(a_i + c_j)(a_i + c_j + 1)}{(a_i + b_i + N_i)(a_i + b_i + N_i + 1)}. \end{aligned}$$

Finally, in order to find the prior covariance between θ_1 and θ_2, we need $\mathrm{E}_0(\theta_1\theta_2)$. This is

$$\begin{aligned} \mathrm{E}_0(\theta_1\theta_2) &= \sum_{j=0}^{m} p_{0j} \int_0^1 \theta_1 f_{01}(a_1 + c_j, b_1 + N_1 - c_j) d\theta_1 \int_0^1 \theta_2 f_{02}(a_2 + c_j, b_2 + N_2 - c_j) d\theta_2 \\ &= \sum_{j=0}^{m} p_{0j} \frac{(a_1 + c_j)}{(a_1 + b_1 + N_1)} \frac{(a_2 + c_j)}{(a_2 + b_2 + N_2)}. \end{aligned}$$

The prior covariance is then $\mathrm{Cov}_0(\theta_1, \theta_2) = \mathrm{E}_0(\theta_1\theta_2) - \mathrm{E}_0(\theta_1)\mathrm{E}_0(\theta_2)$.

Having observed x_1 successes in group 1 and x_2 successes in group 2 we update via Bayes theorem using the likelihood in Equation 2.2 and obtain a joint posterior distribution for θ_1 and θ_2, in the case of positive correlation, of the form

$$f_1(\theta_1, \theta_2) = \sum_{j=0}^{m} p_{1j} f_{11}(\theta_1; A_1 + c_j, B_1 + N_1 - c_j) f_{12}(\theta_2; A_2 + c_j, B_2 + N_2 - c_j),$$

with posterior weights proportional to

$$\begin{aligned} \hat{p}_{1k} = p_{0k} &\frac{\Gamma(a_1 + b_1 + N_1)}{\Gamma(a_1 + c_k)\Gamma(b_1 + N_1 - c_k)} \frac{\Gamma(a_1 + c_k + x_1)\Gamma(b_1 + N_1 - c_k + n_1 - x_1)}{\Gamma(a_1 + b_1 + N_1 + n_1)} \\ &\times \frac{\Gamma(a_2 + b_2 + N_2)}{\Gamma(a_2 + c_k)\Gamma(b_2 + N_2 - c_k)} \frac{\Gamma(a_2 + c_k + x_2)\Gamma(b_2 + N_2 - c_k + n_2 - x_2)}{\Gamma(a_2 + b_2 + N_2 + n_2)}. \end{aligned}$$

The weights are then $p_{1k} = \dfrac{\hat{p}_{1k}}{\sum_{j=0}^{m} p_{1j}}$.

This is the same form as the prior density but with the parameters updated to $A_1 =$

$a_1 + x_1$, $B_1 = b_1 + n_1 - x_1$, $A_2 = a_2 + x_2$ and $B_2 = b_2 + n_2 - x_2$. Similarly, the posterior moments of θ_1 and θ_2 are of the same form as the prior but using A_1, B_1, A_2 and B_2.

2.12.1 Example: Heart attack data

Let us now apply this model to the heart attack data. We have 6 prior parameters with which to make the 5 prior specifications given in Equations 2.15 and 2.16. We settle on values of

$$a_1 = 0.88, b_1 = 18.7, a_2 = 9.8, b_2 = 46.5, m = 32, p^* = 0.12,$$

so that we have a 33 component beta mixture. This gives prior moments for θ_1 and θ_2 of

$$\begin{aligned} \mathrm{E}_0(\theta_1) &= 0.117, \quad & \mathrm{E}_0(\theta_2) &= 0.117 \\ \mathrm{Var}_0(\theta_1) &= 0.0045, \quad & \mathrm{Var}_0(\theta_2) &= 0.0011, \end{aligned}$$

and $\mathrm{Corr}_0(\theta_1, \theta_2) = 0.31$. We can also produce a prior contour plot for this mixture and this is given in Figure 2.10. Having updated with the observed numbers of deaths

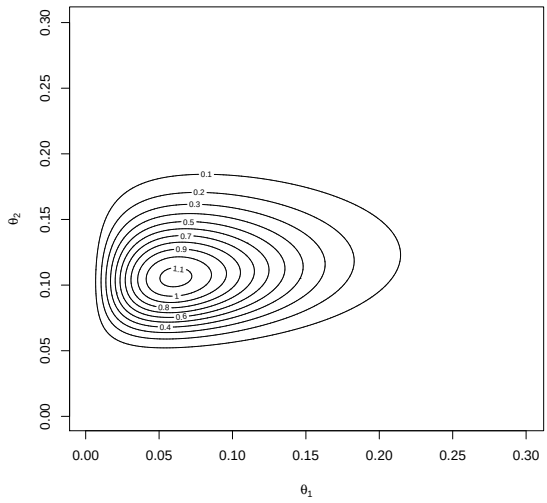

Figure 2.10: A prior contour plot for the finite mixture model, heart attacks example

in the two groups, $x_1 = 44$ and $x_2 = 62$, we obtain the joint posterior distribution with

$$A_1 = 44.88, B_1 = 534.7, A_2 = 71.8, B_2 = 524.5,$$

and $m = 32$, $p^* = 0.12$ as before. A posterior contour plot is given in Figure 2.11. We

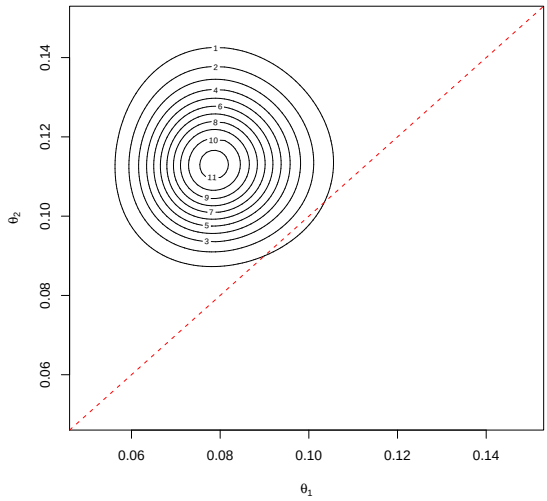

Figure 2.11: A posterior contour plot for the finite mixture model, heart attacks example

can also calculate posterior moments for θ_1 and θ_2. These are

$$\begin{aligned} \mathrm{E}_1(\theta_1) &= 0.0803, \quad \mathrm{Var}_1(\theta_1) = 1.27 \times 10^{-4}, \\ \mathrm{E}_1(\theta_2) &= 0.114, \quad \mathrm{Var}_1(\theta_2) = 1.58 \times 10^{-4}. \end{aligned}$$

These results are very similar to those achieved in earlier models.

2.13 A finite component gamma mixture

Now suppose we wish to create a finite component gamma mixture prior for Poisson models. We can do so in a similar way to the beta-binomial setup in the previous section. If we have a single θ then we can think of observing a events in an interval of length b. This gives a Poisson likelihood proportional to $e^{-b\theta}\theta^a$.

Then a mixture of the form

$$f_0(\theta) = p_0 f_0(\theta; a_{(1)}, b_{(1)}) + (1 - p_0) f_0(\theta; a_{(2)}, b_{(2)}),$$

represents probability p_0 of observing $a_{(1)}$ prior events in an interval of length $b_{(1)}$ and probability $1 - p_0$ of observing $a_{(2)}$ prior events in an interval of length $b_{(2)}$. Here $f_0(\theta; a, b)$ is a prior gamma density for θ with parameters a and b.

Considering the two parameter problem, if we wish to specify positive correlation be-

tween θ_1, θ_2, then a finite mixture prior could be

$$f_0(\theta_1, \theta_2) = \sum_{j=0}^{m} p_{0j} f_{01}(\theta_1; a_1 + c_j, b_1) f_{02}(\theta_2; a_2 + c_j, b_2),$$

for $\sum_{j=0}^{m} p_{0j} = 1$. We can give each c_j a binomial distribution as previously so that $c_j \sim \text{bin}(m, p^*)$, for some p^*, with the weights being $p_{0j} = \Pr(c_j = j)$. If, alternatively, we gave each c_j a Poisson distribution this would lead to an infinite mixture.

If we wished to specify a negative prior correlation then the mixture could take the form

$$f_0(\theta_1, \theta_2) = \sum_{j=0}^{m} p_{0j} f_{01}(\theta_1; a_1 + c_j, b_1) f_{02}(\theta_2; a_2 + m - c_j, b_2),$$

where once again $p_{0j} = \Pr(c_j = j)$ for $\text{bin}(m, p^*)$.

From this specification we can find the prior moments. The prior expectations of θ_1 and θ_2 are

$$\begin{aligned} \mathrm{E}_0(\theta_i) &= \sum_{j=0}^{m} p_{0j} \int_0^\infty \theta_i f_{0i}(\theta_i; a_i + c_j, b_i) d\theta_i \\ &= \sum_{j=0}^{m} p_{0j} \frac{a_i + c_j}{b_i}, \end{aligned}$$

for $i = 1, 2$. To calculate the variances first we find the second moments.

$$\begin{aligned} \mathrm{E}_0(\theta_i^2) &= \sum_{j=0}^{m} p_{0j} \int_0^\infty \theta_i^2 f_{0i}(\theta_i; a_i + c_j, b_i) d\theta_i \\ &= \sum_{j=0}^{m} p_{0j} \frac{(a_i + c_j)(a_i + c_j + 1)}{b_i^2}. \end{aligned}$$

Finally, we find the mixed second moment to obtain the prior covariance between θ_1 and θ_2.

$$\begin{aligned} \mathrm{E}_0(\theta_1 \theta_2) &= \sum_{j=0}^{m} p_{0j} \int_0^\infty \theta_1 f_{01}(\theta_1; a_1 + c_j, b_1) d\theta_1 \int_0^\infty \theta_2 f_{02}(\theta_2; a_2 + c_j, b_2) d\theta_2 \\ &= \sum_{j=0}^{m} p_{0j} \frac{(a_1 + c_j)}{b_1} \frac{(a_2 + c_j)}{b_2}. \end{aligned}$$

If we observe the likelihood in Equation 2.3 then, by Bayes theorem, the posterior joint

density is

$$f_1(\theta_1, \theta_2) = \sum_{j=0}^{m} p_{1j} f_{11}(\theta_1; A_1 + c_j, B_1) f_{12}(\theta_2; A_2 + c_j, B_2),$$

where the posterior weights are proportional to

$$\hat{p}_{1k} = p_{0k} \frac{b_1^{a_1+c_k}}{(b_1+1)^{a_1+c_k+x_1}} \frac{b_2^{a_2+c_k}}{(b_2+1)^{a_2+c_k+x_2}} \frac{\Gamma(a_1+c_k+x_1)}{\Gamma(a_1+c_k)} \frac{\Gamma(a_2+c_k+x_2)}{\Gamma(a_2+c_k)}.$$

The posterior weights are then $p_{1k} = \dfrac{\hat{p}_{1k}}{\sum_{j=0}^{m} p_{1j}}$.

The posterior gamma parameters are $A_1 = a_1 + x_1$, $B_1 = b_1 + 1$, $A_2 = a_2 + x_2$ and $B_2 = b_2 + 1$. Thus posterior moments are as above but using these new parameter values and weights.

2.13.1 Example: piston ring failures

Suppose we wish to apply this model to the Poisson example involving piston ring failures. The prior specifications we wish to make are given in Equation 2.12.

Setting the prior parameter values to $a_1 = a_2 = 18.8$, $b_1 = b_2 = 1.34$, $p^* = 0.37$ and $m = 58$ leads to a 59 component mixture distribution with prior means and variances

$$\mathrm{E}_0(\theta_i) = 30.0, \mathrm{Var}_0(\theta_i) = 30.0,$$

and a prior correlation between θ_1 and θ_2 of 0.251. This gives a prior joint density as in the left hand side of Figure 2.12. Having observed 46 failures in compressor 1 and

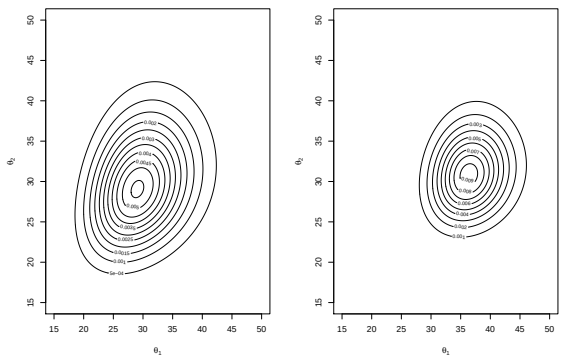

Figure 2.12: Prior and posterior contour plots for the gamma finite mixture model

33 failures in compressor 2 the posterior parameter values are $A_1 = 64.8$, $B_1 = 2.34$, $A_2 = 51.8$ and $B_2 = 1.34$ with p^* and m as before. This leads to a joint posterior

density as in the right hand side of Figure 2.12.

The posterior means and variances of θ_1 and θ_2 are

$$\begin{aligned} \mathrm{E}_1(\theta_1) = 38.02, \quad & \mathrm{E}_1(\theta_2) = 32.47 \\ \mathrm{Var}_1(\theta_1) = 18.21, \quad & \mathrm{Var}_1(\theta_2) = 15.83, \end{aligned}$$

and the posterior correlation is 0.115. With both observed numbers of failures being higher than their prior expectations the means of both parameters have increased. Our uncertainty has been reduced up on observation of the data. The posterior correlation is lower than its prior counterpart.

2.13.2 Discussion

All of the non-copula density multipliers we have investigated have led to mixture distributions. The first, $g(\theta_1, \theta_2) = 1 + k\theta_1^{m_1}\theta_2^{m_2}$ produced a two component mixture in both the beta-binomial and gamma-Poisson cases. The other, which was related to probabilities associated with observing different numbers of "prior successes" had a number of components determined by the parameter m.

Using any of the proposed mixtures prior specification is difficult as the marginal densities for θ_1 and θ_2 are not the beta or gamma densities $f_{01}(\theta_1)$ and $f_{02}(\theta_2)$. For example, for the two component mixture the joint density is

$$f_0(\theta_1, \theta_2) = p_{01} f_{011}(\theta_1) f_{012}(\theta_2) + (1 - p_{01}) f_{021}(\theta_1) f_{022}(\theta_2),$$

and so the marginal density of θ_1 is

$$h_0(\theta_1) \quad = \quad p_{01} f_{011}(\theta_1) + (1 - p_{01}) f_{021}(\theta_1),$$

This lack of marginality and subsequent difficulty in prior specification would become a real problem in high dimensions. Some method to overcome this will have to be found for density multipliers which lead to mixtures to be widely applicable.

Clearly a density multiplier which leads to a two component mixture results in simpler calculations than that which produces a mixture of more than two components. However, with relatively few components specification of a fairly strong relationship between θ_1 and θ_2 can lead to a joint density which is bimodal. This bimodality can be overcome by increasing the number of components in the mixture.

2.14 Dirichlet approaches to the binomial case

Let us consider the case of correlated binomial parameters. Thinking about the number of successes in group j, X_j, they can be separated into n_j Bernoulli trials. We will, without loss of generality, consider the groups to be different treatments. For each such trial, $i = 1, \ldots, n_j$,

$$X_{ij} = \begin{cases} 1, & \text{with probability } \theta_{ij}, \\ 0, & \text{with probability } 1 - \theta_{ij}. \end{cases}$$

We could then imagine a population of individuals of four types as in Table 2.7.

Outcome		
Treatment 1	Treatment 2	Probability
1	1	π_{11}
1	0	π_{10}
0	1	π_{01}
0	0	π_{00}

Table 2.7: Truth table for successes in 2 Bernoulli trials

That is, an individual would record a success under both treatments with probability π_{11}, a success in treatment 1 and failure in treatment 2 with probability π_{10}, a failure in treatment 1 and success in treatment 2 with probability π_{01} and failures in both treatments with probability π_{00}.

Clearly $\pi_{11} + \pi_{10} + \pi_{01} + \pi_{00} = 1$ and so we can give the four probabilities a Dirichlet prior distribution,

$$\boldsymbol{\pi} \sim \text{Dir}(\boldsymbol{a}),$$

for some parameter vector $\boldsymbol{a} = (a_{11}, a_{10}, a_{01}, a_{00})'$, so that their joint density is $f_0(\boldsymbol{\pi}) \propto \prod_{k,l=0,1} \pi_{kl}^{a_{kl}-1}$. The probability an individual would record a success under treatment 1 (θ_1) and treatment 2 (θ_2) are then

$$\begin{aligned} \theta_1 &= \pi_{11} + \pi_{10} \quad 1 - \theta_1 = \pi_{01} + \pi_{00}, \\ \theta_2 &= \pi_{11} + \pi_{01} \quad 1 - \theta_2 = \pi_{10} + \pi_{00}. \end{aligned}$$

Given x_1 successes out of n_1 trials under treatment 1 and x_2 successes out of n_2 trials under treatment 2 the likelihood is

$$L(\theta_1, \theta_2) = \begin{pmatrix} n_1 \\ x_1 \end{pmatrix} \theta_1^{x_1} (1 - \theta_1)^{n_1 - x_1} \begin{pmatrix} n_2 \\ x_2 \end{pmatrix} \theta_2^{x_2} (1 - \theta_2)^{n_2 - x_2},$$

which becomes

$$L(\boldsymbol{\pi}) = \begin{pmatrix} n_1 \\ x_1 \end{pmatrix} (\pi_{11} + \pi_{10})^{x_1} (\pi_{01} + \pi_{00})^{n_1 - x_1} \begin{pmatrix} n_2 \\ x_2 \end{pmatrix} (\pi_{11} + \pi_{01})^{x_2} (\pi_{10} + \pi_{00})^{n_2 - x_2}. \tag{2.18}$$

Thus the posterior distribution will be a finite mixture of Dirichlet distributions. Unfortunately, this model isn't sufficiently flexible: in the heart attack example we wish to make five prior specifications and we have just four parameters. In order for a Dirichlet model to be useful then a more flexible form will need to be found.

2.14.1 Aitchison A-class distributions

The Dirichlet distribution is a special case of the more general Aitchison A-class of distributions (Aitchison, 1986). The log-density of the Dirichlet distribution is

$$\log [f(\boldsymbol{\pi})] \propto \sum_{i=1}^{p} (a_i - 1) \log \pi_i,$$

for parameter vector $\boldsymbol{\pi}$. The new class is defined to incorporate both the Dirichlet and logistic-Normal distributional forms. Its probability density function is $f(\boldsymbol{\pi})$ where

$$\log [f(\boldsymbol{\pi})] \propto \sum_{i=1}^{p} (a_i - 1) \log \pi_i - \frac{1}{2} \sum_{i=1}^{p} \sum_{j=i+1}^{p} \beta_{ij} (\log \pi_i - \log \pi_j)^2,$$

Thus extra parameters β_{ij} are introduced. In order for this to be a proper density one of two conditions must hold (Aitchison, 1986). Either

(i) the quadratic form must be positive definite and $a_1 + \ldots + a_p \geq 0$, or

(ii) the quadratic form must be non-negative definite and $a_i > 0$ for all $i = 1, \ldots, p$.

The class is a conjugate prior to the multinomial distribution but its moments are not analytically tractable. The density is

$$f(\boldsymbol{\pi}) \propto \left[\prod_{i=1}^{p} \pi_i^{a_i - 1} \right] \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^{p} \sum_{j=i+1}^{p} \beta_{ij} (\log \pi_i - \log \pi_j)^2 \right\}.$$

This is not conjugate to the likelihood given in Equation 2.18 and so this class is not suitable for our needs.

2.15 Hierarchical modelling

Another possible fully Bayesian approach to the beta-binomial case would be to take a hierarchical model. If the prior distributions for θ_1 and θ_2 were

$$\theta_1 \sim \text{beta}(a_1\mu, b_1[1-\mu]), \quad \theta_2 \sim \text{beta}(a_2\mu, b_2[1-\mu]),$$

for some parameter μ, then μ could also be given a beta prior distribution so that

$$\mu \sim \text{beta}(a_\mu, b_\mu),$$

for hyperparameters a_μ and b_μ. The joint density for $\boldsymbol{\theta} = (\theta_1, \theta_2, \mu)'$ takes the form

$$\begin{aligned} f(\boldsymbol{\theta}) = \frac{\Gamma(a_\mu + b_\mu)}{\Gamma(a_\mu)\Gamma(b_\mu)}\mu^{a_\mu - 1}(1-\mu)^{b_\mu - 1} \times \frac{\Gamma(a_1\mu + b_1[1-\mu])}{\Gamma(a_1\mu)\Gamma(b_1[1-\mu])}\mu^{a_1\mu - 1}(1-\mu)^{b_1[1-\mu]-1} \\ \times \frac{\Gamma(a_2\mu + b_2[1-\mu])}{\Gamma(a_2\mu)\Gamma(b_2[1-\mu])}\mu^{a_2\mu - 1}(1-\mu)^{b_2[1-\mu]-1} \end{aligned}$$

Thus updates using the likelihood in Equation 2.2 will be conjugate. However, due to the μ terms inside the gamma functions associated with θ_1 and θ_2, calculations to obtain such things as marginal distributions and moments for θ_1 and θ_2 are analytically intractable.

We could define a similar sort of structure in the gamma-Poisson case. If we define our Poisson parameters to be

$$\begin{aligned} \theta_1 &= U + E_1 \\ \theta_2 &= U + E_2, \end{aligned}$$

then we could give each of U, E_1, E_2 gamma prior distributions. Thus,

$$U \sim \text{gamma}(a_U, b), \quad E_i \sim \text{gamma}(a_i, b),$$

for $i = 1, 2$. We see that it is necessary to use the same scale parameter b for each of the distributions, however. This leaves only 4 free parameters in the model which is a little overly restrictive.

2.16 Conclusions

In this chapter we have investigated fully Bayesian methods for modelling correlated binomial probabilities or Poisson parameters without the need for intensive numerical

calculations such as MCMC. Initially we considered copula functions and it became apparent that the only family suitable to our needs was the FGM family. This copula family produced a solution to the problem in both the beta-binomial and gamma-Poisson cases but the usefulness of the FGM copula is limited by the restriction in prior correlation it is possible to specify. To overcome this two extensions were investigated which increased the range of possible correlations, if not by much. The marginality property is also lost in the posterior.

We then considered mixtures. Two mixtures were used to model both the beta-binomial and gamma-Poisson cases and both produced solutions to the problem in which posterior joint densities were very similar to those produced for the copula model. It was felt that whilst keeping the number of components low in the mixture distribution would allow simpler calculations there was an issue with bimodality in joint densities using mixtures with few components.

Prior specifications could not be made simply in general with density multipliers, copulas being the exception, due to a lack of marginality and it is felt that this is something which would have to be overcome before they could be of practical use in higher dimensions.

As a comparison between the two models Tables 2.8 and 2.9 give posterior expectations and variances for the two parameters using each model for the heart attack and piston ring examples respectively.

Method	$E_1(\theta_1)$	$E_1(\theta_2)$	$Var_1(\theta_1)$	$Var_1(\theta_2)$
FGM copula	0.081	0.114	1.26×10^{-4}	1.56×10^{-4}
2-component	0.080	0.114	1.28×10^{-4}	1.59×10^{-4}
33-component	0.080	0.114	1.27×10^{-4}	1.58×10^{-4}

Table 2.8: A comparison of the different fully Bayesian models in the beta-binomial case

In Table 2.8 all of the posterior moments are very similar using the three methods. This would appear to be due to the large number of observations and relatively weak correlation.

Method	$E_1(\theta_1)$	$E_1(\theta_2)$	$Var_1(\theta_1)$	$Var_1(\theta_2)$
FGM copula	37.50	30.89	19.83	15.38
2-component	38.28	32.72	18.67	16.30
59-component	38.02	32.47	18.21	15.83

Table 2.9: A comparison of the different fully Bayesian models in the gamma-Poisson case

In contrast in Table 2.9 there are some differences between the posterior moments using

the three methods. The results for the two mixtures are very similar but generally the posterior estimates for the two groups are closer together than when using the FGM copula.

Chapter 3

Bayes linear approaches

3.1 Introduction

In this chapter we apply Bayes linear approaches to the problem of inference for two correlated binomial or Poisson parameters. Initially we consider a model for the binomial case which takes advantage of properties of second order exchangeability between Bernoulli trials within a group. We then consider models based upon the idea of Bayes linear kinematics, a form of Bayes linear analysis in which changes in belief about some quantities are propagated through to others within a Bayes linear structure. The use of transformations of the unknown binomial or Poisson parameters is proposed. We investigate several models involving such transformations and then one in which the parameter is not transformed to allow for comparisons between models to be made.

3.2 Bayes linear methods

In a traditional Bayesian analysis a full joint prior distribution is specified for all observables and unknown quantities such as parameters. Prior beliefs are then updated, by conditioning on the observations and using Bayes theorem, and posterior distributions are calculated.

A Bayes linear analysis (Goldstein & Wooff, 2007) differs from a full Bayesian analysis in that only first and second order moments are specified in the prior. Posterior (termed *adjusted*) moments are then calculated. For example, for each quantity Q in the analysis we specify its prior expectation and variance

$$\mathrm{E}_0(Q), \quad \mathrm{Var}_0(Q),$$

and for every two quantities Q_1 and Q_2, a prior covariance

$$\mathrm{Cov}_0(Q_1, Q_2),$$

is specified. Consider two vector random quantities $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)'$, where $\boldsymbol{\alpha}$ is a collection of quantities which shall be observed and $\boldsymbol{\beta}$ is a collection of quantities about which inferences are to be made. Suppose that a full second order prior specification has been made for the set $\boldsymbol{A} = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$.

Bayes linear methods (Goldstein & Wooff, 2007) offer a procedure by which beliefs about $\boldsymbol{\beta}$ are updated by a process of linear fitting on $\boldsymbol{\alpha}$. To do this, we minimise the expected squared loss. That is, we minimise

$$\mathrm{E}_0\left(\left[\boldsymbol{\beta} - \sum_{i=0}^{p} c_i \alpha_i\right]^2\right),$$

for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ with respect to $c_0, c_1, \ldots, c_p$, where $\alpha_0 = 1$. This gives the Bayes linear updating equations for the adjusted expectation and variance of $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$:

$$\begin{aligned} \mathrm{E}_1(\boldsymbol{\beta}; \boldsymbol{\alpha}) &= \mathrm{E}_0(\boldsymbol{\beta}) + \mathrm{Cov}_0(\boldsymbol{\beta}, \boldsymbol{\alpha}) \mathrm{Var}_0^{-1}(\boldsymbol{\alpha})[\boldsymbol{\alpha} - \mathrm{E}_0(\boldsymbol{\alpha})] \\ \mathrm{Var}_1(\boldsymbol{\beta}; \boldsymbol{\alpha}) &= \mathrm{Var}_0(\boldsymbol{\beta}) - \mathrm{Cov}_0(\boldsymbol{\beta}, \boldsymbol{\alpha}) \mathrm{Var}_0^{-1}(\boldsymbol{\alpha}) \mathrm{Cov}_0(\boldsymbol{\alpha}, \boldsymbol{\beta}), \end{aligned}$$

when $\mathrm{Var}_0(\boldsymbol{\alpha})$ is invertible. When this matrix is not invertible a suitable generalised inverse such as the Moore-Penrose inverse can be used.

In real world problems there are often many quantities, both unknown and observable, for which it is necessary to make full second order prior specifications. In practice it is often unrealistic to make all of these specifications individually and so properties of exchangeability between quantities can be utilised.

Definition. If we have a collection of random quantities (say vectors) $\boldsymbol{Q} = (\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_k)$ then we say that this set is **exchangeable** if any other ordering of the set would not alter our beliefs, in the form of the joint density $f_0(\boldsymbol{Q})$, about $\boldsymbol{Q}$.

This requirement is fairly restricive and so in Bayes linear statistics a weaker form, second order exchangeability, has been utilised. This second order exchangeability is often sufficient to reduce the burdens associated with prior specifications whilst maintaining a representation which is consistent with an expert's prior beliefs.

Definition. We say that the collection $\boldsymbol{Q}$ is **second order exchangeable** if its first and second order belief specifications would not alter under any re-ordering of $\boldsymbol{Q}$. That

is, for some constants $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$,

$$\mathrm{E}_0(\boldsymbol{Q}_i) = \boldsymbol{c}_1, \tag{3.1}$$

$$\mathrm{Var}_0(\boldsymbol{Q}_i) = \boldsymbol{c}_2, \tag{3.2}$$

$$\mathrm{Cov}_0(\boldsymbol{Q}_i, \boldsymbol{Q}_j) = \boldsymbol{c}_3, \tag{3.3}$$

for all $i, j \in \boldsymbol{Q}$, where $i \neq j$.

Thus in second order exchangeable collections of quantities we assume equal expectations, variances and covariances for the random quantities in $\boldsymbol{Q}$.

Now let us suppose we assume a second order exchangeable structure for our vector of unknown quantities $\boldsymbol{Q}$. Goldstein (1986) gives a representation theorem, repeated in the following form in Goldstein & Wooff (2007), which shows how models for such second-order exchangeable structures can be created.

Theorem 3.1. *If $\boldsymbol{Q} = \boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots$ is an infinite second-order exchangeable sequence of random vectors, with mean and variance structure given as in Equations 3.1-3.3 then we may introduce the further random vector $\boldsymbol{M}(\boldsymbol{Q})$, the population mean vector, and the infinite sequence $\boldsymbol{R}_1(\boldsymbol{Q}), \boldsymbol{R}_2(\boldsymbol{Q}), \ldots$, termed the individual residual vectors, which satisfy the following properties.*

(i) For each individual i

$$\boldsymbol{Q}_i = \boldsymbol{M}(\boldsymbol{Q}) + \boldsymbol{R}_i(\boldsymbol{Q}),$$

where $\boldsymbol{M}(\boldsymbol{Q})$ has mean $E[\boldsymbol{M}(\boldsymbol{Q})] = \boldsymbol{c}_1$ and variance $Var(\boldsymbol{M}(\boldsymbol{Q})) = \boldsymbol{c}_3$.

(ii) The collection $\boldsymbol{R}_1(\boldsymbol{Q}), \boldsymbol{R}_2(\boldsymbol{Q}), \ldots$ is second-order exchangeable with

$$E[\boldsymbol{R}_i(\boldsymbol{Q})] = \boldsymbol{0}, \quad Var(\boldsymbol{R}_i(\boldsymbol{Q})) = \boldsymbol{c}_2 - \boldsymbol{c}_3 \quad Cov(\boldsymbol{R}_i(\boldsymbol{Q}), \boldsymbol{R}_j(\boldsymbol{Q})) = \boldsymbol{0},$$

for $i \neq j$. Also $Cov(\boldsymbol{R}_i(\boldsymbol{Q}), \boldsymbol{M}(\boldsymbol{Q})) = \boldsymbol{0}$.

3.3 Second order exchangeable model

The methodology in this section follows closely to that used by Coolen *et al.* (2001) in their application of Bayes linear methods to software partition testing.

Let us suppose that we have a number of individuals split into 2 groups with n_1 individuals in group 1 and n_2 in group 2. Each undertakes the same task and a success

or failure is recorded. So, for individual j in group i,

$$X_{ij} = \begin{cases} 1, & \text{with probability } p_{ij}, \\ 0, & \text{with probability } 1 - p_{ij}, \end{cases}$$

for $i = 1, 2$ where p_{ij} is the probability of a success. Thus we have $X_{1j} \sim \text{Bern}(p_{1j})$ and $X_{2j} \sim \text{Bern}(p_{2j})$. Within each group we can count the number of successes, X_i, so that

$$X_i = \sum_{j=1}^{n_i} X_{ij}.$$

Now, in order to perform a Bayes linear analysis directly on the X_{ij}'s, it would be necessary to specify $N = n_1 + n_2$ expectations and $\frac{N(N+1)}{2}$ variances and covariances. In practice this would become infeasible for reasonably large N. Thus instead we shall assume that individuals within a group are second-order exchangeable so that their prior moments are given by

$$\begin{aligned} \text{E}_0(X_{ij}) &= p_i, \\ \text{Var}_0(X_{ij}) &= p_i(1 - p_i), \\ \text{Cov}_0(X_{ij}, X_{il}) &= v_i, \end{aligned}$$

for $l \neq j$. Let us further suppose that there is some constant covariance between individuals in different groups, namely

$$\text{Cov}_0(X_{1j}, X_{2l}) = v_{12}.$$

Now, as we have specified a second-order exchangeable structure for individuals in each group we can apply the representation theorem, 3.1. Thus, for X_{ij},

$$X_{ij} = M_i + \epsilon_{ij},$$

with mean component M_i and residual component ϵ_{ij}. Therefore the prior means and variances of M_i and ϵ_{ij} are

$$\begin{aligned} &\text{E}_0(M_i) = p_i, \quad \text{Var}_0(M_i) = v_i, \\ &\text{E}_0(\epsilon_{ij}) = 0, \quad \text{Var}_0(\epsilon_{ij}) = p_i(1 - p_i) - v_i, \end{aligned}$$

where the residual components are mutually uncorrelated and are uncorrelated with the mean components so that $\text{Cov}_0(M_i, \epsilon_{ij}) = 0$ and $\text{Cov}_0(\epsilon_{ij}, \epsilon_{kl}) = 0$. The mean

components are related, however, with

$$\begin{aligned} \mathrm{Cov}_0(M_1, M_2) &= \mathrm{Cov}_0(X_{1j} - \epsilon_{1j}, X_{2l} - \epsilon_{2l}) \\ &= v_{12}. \end{aligned}$$

We can think of M_i as the unknown population average proportion of successes in group i and this is the quantity in which we shall be interested. Thus we will choose to learn about the set $\boldsymbol{M} = (M_1, M_2)'$.

Given the above setup we can also calculate some other covariances.

$$\begin{aligned} \mathrm{Cov}_0(M_i, X_{ij}) &= \mathrm{Cov}_0(X_{il} - \epsilon_{il}, X_{ij}) = v_i, \\ \mathrm{Cov}_0(M_i, X_{kj}) &= \mathrm{Cov}_0(X_{il} - \epsilon_{il}, X_{kj} - \epsilon_{kj}) = v_{12}, \end{aligned}$$

for $k \neq i$. As a result of the exchangeable structure we have assumed within each group X_1 and X_2 are sufficient statistics for all of the information found in the individual Bernoulli trials X_{ij}. To see this let us return to the definition of second-order exchangeability. As all of the Bernoulli trials in a group are second-order exchangeable this means that a relabelling of the trials would not affect our beliefs up to second order about them. Thus the order of the trials does not contain any information and so the only information is to be found in the number of successful trials. We shall prove this formally below.

Thus we shall use Bayes linear methods to update $\boldsymbol{M} = (M_1, M_2)'$ using $\boldsymbol{X} = (X_1, X_2)'$. In order to carry out these updates we shall need prior expectations, variances and a covariance for X_1 and X_2. Their expectations are given by

$$\mathrm{E}_0(X_i) = \mathrm{E}_0\left(\sum_{j=1}^{n_i} X_{ij}\right) = \sum_{j=1}^{n_i} \mathrm{E}_0(X_{ij}) = n_i p_i,$$

with prior variances,

$$\begin{aligned} \mathrm{Var}_0(X_i) &= \mathrm{Var}_0\left(\sum_{j=1}^{n_i} X_{ij}\right) \\ &= \sum_{j=1}^{n_i} \mathrm{Var}_0(X_{ij}) + \sum\sum_{j \neq k} \mathrm{Cov}_0(X_{ij}, X_{ik}) \\ &= n_i p_i (1 - p_i) + 2 \binom{n_i}{2} v_i. \end{aligned}$$

The necessary covariances can also be calculated. The within-group covariance is

$$\mathrm{Cov}_0(M_i, X_i) = \mathrm{Cov}_0\left(M_i, \sum_{j=1}^{n_i} X_{ij}\right) = \sum_{j=1}^{n_i} \mathrm{Cov}_0(M_i, X_{ij}) = n_i v_i,$$

and the between-group covariance is found to be

$$\mathrm{Cov}_0(M_i, X_k) = \mathrm{Cov}_0\left(M_i, \sum_{j=1}^{n_k} X_{kj}\right) = \sum_{j=1}^{n_k} \mathrm{Cov}_0(M_i, X_{kj}) = n_k v_{12}.$$

Finally we need the prior covariance between X_1 and X_2. This is

$$\mathrm{Cov}_0(X_1, X_2) = \mathrm{Cov}_0\left(\sum_{j=1}^{n_1} X_{1j}, \sum_{l=1}^{n_2} X_{2l}\right) = \sum_{j=1}^{n_1}\sum_{l=1}^{n_2} \mathrm{Cov}_0(X_{1j}, X_{2l}) = n_1 n_2 v_{12}.$$

Thus all of the prior specifications for X_1 and X_2 can be made in terms of the values already specified for the X_{ij}'s.

We shall now show the sufficiency more formally. Bayes linear sufficiency is defined in Goldstein & Wooff (2007).

Definition. If $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are 3 collections of random quantities then $\boldsymbol{\gamma}$ is Bayes linear sufficient for $\boldsymbol{\alpha}$ for adjusting $\boldsymbol{\beta}$ if $\mathrm{E}_1(\boldsymbol{\beta}; \boldsymbol{\alpha} \cup \boldsymbol{\gamma}) = \mathrm{E}_1(\boldsymbol{\beta}; \boldsymbol{\gamma})$.

We can then show the sufficiency of X_1 and X_2 using the following theorem, also from Goldstein & Wooff (2007).

Theorem 3.2. *If $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are 3 belief structures then $E_1(\boldsymbol{\beta}; \boldsymbol{\alpha} \cup \boldsymbol{\gamma}) = E_1(\boldsymbol{\beta}; \boldsymbol{\gamma})$ is equivalent to*

$$Cov(\boldsymbol{\beta}, \boldsymbol{\alpha}) = Cov(\boldsymbol{\beta}, \boldsymbol{\gamma})\, Var^{-1}(\boldsymbol{\gamma})\, Cov(\boldsymbol{\gamma}, \boldsymbol{\alpha}). \tag{3.4}$$

We consider, without loss of generality, the case of $i = 1$. Set

$$\begin{array}{rcccl} \boldsymbol{\alpha} & = & \boldsymbol{X}_1 & = & (X_{11}, \ldots, X_{1n_1})', \\ \boldsymbol{\beta} & = & \boldsymbol{M} & = & (M_1, M_2)', \\ \boldsymbol{\gamma} & = & X_1 & = & \sum_{j=1}^{n_1} X_{1j}. \end{array}$$

Then

$$\mathrm{Cov}_0(\boldsymbol{M}, X_1) = \begin{bmatrix} \mathrm{Cov}_0(M_1, \sum_j X_{1j}) \\ \mathrm{Cov}_0(M_2, \sum_j X_{1j}) \end{bmatrix} = \begin{bmatrix} n_1 v_1 \\ n_1 v_{12} \end{bmatrix}.$$

$\mathrm{Cov}_0(\boldsymbol{M}, \boldsymbol{X}_1)$ is a $(2 \times n_1)$ matrix the $[i, j]$'th element of which is

$$\mathrm{Cov}_0(M_i, X_{1j}) = \begin{cases} v_1, \text{ if } i = 1, \\ v_{12}, \text{ if } i = 2. \end{cases}$$

$\mathrm{Cov}_0(X_1, \boldsymbol{X}_1)$ is a $(1 \times n_1)$ vector all of the elements of which are $p_1(1-p_1)+(n_1-1)v_1$. Thus the right hand side of Equation 3.4 becomes

$$\begin{aligned} \mathrm{Cov}_0(\boldsymbol{M}, X_1)\mathrm{Var}_0^{-1}(X_1)\mathrm{Cov}_0(X_1, \boldsymbol{X}_1) &= \frac{1}{k_1}\mathrm{Cov}_0(\boldsymbol{M}, X_1)\mathrm{Cov}_0(X_1, \boldsymbol{X}_1) \\ &= \frac{k_2}{k_1}\mathrm{Cov}_0(\boldsymbol{M}, \boldsymbol{X}_1), \end{aligned}$$

where $\dfrac{k_2}{k_1} = \dfrac{n_1 p_1(1-p_1) + n_1(n_1-1)v_1}{n_1 p_1(1-p_1) + 2\mathrm{C}_2^{n_1} v_1} = 1$. Therefore, by Theorem 3.2, X_1 is Bayes linear sufficient for $\boldsymbol{X}_1$ for adjusting $\boldsymbol{M}$.

3.3.1 Example: Heart attack data

We are now in a position to apply our Bayes linear modelling approach to the heart attack data given in Section 2.6.3. Our first task is to specify the prior parameters p_i and v_i for $i = 1, 2$ and v_{12}. We shall make similar specifications to the models in Chapter 2 for comparability.

We find p_1 and p_2, the probabilities of patients dying in groups 1 and 2 respectively, by setting them to the values of their prior expectations in previous models.

$$\mathrm{E}_0(X_{ij}) = p_i = 0.125 \;\Rightarrow\; \mathrm{Var}_0(X_{ij}) = p_i(1-p_i) = 0.109375,$$

for $i = 1, 2$. The other prior values can be specified by noting that for binary random variables Y_1 and Y_2 their covariance can be expressed as

$$\begin{aligned} \mathrm{Cov}(Y_1, Y_2) &= \Pr(Y_2 = 1)\Pr(Y_1 = 1 \mid Y_2 = 1) - \Pr(Y_1 = 1)\Pr(Y_2 = 1) \\ &= \Pr(Y_2 = 1)[\Pr(Y_1 = 1 \mid Y_2 = 1) - \Pr(Y_1 = 1)]. \end{aligned}$$

Thus we can assess the 3 remaining covariances by eliciting probabilities in the usual subjective manner in terms of the fair price of gambles (De Finetti, 1974, 1975). Suppose we do this and in particular we wish to have stronger correlation between variables within a group than between variables in different groups. Thus, following the elicitation process, it is decided that

$$\mathrm{Cov}_0(X_{1j}, X_{1l}) = v_1 = 0.003125, \quad \mathrm{Cov}(X_{2j}, X_{2l}) = v_2 = 0.001875$$

and

$$\mathrm{Cov}(X_{1j}, X_{2l}) = v_{12} = 0.00125.$$

From these 5 quantities we can calculate the prior moments for M_1 and M_2. These are $\mathrm{E}_0(M_1) = \mathrm{E}_0(M_2) = 0.125$, $\mathrm{Var}_0(M_1)$=0.003125, $\mathrm{Var}_0(M_2)$=0.001875 and $\mathrm{Cov}_0(M_1, M_2)$=0.00125.

To complete our prior specifications we need the prior moments for X_1 and X_2 and the strengths of their relationships with M_1 and M_2. Noting that $n_1 = 560$ and $n_2 = 540$, they are given by $\mathrm{E}_0(X_1) = 70$, $\mathrm{Var}_0(X_1) = 1039.5$, $\mathrm{Cov}_0(M_1, X_1) = 1.75$ and $\mathrm{Cov}_0(M_2, X_1) = 0.7$ for X_1 and $\mathrm{E}_0(X_2) = 67.5$, $\mathrm{Var}_0(X_2) = 604.8$, $\mathrm{Cov}_0(M_1, X_2) = 0.675$ and $\mathrm{Cov}_0(M_2, X_2) = 1.0125$ for X_2. The prior covariance between X_1 and X_2 is 378.

We are now in a position to carry out a Bayes linear analysis on the heart attack data. All of the following calculations have been performed in the computer package designed for carrying out Bayes linear analyses [B/D] (Wooff & Goldstein, 2000; Wooff, 2000). Using this software the adjusted expectations of M_1 and M_2 having observed $\boldsymbol{X} = (X_1, X_2)'$ are found to be

$$\begin{aligned}\mathrm{E}_1(M_1; \boldsymbol{X}) &= 0.0017X_1 + 0.0001X_2 + 0.0037\\ \mathrm{E}_1(M_2; \boldsymbol{X}) &= 0.0001X_1 + 0.0016X_2 + 0.0097.\end{aligned}$$

The standardised adjusted expectations are then calculated as

$$\begin{aligned}\mathrm{E}_1(M_1; \boldsymbol{X}) &= 0.0533X_1^* + 0.0020X_2^* + 0.125\\ \mathrm{E}_1(M_2; \boldsymbol{X}) &= 0.0027X_1^* + 0.0399X_2^* + 0.125,\end{aligned}$$

where $X_i^* = \frac{X_i - \mathrm{E}_0(X_i)}{\sqrt{\mathrm{Var}_0(X_i)}}$ is the standardised quantity of X_i. It can be seen from the standardised adjusted expectations that we learn most about M_1 through X_1 and we learn most about M_2 through X_2. However the increases in knowledge about M_1 through X_2 and M_2 through X_1 though smaller are not insignificant. That is, our beliefs about the death rate in group 1 are being updated by the data from group 1 and group 2, as are our beliefs about the death rate in group 2.

The adjusted variances can then be found and are given by

$$\begin{aligned}\mathrm{Var}_1(M_1; \boldsymbol{X}) &= 0.0002\\ \mathrm{Var}_1(M_2; \boldsymbol{X}) &= 0.0002,\end{aligned}$$

leading, through the equation $\mathrm{Var}_0(M_i)$ =$\mathrm{Var}_1(M_i; \boldsymbol{X})$+$\mathrm{RVar}_1(M_i; \boldsymbol{X})$, to resolved

variances of

$$\begin{aligned}\mathrm{RVar}_1(M_1; \boldsymbol{X}) &= 0.0029\\ \mathrm{RVar}_1(M_2; \boldsymbol{X}) &= 0.0017,\end{aligned}$$

and resolutions of

$$\begin{aligned}\mathrm{R}_1(M_1; \boldsymbol{X}) &= 0.9438\\ \mathrm{R}_1(M_2; \boldsymbol{X}) &= 0.9009.\end{aligned}$$

The resolution $R_1(M_i; \boldsymbol{X}) = \frac{\mathrm{RVar}_1(M_i; \boldsymbol{X})}{\mathrm{Var}_0(M_i)}$ and represents the proportion of uncertainty about M_i which is resolved by observing $\boldsymbol{X}$. Thus 94.38% of the uncertainty about M_1 and 90.09% of the uncertainty about M_2 have been resolved and so a great deal has been learnt about the efficacy of the new drug. The two canonical directions, the linear combinations of M_1 and M_2 which lead to the largest and smallest resolutions, C_1 and C_2, for the heart attack data are given by

$$\begin{aligned}C_1 &= 13.8714M_1 + 8.0196M_2 - 2.7364\\ C_2 &= 15.6188M_1 - 25.7480M_2 + 1.2662,\end{aligned}$$

and their resolutions are

$$\begin{aligned}\mathrm{R}_1(C_1; \boldsymbol{X}) &= 0.9530\\ \mathrm{R}_1(C_2; \boldsymbol{X}) &= 0.8487.\end{aligned}$$

The canonical directions are automatically rescaled by [B/D] to have mean zero and unit variance. Hence we can expect to resolve at most 95.3% of the uncertainty about any linear combination of M_1 and M_2 and we will always resolve at least 84.87%. That is, we will always significantly reduce the uncertainty about whichever combination of M_1 and M_2 we are interested in.

In our heart attack data 44 patients died out of a total of 560 in group 1 and 62 patients died out of a total of 540 in group 2. This means that our observations are $x_1 = 44$ and $x_2 = 62$. When we carry out the Bayes linear analysis using these values we see that our adjusted expectations having observed $\boldsymbol{x} = (x_1, x_2)'$ are

$$\begin{aligned}\mathrm{E}_1(M_1; \boldsymbol{x}) &= 0.0816\\ \mathrm{E}_1(M_2; \boldsymbol{x}) &= 0.1139.\end{aligned}$$

These are both lower than their corresponding prior expectations of 0.125. The adjusted expectation for M_1 is also quite a lot smaller than the adjusted expectation for M_2 implying that there could be a difference between the death rates for the two groups.

We are investigating the heart attack data in order to try and answer the question 'is the drug sulfinpyrazone reducing the number of patients dying following heart attacks?' In order to answer this question it will be useful to look at the difference between the

two death rates

$$L = M_1 - M_2.$$

As this is a linear combination of M_1 and M_2 we expect to reduce our uncertainty about it by between 85% and 95% (our minimum and maximum resolutions).

The prior expectation and variance of L are given by

$$\mathrm{E}_0(L) = 0, \quad \mathrm{Var}_0(L) = 0.0025$$

and we see that the initial expectation of L is zero. The adjusted expectation of L is given by

$$\mathrm{E}_1(L; \boldsymbol{X}) = 0.00165X_1 - 0.0015X_2 - 0.06$$

and the observed adjusted expectation is found to be

$$\mathrm{E}_1(L; \boldsymbol{x}) = -0.0323.$$

That is, we now expect the number of patients dying given the drug to be lower than the number dying given the placebo. The adjusted variance is then

$$\mathrm{Var}_1(L; \boldsymbol{X}) = 0.0003.$$

We can take intervals of two standard deviations from the adjusted expectation to give us an idea of where L might reasonably be expected to lie;

$$\mathrm{E}_1(L; \boldsymbol{x}) \pm 2\sqrt{\mathrm{Var}_1(L; \boldsymbol{X})} = (-0.067, 0.00234).$$

Zero is in this interval and so on this basis we cannot conclude with any certainty that sulfinpyrazone is reducing the number of patients dying following heart attacks.

3.3.2 Discussion

We have used the representation theorem for second-order exchangeable quantities to model the heart attack data via Bayes linear updating. However, with binary (or binomial) variables, we know the relationship between the mean and the variance. The variance of a Bernoulli variable with mean p is $p(1-p)$. We can set up our prior to respect this relationship, as indeed we have in the above analysis, but the adjusted means and variances will not satisfy this relationship. The adjusted expectation of X_{ij} is given by

$$\mathrm{E}_1(X_{ij}; \boldsymbol{x}) = p_i^{'},$$

where p_i' is the adjusted expectation for M_i. The adjusted variance of X_{ij} is then

$$\begin{aligned} \mathrm{Var}_1(X_{ij}; \boldsymbol{x}) &= v_i^{'} - v_i + p_i(1-p_i) \\ &\neq p_i^{'}(1-p_i^{'}) \qquad \text{(in general)}, \end{aligned}$$

where $v_i^{'}$ is the adjusted variance of M_i. One possible way to overcome this would be to calculate the adjusted means and variances for the M_i's and calculate the variances for the unobserved X_{ij}'s using $p_i'(1-p_i')$, where p_i' is the adjusted expectation for M_i. However this can lead to the case where posterior variances are larger than prior variances.

While this is possible in the beta-binomial case it only tends to happen when most of the beta density is close to a boundary. However, the prior expectation of a posterior variance cannot be greater than the prior variance. We know that, for p as above and binomially distributed $Y \mid p$, a priori,

$$\mathrm{Var}(p) = \mathrm{E}_Y[\mathrm{Var}(p \mid Y)] + \mathrm{Var}_Y[\mathrm{E}(p \mid Y)],$$

so that the prior expectation of the posterior variance, that is the expectation over the Y distribution of the conditional variance of p given Y, is

$$\mathrm{E}_Y[\mathrm{Var}(p \mid Y)] = \mathrm{Var}(p) - \mathrm{Var}_Y[\mathrm{E}(p \mid Y)] \leq \mathrm{Var}(p).$$

Another issue is that, in this case where we know the mean-variance relationship, the Bayes linear updating is (arguably) not using the information in the most efficient way. Bayes linear updates have the property that our update of the mean is unaffected by anything which we learn about the variance. This is not an obviously reasonable property with Bernoulli or binomial variables where, once we change our expectation for the mean, we clearly change what we think about the variance and therefore change what we think about how we adjusted the mean.

Finally, the Bayes linear method is (as the name implies) a linear fitting procedure. This works very well on quantities defined on $(-\infty, \infty)$ but arguably less well on quantities such as probabilities as we have here. It would seem sensible to consider a suitable transformation of such quantities.

3.4 Bayes linear kinematics

3.4.1 Probability kinematics

Probability kinematics (Jeffrey, 1965) is a method for updating probabilities of events when beliefs over elements in a partition change in some way. Let us suppose that we have a partition $A = (A_1, A_2, \ldots, A_n)$ and that the A_i's have probabilities $\Pr_0(A_i) = p_i$ with $\sum_{i=1}^{n} p_i = 1$.

Now, suppose that we receive some information which causes us to update the probabilities of these events to $\Pr_1(A_1), \ldots, \Pr_1(A_n)$. We can impose the condition that, for any future event B,

$$\Pr_0(B \mid A_i) = \Pr_1(B \mid A_i), \quad \forall i. \tag{3.5}$$

The 'new' marginal probability of B is found by probability kinematics on $\Pr_1(A_1), \ldots, \Pr_1(A_n)$. It is

$$\Pr_1(B) = \sum_{i=1}^{n} \Pr_0(B \mid A_i) \Pr_1(A_i).$$

Successive probability kinematics are not necessarily commutative, however. A great deal of work has been carried out to determine conditions for commutativity of probability kinematics. See, for example, Field (1978), Diaconis & Zabell (1982) and Doring (1999).

To understand this lack of commutativity consider a simple case where we have two unknowns (partitions) A, B, each of which can take only two values, 0,1. The initial joint probability distribution is as follows.

		B	
		0	1
A	0	p_{00}	p_{01}
	1	p_{10}	p_{11}

Let A_0 be the event that "$A = 0$" etc. The conditional probabilities for $B \mid A$ are

$$\Pr(B_0 \mid A_0) = \frac{p_{00}}{p_{00} + p_{01}}, \quad \Pr(B_1 \mid A_0) = \frac{p_{01}}{p_{00} + p_{01}},$$
$$\Pr(B_0 \mid A_1) = \frac{p_{10}}{p_{10} + p_{11}}, \quad \Pr(B_1 \mid A_1) = \frac{p_{11}}{p_{10} + p_{11}}.$$

Now suppose we gain information which causes us to change our marginal probabilities

for A to $\Pr(A_0) = q_0$, $\Pr(A_1) = q_1$. Hence, by probability kinematics, the joint distribution would become as in Table 3.1.

		B	
		0	1
A	0	$q_0 p_{00}/(p_{00}+p_{01})$	$q_0 p_{01}/(p_{00}+p_{01})$
	1	$q_1 p_{10}/(p_{10}+p_{11})$	$q_1 p_{11}/(p_{10}+p_{11})$

Table 3.1: Joint distribution of A and B

Now the conditional probabilities for $A \mid B$ are

$$\begin{aligned}
\Pr(A_0 \mid B_0) &= \left(\frac{q_0 p_{00}}{p_{00}+p_{01}}\right) \Big/ \left\{\left(\frac{q_0 p_{00}}{p_{00}+p_{01}}\right) + \left(\frac{q_1 p_{10}}{p_{10}+p_{11}}\right)\right\}, \\
\Pr(A_1 \mid B_0) &= \left(\frac{q_1 p_{10}}{p_{10}+p_{11}}\right) \Big/ \left\{\left(\frac{q_0 p_{00}}{p_{00}+p_{01}}\right) + \left(\frac{q_1 p_{10}}{p_{10}+p_{11}}\right)\right\}, \\
\Pr(A_0 \mid B_1) &= \left(\frac{q_0 p_{01}}{p_{00}+p_{01}}\right) \Big/ \left\{\left(\frac{q_0 p_{00}}{p_{00}+p_{01}}\right) + \left(\frac{q_1 p_{10}}{p_{10}+p_{11}}\right)\right\}, \\
\Pr(A_1 \mid B_1) &= \left(\frac{q_1 p_{11}}{p_{10}+p_{11}}\right) \Big/ \left\{\left(\frac{q_0 p_{00}}{p_{00}+p_{01}}\right) + \left(\frac{q_1 p_{10}}{p_{10}+p_{11}}\right)\right\}.
\end{aligned}$$

Suppose we gain information which causes us to change our marginal probabilities for B to $\Pr(B_0) = r_0$, $\Pr(B_1) = r_1$. Hence, by probability kinematics, $\Pr(A_0, B_0)$ would become

$$\begin{aligned}
\Pr(A_0, B_0) &= \Pr(B_0)\Pr(A_0 \mid B_0) \\
&= \left(\frac{r_0 q_0 p_{00}}{p_{00}+p_{01}}\right) \Big/ \left\{\left(\frac{q_0 p_{00}}{p_{00}+p_{01}}\right) + \left(\frac{q_1 p_{10}}{p_{10}+p_{11}}\right)\right\} \\
&= \frac{r_0 q_0 p_{00}}{q_0 p_{00} + q_1 p_{10}\left(\dfrac{p_{00}+p_{01}}{p_{10}+p_{11}}\right)}.
\end{aligned}$$

If we had received the information in the opposite order (about B then A) we would have obtained

$$\Pr(A_0, B_0) = \frac{r_0 q_0 p_{00}}{r_0 p_{00} + r_1 p_{01}\left(\dfrac{p_{00}+p_{10}}{p_{01}+p_{11}}\right)}.$$

Clearly the updates are not commutative.

Papathomas & O'Hagan (2005) utilise probability kinematics (called in their terminology Jeffrey's conditionalization) to update beliefs for binary variables when information received is of varying quality. The dependence between the variables is represented by a threshold copula and simulation methods are used to find posterior quantities.

3.4.2 Bayes linear kinematics

Bayes linear kinematics is the kinematic form of a Bayes linear analysis in which the effects of changes in belief about some quantities, rather than actual observations on them, are propagated through to others within a Bayes linear structure. It was developed in Goldstein & Shaw (2004).

Define the full second-order prior specification for some vector random quantity $\boldsymbol{Q}$ to be

$$S_0(\boldsymbol{Q}) = [\mathrm{E}_0(\boldsymbol{Q}), \mathrm{Var}_0(\boldsymbol{Q})],$$

where $\mathrm{E}_0(\boldsymbol{Q})$ is a vector of prior expectations and $\mathrm{Var}_0(\boldsymbol{Q})$ is a prior variance matrix.

Suppose that, rather than directly observing $\boldsymbol{\alpha}$ in Section 3.2, information $I_{\boldsymbol{\alpha}}$ is received which causes our beliefs about $\boldsymbol{\alpha}$ to be updated to $S_1(\boldsymbol{\alpha}) = [\mathrm{E}_1(\boldsymbol{\alpha}), \mathrm{Var}_1(\boldsymbol{\alpha})]$ rather than $S_0(\boldsymbol{\alpha})$.

Then the specification $S_1(\boldsymbol{A})$, for $\boldsymbol{A} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$, is a Bayes linear kinematic update (Goldstein & Shaw, 2004) if it satisfies

$$\mathrm{E}_0(\boldsymbol{\beta}; \boldsymbol{\alpha}) = \mathrm{E}_1(\boldsymbol{\beta}; \boldsymbol{\alpha}), \quad \mathrm{Var}_0(\boldsymbol{\beta}; \boldsymbol{\alpha}) = \mathrm{Var}_1(\boldsymbol{\beta}; \boldsymbol{\alpha}),$$

where $\mathrm{E}_i(\boldsymbol{\beta}; \boldsymbol{\alpha})$ and $\mathrm{Var}_i(\boldsymbol{\beta}; \boldsymbol{\alpha})$ are the Bayes linear adjusted expectation and variance of $\boldsymbol{\beta}$ by $\boldsymbol{\alpha}$ using $S_i(\boldsymbol{A})$. These are the Bayes linear equivalents of Equation 3.5. This yields the Bayes linear kinematic updating equations

$$\mathrm{E}_1(\boldsymbol{A}) = \mathrm{E}_0(\boldsymbol{A}) + \mathrm{Cov}_0(\boldsymbol{A}, \boldsymbol{\alpha})\mathrm{Var}_0^{-1}(\boldsymbol{\alpha})[\mathrm{E}_1(\boldsymbol{\alpha}) - \mathrm{E}_0(\boldsymbol{\alpha})], \tag{3.6}$$

$$\mathrm{Var}_1(\boldsymbol{A}) = \mathrm{Var}_0(\boldsymbol{A}; \boldsymbol{\alpha}) + \mathrm{Cov}_0(\boldsymbol{A}, \boldsymbol{\alpha})\mathrm{Var}_0^{-1}(\boldsymbol{\alpha})\mathrm{Var}_1(\boldsymbol{\alpha})\mathrm{Var}_0^{-1}(\boldsymbol{\alpha})\mathrm{Cov}_0(\boldsymbol{\alpha}, \boldsymbol{A}) \tag{3.7}$$

This is also true if $\boldsymbol{A}$ is replaced by $\boldsymbol{\beta}$ in the above equations. Taking the case of $\boldsymbol{\beta}$ if we observe $\boldsymbol{\alpha}$ directly we return to the usual Bayes linear update.

Let us suppose that we wish to make multiple updates. If we initially observe information $I_{\boldsymbol{\alpha}}$ this updates our beliefs about $\boldsymbol{A}$ to $S_1(\boldsymbol{A}; I_{\boldsymbol{\alpha}})$ using Equations 3.6 and 3.7 as before. If we then observe information $I_{\boldsymbol{\beta}}$ which updates our beliefs about $\boldsymbol{\beta}$ to $S_2(\boldsymbol{\beta}; I_{\boldsymbol{\alpha}}, I_{\boldsymbol{\beta}})$ we can use Bayes linear kinematics a second time to obtain $S_2(\boldsymbol{A}; I_{\boldsymbol{\alpha}}, I_{\boldsymbol{\beta}})$.

Now suppose that we observe the 2 pieces of information in the opposite order. Thus initially we observe $I_{\boldsymbol{\beta}}$ which updates our beliefs about $\boldsymbol{\beta}$ to $S_1(\boldsymbol{\beta}; I_{\boldsymbol{\beta}})$ and we perform Bayes linear kinematics to obtain $S_1(\boldsymbol{A}; I_{\boldsymbol{\beta}})$. We then observe $I_{\boldsymbol{\alpha}}$ which updates our beliefs over $\boldsymbol{\alpha}$ to $S_2(\boldsymbol{\alpha}; I_{\boldsymbol{\beta}}, I_{\boldsymbol{\alpha}})$ and use Equations 3.6 and 3.7 to obtain $S_2(\boldsymbol{A}; I_{\boldsymbol{\beta}}, I_{\boldsymbol{\alpha}})$.

We wish to know when these two updates are commutative, i.e., when $S_2(\boldsymbol{A}; I_{\boldsymbol{\alpha}}, I_{\boldsymbol{\beta}}) = S_2(\boldsymbol{A}; I_{\boldsymbol{\beta}}, I_{\boldsymbol{\alpha}})$. Goldstein & Shaw (2004) give necessay and sufficient conditions for a

unique, commutative, Bayes linear kinematic update.

They show, in Theorem 5, that there is a unique commutative solution if and only if any of the following hold.

(i) $\lambda_{max} < 1$, where λ_{max} is the largest eigenvalue of $\mathrm{Var}_1^{-1}(\boldsymbol{\alpha};\boldsymbol{\beta})\mathrm{Var}_1(\mathrm{E}_1(\boldsymbol{\alpha};\boldsymbol{\beta});\boldsymbol{\alpha})$,

(ii) $\mathrm{Var}_1^{-1}(\boldsymbol{\alpha};I_{\boldsymbol{\alpha}}) + \mathrm{Var}_1^{-1}(\boldsymbol{\alpha};I_{\boldsymbol{\beta}}) - \mathrm{Var}_0^{-1}(\boldsymbol{\alpha})$ is positive definite,

(iii) $\mathrm{Var}_1^{-1}(\boldsymbol{\beta};I_{\boldsymbol{\alpha}}) + \mathrm{Var}_1^{-1}(\boldsymbol{\beta};I_{\boldsymbol{\beta}}) - \mathrm{Var}_0^{-1}(\boldsymbol{\beta})$ is positive definite.

From this a sufficiency condition is derived. Consider (ii) above. Clearly if $\mathrm{Var}_1^{-1}(\boldsymbol{\alpha};I_{\boldsymbol{\alpha}}) - \mathrm{Var}_0^{-1}(\boldsymbol{\alpha})$ is positive definite then the condition holds. This will be positive definite if $\mathrm{Var}_0(\boldsymbol{\alpha}) - \mathrm{Var}_1(\boldsymbol{\alpha};I_{\boldsymbol{\alpha}})$ is positive definite. This leads directly to Corollory 3 of Goldstein & Shaw (2004), which says if

$$\mathrm{Var}_1(\boldsymbol{\alpha};I_{\boldsymbol{\alpha}}) < \mathrm{Var}_0(\boldsymbol{\alpha}) \quad \text{or} \quad \mathrm{Var}_1(\boldsymbol{\beta};I_{\boldsymbol{\beta}}) < \mathrm{Var}_0(\boldsymbol{\beta}), \tag{3.8}$$

then there is a unique commutative Bayes linear kinematic update. Thus, if recieving information causes the variance of either or both quantities to reduce then there will be a commutative update. Thus if we find a model in which variances always reduce on observation of data then we will always be able to use Bayes linear kinematics to provide a commutative solution.

This would allow us to provide general models in which commutativity does not have to be considered and which are always applicable in the analysis of related quantities in two dimensions.

When this unique solution exists it is given by

$$\begin{aligned}\mathrm{E}_{(2)}(\boldsymbol{A}) = \mathrm{Var}_{(2)}(\boldsymbol{A})\{&\mathrm{Var}_1^{-1}(\boldsymbol{A};I_{\boldsymbol{\alpha}})\mathrm{E}_1(\boldsymbol{A};I_{\boldsymbol{\alpha}}) \\ &+ \mathrm{Var}_1^{-1}(\boldsymbol{A};I_{\boldsymbol{\beta}})\mathrm{E}_1(\boldsymbol{A};I_{\boldsymbol{\beta}}) - \mathrm{Var}_0^{-1}(\boldsymbol{A})\mathrm{E}_0(\boldsymbol{A})\},\end{aligned} \tag{3.9}$$

and

$$\mathrm{Var}_{(2)}(\boldsymbol{A}) = \{\mathrm{Var}_1^{-1}(\boldsymbol{A};I_{\boldsymbol{\alpha}}) + \mathrm{Var}_1^{-1}(\boldsymbol{A};I_{\boldsymbol{\beta}}) - \mathrm{Var}_0^{-1}(\boldsymbol{A})\}^{-1}, \tag{3.10}$$

where $\boldsymbol{A}$ can be replaced by $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$. The above solution is clearly commutative as swapping the updates in the equations would not alter the solution. We shall consider the case of more than two related quantities in Section 4.3.

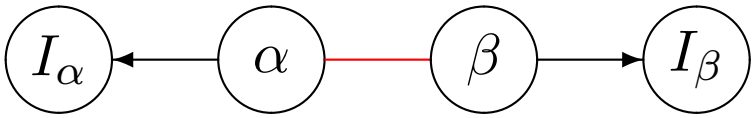

Figure 3.1: Bayes linear Bayes graphical model in 2 dimensions

3.5 Bayes linear Bayes structures

So far all of the models we have considered have been either completely full Bayesian or Bayes linear. It is also possible to combine these two approaches to obtain a structure in which some unknowns have full Bayesian relationships and some have Bayes linear relationships. Following Goldstein & Shaw (2004), we call these Bayes linear Bayes structures.

Suppose that we have unknowns $\boldsymbol{\alpha}, \boldsymbol{\beta}, I_{\boldsymbol{\alpha}}, I_{\boldsymbol{\beta}}$ as in the previous section and give a Bayes linear belief structure to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. We could then give a full Bayesian probability specification to the pairs $(\boldsymbol{\alpha}, I_{\boldsymbol{\alpha}})$ and $(\boldsymbol{\beta}, I_{\boldsymbol{\beta}})$.

We impose the condition that given $\boldsymbol{\alpha}$, $I_{\boldsymbol{\alpha}}$ is conditionally independent of everything in $(\boldsymbol{\beta}, I_{\boldsymbol{\beta}})$ and given $\boldsymbol{\beta}$, $I_{\boldsymbol{\beta}}$ is conditionally independent of everything in $(\boldsymbol{\alpha}, I_{\boldsymbol{\alpha}})$. These three conditions define a Bayes linear Bayes structure.

We can represent situations satisfying the above relationships using a Bayes linear Bayes graphical model. This is a combination of Bayesian graphical models (Lauritzen, 1996) and Bayes linear graphical models (Goldstein & Wilkinson, 2000). In such a model unknowns are represented using nodes and relationships between them using arcs (or edges).

Full Bayesian relationships take the form of black arcs and Bayes linear relationships red arcs. In the case of full Bayesian relationships the arc will be directed if the distribution of one of the quantities is conditional on the other quantity. If the distribution of X_1 were conditional on X_2 then there would be an arrow on the arc between these two variables pointing from X_1 to X_2.

The structure above has the Bayes linear Bayes graphical representation as in Figure 3.1.

We can represent more complex structures using Bayes linear Bayes models than that

given above. Suppose now we have a set of unknowns given by

$$B = \{\boldsymbol{Y}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_s, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_s\}.$$

Here $D_1, \ldots, D_s$ are quantities which shall be observed and these will directly update our beliefs about $X_1, \ldots, X_s$.

The conditions for a Bayes linear Bayes structure are now;

- The collection of quantities $(\boldsymbol{Y}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_s)$ is given a Bayes linear belief structure.
- A full Bayesian probability specification is given to each $(\boldsymbol{X}_i, \boldsymbol{D}_i)$ for $i = 1, \ldots, s$.
- Each $\boldsymbol{D}_i$ is conditionally independent of $B \backslash \{\boldsymbol{X}_i, \boldsymbol{D}_i\}$ given $\boldsymbol{X}_i$.

In such a situation, for $s = 3$, the Bayes linear Bayes graphical model is as in Figure 3.2.

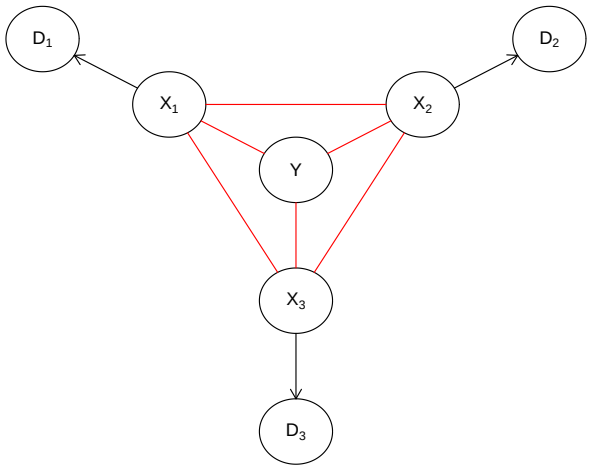

Figure 3.2: A second Bayes linear Bayes graphical model

3.6 Bayes linear kinematics for counts

Consider the two parameter problem from Chapter 2. We have counts X_1 and X_2 such that either

$$X_i \mid \theta_i \sim \text{bin}(n_i, \theta_i), \quad \text{or} \quad X_i \mid \theta_i \sim \text{Po}(\theta_i),$$

for unknown parameters θ_i, $i = 1, 2$. Conditional on the values of θ_1, θ_2, our counts X_1, X_2 are independent. Each θ_i is given the conjugate prior distribution so that

$$\theta_i \sim \text{beta}(a_i, b_i) \quad \text{or} \quad \theta_i \sim \text{gamma}(a_i, b_i),$$

for binomial and Poisson X_i respectively. We shall embed these fully Bayesian updates within a Bayes linear Bayes structure and utilise Bayes linear kinematics to solve this problem.

Observation of x_i successes in group i leads to a conjugate fully Bayesian update within this group. The resulting posterior distributions in the two models are then

$$\theta_i \mid x_i \sim \text{beta}(a_i + x_i, b_i + n_i - x_i) \quad \text{or} \quad \theta_i \mid x_i \sim \text{gamma}(a_i + x_i, b_i + 1),$$

respectively. Thus the prior mean and variance of θ_i in the beta-binomial model are

$$\text{E}_0(\theta_i) = \frac{a_i}{a_i + b_i}, \quad \text{Var}_0(\theta_i) = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)}, \tag{3.11}$$

and the posterior mean and variance are

$$\text{E}_1(\theta_i) = \frac{a_i + x_i}{a_i + b_i + n_i}, \quad \text{Var}_1(\theta_i) = \frac{(a_i + x_i)(b_i + n_i - x_i)}{(a_i + b_i + n_i)^2(a_i + b_i + n_i + 1)}. \tag{3.12}$$

In the gamma-Poisson case the prior mean and variance of θ_i are

$$\text{E}_0(\theta_i) = \frac{a_i}{b_i}, \quad \text{Var}_0(\theta_i) = \frac{a_i}{b_i^2}, \tag{3.13}$$

and, having observed x_i successes, the posterior mean and variance are

$$\text{E}_1(\theta_i) = \frac{a_i + x_i}{b_i + 1}, \quad \text{Var}_1(\theta_i) = \frac{a_i + x_i}{(b_i + 1)^2}. \tag{3.14}$$

In both cases it would be possible to proceed by linking the parameters θ_1, θ_2 in a Bayes linear structure and propagating the within-group updates through to the other group directly using Bayes linear kinematics. Indeed, this is exactly what Goldstein & Shaw (2004) do.

However, we believe that this is not the most effective way to perform such an analysis. The quantity in the Bayes linear structure, about which we recieve information, need not be θ_i itself. It could be some function of θ_i or even something more loosely associated with θ_i.

We propose, rather than choosing to learn about θ_i directly, we transform θ_i to a new quantity η_i on $(-\infty, \infty)$ and embed this into the Bayes linear structure. We believe that this will lead to more effective Bayes linear updates and so is the appropriate way to proceed in this situation. We now discuss the reasons for proposing such transformations.

3.7 The use of transformations

As we mentioned in the previous section, it would be possible to proceed by linking the parameters θ_1, θ_2 in a Bayes linear structure. However there are advantages in transforming the parameters first. The transformed parameters η_1, η_2 are then linked in a Bayes linear structure. The reasons for using the transformation are as follows. Firstly, the range of θ_i is bounded to $0 < \theta_i < 1$ in the binomial case and $0 < \theta_i < \infty$ in the Poisson case. The combination of linear updates with bounded parameter spaces seems undesirable both in terms of first and second moments. If information leads to adjustment of the expectation for a quantity towards a boundary, it seems clear that this adjustment should not continue to be linear as the boundary is approached. It is to be expected that variances will be affected by the proximity of a boundary and beliefs, when the mean is close to a boundary, will no longer be symmetric in the sense that deviations from the mean in either direction would be regarded in the same way. Similarly there are difficulties with covariances in bounded spaces where the tendency would be to imagine rather nonlinear relationships between unknowns close to boundaries. So it is desirable to transform the parameters onto unbounded spaces.

Secondly it is possible for the variances of the untransformed parameters θ_i to increase when data are observed. For example, in the beta-binomial case above, when $a_i = 7$, $b_i = 1$, $n_i = 4$ and $x_i = 2$. In the gamma-Poisson case the posterior variance can be greater than the prior variance if x_i is sufficiently large. While Goldstein & Shaw (2004) (Theorem 5) give conditions for the existence of unique Bayes linear kinematic updates which allow some such variance increase, the transformations have the effect of making reductions in variance of the transformed parameters occur when observations are made, at least in most circumstances, and therefore allow use of the simpler sufficient condition given in Corollary 3 of Goldstein & Shaw (2004).

Bayes linear kinematics, without transformation, gives a rule for adjusting beliefs about

θ_1, θ_2 by Bayes linear updates. Similarly Bayes linear kinematics, with the transformation, gives a Bayes linear rule for updating beliefs about η_1, η_2, where there is a 1 - 1 relationship between η_i and θ_i. Any further use of conjugate Bayesian updating of beliefs about θ_j, given observation of X_j, after already adjusting by observation of X_i, relies on the idea that θ_j still has a distribution of the required conjugate form, whether or not a transformation is used. Similarly evaluating predictive distributions for new observations or credible intervals for θ_1, θ_2 depends on such an idea. Additionally, when a transformation is used, this preserved conjugate form is required in order to convert back from the adjusted moments of η_j to the new distribution for θ_j.

Clearly, if adjustments were only ever made in one direction, eg. of beliefs about θ_j by observing X_i, and this was never reversed to adjust beliefs about θ_i by observing X_j, then it could simply be declared that the conditional distribution was the required conjugate distribution. Such one-way belief adjustment might be appropriate, for example, in a time-series forecasting context, as in West *et al.* (1985). Even in this case, however, we would be saying that the conjugate distribution holds both when we make the update and for forecasts to time $t + k$.

When commutativity, in the strong sense that conjugate updates of the marginal distributions of θ_1, θ_2 are always appropriate, is required then this might be regarded as a pragmatic approximation which does not correspond exactly to a full Bayesian conditioning analysis. With no transformation, this assumption is made directly on the distributions of θ_1, θ_2 under Bayes linear kinematic updates. With transformation, the assumption applies to the corresponding distributions of η_1, η_2, in the same way.

3.7.1 The transformed approach

Having decided on the use of transformations of the binomial and Poisson parameters we represent them using the function $g()$, where

$$\eta_i = g(\theta_i),$$

for $i = 1, 2$. The transformation $g()$ is such that for either $0 < \theta_i < 1$ in the beta-binomial case or $\theta_i > 0$ in the gamma-Poisson model then $\eta_i \in (-\infty, \infty)$. We then link η_1, η_2, rather than θ_1, θ_2, in a Bayes linear structure.

In order to perform Bayes linear kinematics we shall need the prior means and variances

of η_1, η_2. In the beta-binomial model they are

$$\begin{aligned} \mathrm{E}_0(\eta_i) &= \int_0^1 g(\theta_i) f_{0i}(\theta_i) d\theta_i, \\ \mathrm{Var}_0(\eta_i) &= \int_0^1 [g(\theta_i)]^2 f_{0i}(\theta_i) d\theta_i - [\mathrm{E}_0(\theta_i)]^2, \end{aligned}$$

where $f_{0i}(\theta_i)$ is the prior beta density for θ_i. After the conjugate updates the expressions for the mean and variance, $\mathrm{E}_1(\eta_i)$ and $\mathrm{Var}_1(\eta_i)$, remain of the same form but with a_i and b_i replaced by $a_i + x_i$ and $b_i + n_i - x_i$ respectively.

In the gamma-Poisson model the prior means and variances of each η_i are

$$\begin{aligned} \mathrm{E}_0(\eta_i) &= \int_0^\infty g(\theta_i) f_{0i}(\theta_i) d\theta_i, \\ \mathrm{Var}_0(\eta_i) &= \int_0^\infty [g(\theta_i)]^2 f_{0i}(\theta_i) d\theta_i - [\mathrm{E}_0(\theta_i)]^2, \end{aligned}$$

where now $f_{0i}(\theta_i)$ is the gamma density associated with θ_i. Having observed x_i, $\mathrm{E}_1(\eta_i)$ and $\mathrm{Var}_1(\eta_i)$ are obtained in the same form but using the new parameter values $a_i + x_i$ and $b_i + 1$.

We can propagate these changes in belief through to the other group via the Bayes linear kinematic updating equations.

$$\mathrm{E}_1(\eta_j; x_i) = \mathrm{E}_0(\eta_j) + \frac{\mathrm{Cov}_0(\eta_i, \eta_j)}{\mathrm{Var}_0(\eta_i)} \left[\mathrm{E}_1(\eta_i) - \mathrm{E}_0(\eta_i)\right], \tag{3.15}$$

$$\mathrm{Var}_1(\eta_j; x_i) = \mathrm{Var}_0(\eta_j) - \frac{\mathrm{Cov}_0(\eta_i, \eta_j)^2}{\mathrm{Var}_0(\eta_i)} \left[1 - \frac{\mathrm{Var}_1(\eta_i)}{\mathrm{Var}_0(\eta_i)}\right], \tag{3.16}$$

for $i \neq j$. We wish to know when a unique, commutative Bayes linear kinematic solution exists. A sufficient condition for uniqueness, using Equation 3.8 is

$$\mathrm{Var}_1(\eta_i) < \mathrm{Var}_0(\eta_i) \tag{3.17}$$

for $i = 1$ or 2 or both. If this condition holds then the Bayes linear kinematic adjusted expectation and variance of η_i are given by

$$\begin{aligned} \mathrm{E}_{(2)}(\eta_i; x_i, x_j) = \mathrm{Var}_{(2)}(\eta_i; x_i, x_j)[&\mathrm{Var}_1^{-1}(\eta_i)\mathrm{E}_1(\eta_i) \\ &+ \mathrm{Var}_1^{-1}(\eta_i; x_j)\mathrm{E}_1(\eta_i; x_j) - \mathrm{Var}_0^{-1}(\eta_i)\mathrm{E}_0(\eta_i)], \end{aligned} \tag{3.18}$$

and

$$\mathrm{Var}_{(2)}(\eta_i; x_i, x_j) = \left[\mathrm{Var}_1^{-1}(\eta_i) + \mathrm{Var}_1^{-1}(\eta_i; x_j) - \mathrm{Var}_0^{-1}(\eta_i)\right]^{-1}. \tag{3.19}$$

In the notation above $\mathrm{E}_{(2)}(\eta_i; x_i, x_j)$ and $\mathrm{Var}_{(2)}(\eta_i; x_i, x_j)$ represent the Bayes linear kinematic commutative expectation and variance (Equations 3.9 and 3.10) of η_i having made 2 observations (given in brackets in the subscript). The quantities after the semi-colon indicate that these are the adjusted expectation and variance having observed x_i and x_j.

3.7.2 Predictive distributions

Suppose now we have $p \geq 2$ groups $X_1, \ldots, X_p$ where each is a binomial or Poisson count. Imagine we have performed a Bayes linear kinematic update of the form in the previous section on $X_1, \ldots, X_p$ and obtained adjusted expectations and variances for $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$.

Now suppose we imagine updating by $X_1, \ldots, X_{p-1}$. The η_p, X_p structure has to be such that we would get the "correct" update by $X_1, \ldots, X_p$. This means that it has to be the conjugate beta-binomial or gamma-Poisson structure.

Therefore, to be consistent with potential future updates, predictive distributions are calculated on the basis of the same structure.

3.8 Logistic transformation

3.8.1 Expectation and variance

Clearly with a beta-binomial setup several transformations are possible, being those used as link functions in generalised linear models. Those commonly chosen are the logit, $\eta_i = \log(\theta_i/[1-\theta_i])$, probit, $\eta_i = \Phi^{-1}(\theta_i)$, and complementary log-log, $\eta_i = \log(-\log(1-\theta_i))$, link functions.

Initially we consider the logistic transformation so that

$$\eta_1 = \log\left(\frac{\theta_1}{1-\theta_1}\right), \quad \eta_2 = \log\left(\frac{\theta_2}{1-\theta_2}\right).$$

We see that $\theta_i \in (0,1) \Rightarrow \eta_i \in (-\infty, \infty)$ and so η_i is unbounded. We can then give η_1 and η_2 a Bayes linear relationship. The prior expectations and variances of θ_1 and θ_2 are given in Equation 3.11. Having made within-group updates, the moments of θ_1 and θ_2 are found from Equation 3.12.

In order to perform Bayes linear kinematics we require the prior and posterior expec-

tations and variances of η_1 and η_2. The prior expectation and variance of η_i are

$$\mathrm{E}_0(\eta_i) = \psi(a_i) - \psi(b_i), \quad \mathrm{Var}_0(\eta_i) = \psi_1(a_i) + \psi_1(b_i),$$

where $\psi(x) = \frac{d}{dx}\log(\Gamma(x))$ is the digamma function and $\psi_1(x) = \frac{d}{dx}\psi(x)$ is the trigamma function.

Proof. The beta function is defined as

$$\beta(a_i, b_i) = \int_0^1 z^{a_i-1}(1-z)^{b_i-1}dz.$$

Differentiating this successively with respect to a_i and b_i gives

$$\begin{aligned}\frac{\partial^{s+t}}{\partial a_i^s \partial b_i^t}\beta(a_i, b_i) &= \int_0^1 (\log z)^s (\log(1-z))^t z^{a_i-1}(1-z)^{b_i-1}dz \\ &= \beta(a_i, b_i)\mathrm{E}_0[(\log\theta_i)^s(\log(1-\theta_i))^t].\end{aligned}$$

The beta function can be expressed as $\beta(a_i, b_i) = \Gamma(a_i)\Gamma(b_i)/\Gamma(a_i+b_i)$. Differentiating this with respect to a_i gives

$$\begin{aligned}\frac{\partial}{\partial a_i}\beta(a_i, b_i) &= \Gamma(b_i)\left\{\frac{\Gamma'(a_i)}{\Gamma(a_i+b_i)} - \frac{\Gamma'(a_i+b_i)\Gamma(a_i)}{\Gamma(a_i+b_i)^2}\right\} \\ &= \frac{\Gamma(a_i)\Gamma(b_i)}{\Gamma(a_i+b_i)}\left\{\frac{\Gamma'(a_i)}{\Gamma(a_i)} - \frac{\Gamma'(a_i+b_i)}{\Gamma(a_i+b_i)}\right\} \\ &= \beta(a_i, b_i)[\psi(a_i) - \psi(a_i+b_i)],\end{aligned}$$

as $\psi(x) = \frac{d}{dx}\log(\Gamma(x)) = \frac{1}{\Gamma(x)}\frac{d}{dx}\Gamma(x)$. We can use a property of the beta function, $\beta(a_i, b_i) = \beta(b_i, a_i)$, to see that $\frac{\partial}{\partial b_i}\beta(a_i, b_i) = \beta(a_i, b_i)[\psi(b_i) - \psi(a_i+b_i)]$. Thus the prior expectation of η_i is

$$\begin{aligned}\mathrm{E}_0(\eta_i) &= \mathrm{E}_0[\log\theta_i] - \mathrm{E}_0[\log(1-\theta_i)] \\ &= \psi(a_i) - \psi(a_i+b_i) - \psi(b_i) + \psi(a_i+b_i) \\ &= \psi(a_i) - \psi(b_i).\end{aligned}$$

In order to find the variance of η_i we shall require

$$\mathrm{E}_0[\eta_i^2] = \mathrm{E}_0\left\{\left[\log\left(\frac{\theta_i}{1-\theta_i}\right)\right]^2\right\} = \mathrm{E}_0[(\log\theta_i)^2] - 2\mathrm{E}_0[\log\theta_i\log(1-\theta_i)] + \mathrm{E}_0[(\log(1-\theta_i))^2].$$

To calculate these quantities we need the three partial second derivatives, which are

$$\begin{aligned}\frac{\partial^2}{\partial a_i^2}\beta(a_i,b_i) &= \frac{\partial}{\partial a_i}\left\{\beta(a_i,b_i)[\psi(a_i)-\psi(a_i+b_i)]\right\}\\ &= \beta(a_i,b_i)\{[\psi(a_i)-\psi(a_i+b_i)]^2+\psi_1(a_i)-\psi_1(a_i+b_i)\}.\end{aligned}$$

Similarly $\frac{\partial^2}{\partial b_i^2}\beta(a_i,b_i) = \beta(a_i,b_i)\{[\psi(b_i)-\psi(a_i+b_i)]^2+\psi_1(b_i)-\psi_1(a_i+b_i)\}$ and the cross derivative is

$$\begin{aligned}\frac{\partial^2}{\partial a_i \partial b_i}\beta(a_i,b_i) &= \frac{\partial}{\partial b_i}\beta(a_i,b_i)[\psi(a_i)-\psi(a_i+b_i)]\\ &= \beta(a_i,b_i)\{[\psi(b_i)-\psi(a_i+b_i)][\psi(a_i)-\psi(a_i+b_i)]-\psi_1(a_i+b_i)\}.\end{aligned}$$

Thus, after cancellation, the required expectation is

$$\mathrm{E}_0[\eta_i^2] = [\psi(a_i)-\psi(b_i)]^2+\psi_1(a_i)+\psi_1(b_i).$$

The prior variance of η_i is therefore

$$\begin{aligned}\mathrm{Var}_0(\eta_i) &= \mathrm{E}_0[\eta_i^2]-\mathrm{E}_0[\eta_i]^2\\ &= \psi_1(a_i)+\psi_1(b_i).\end{aligned}$$

□

The expectation and variance of η_i having observed x_i successes out of n_i trials in group i are of the same form but with new parameter values $A_i = a_i + x_i$ and $B_i = b_i + n_i - x_i$ so that

$$\begin{aligned}\mathrm{E}_1(\eta_1) = \psi(A_1)-\psi(B_1), \quad \mathrm{E}_1(\eta_2) = \psi(A_2)-\psi(B_2),\\ \mathrm{Var}_1(\eta_1) = \psi_1(A_1)+\psi_1(B_1), \quad \mathrm{Var}_1(\eta_2) = \psi_1(A_2)+\psi_1(B_2).\end{aligned}$$

We can propagate these changes in belief through to the other group using Bayes linear kinematic updating Equations, 3.15 and 3.16, to obtain $\mathrm{E}_1(\eta_1; x_2)$, $\mathrm{Var}_1(\eta_1; x_2)$, $\mathrm{E}_1(\eta_2; x_1)$ and $\mathrm{Var}_1(\eta_2; x_1)$.

We must now find out whether a unique, commutative solution exists. The sufficient condition for uniqueness, Equation 3.17, in this case is

$$\psi_1(A_i)+\psi_1(B_i) < \psi_1(a_i)+\psi_1(b_i).$$

As long as we observe at least one Bernoulli trial $A_i > a_i$ or $B_i > b_i$ or both. This means that $\psi_1(A_i) \leq \psi_1(a_i)$ and $\psi_1(B_i) \leq \psi_1(b_i)$ with at least one of them strictly increasing

as the trigamma function is monotonically decreasing on $\mathbb{R}^+$. We can see this property in Figure 3.3. Hence the uniqueness condition always holds when we observe data and so there is always a unique commutative solution when this transformation is taken.

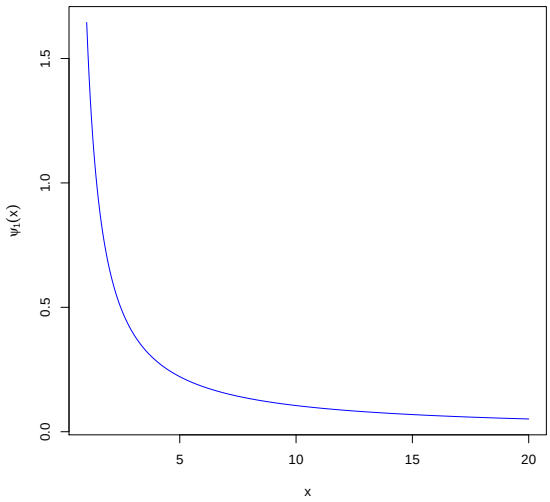

Figure 3.3: The trigamma function

This solution is given in Equations 3.18 and 3.19 and gives us our posterior expectations and variances; $\mathrm{E}_{(2)}(\eta_1; x_1, x_2)$, $\mathrm{E}_{(2)}(\eta_2; x_1, x_2)$, $\mathrm{Var}_{(2)}(\eta_1; x_1, x_2)$ and $\mathrm{Var}_{(2)}(\eta_2; x_1, x_2)$.

If our assumption of beta marginals for θ_1 and θ_2 still holds then we can return to quantities involving θ_1 and θ_2 by solving the following 2 equations numerically for the posterior parameter values a_i^* and b_i^*.

$$\begin{aligned} \mathrm{E}_{(2)}(\eta_i; x_1, x_2) &= \psi(a_i^*) - \psi(b_i^*), \\ \mathrm{Var}_{(2)}(\eta_i; x_1, x_2) &= \psi_1(a_i^*) + \psi_1(b_i^*). \end{aligned}$$

Thus the posterior distributions for θ_1 and θ_2 are $\theta_i; x_1, x_2 \sim \mathrm{beta}(a_i^*, b_i^*)$ and their posterior expectations and variances are

$$\mathrm{E}_{(2)}(\theta_i; x_1, x_2) = \frac{a_i^*}{a_i^* + b_i^*}, \quad \mathrm{Var}_{(2)}(\theta_i; x_1, x_2) = \frac{a_i^* b_i^*}{(a_i^* + b_i^*)^2 (a_i^* + b_i^* + 1)}.$$

3.8.2 Example: Heart attack data

Let us now perform all of the above calculations for the heart attack data. First we must make some prior specifications. We shall use the same values of a_1, a_2, b_1 and b_2 as in the other method that preserved the marginal beta distributions in the prior, the

copula model of Section 2.4. These are

$$a_1 = 2.59, \quad b_1 = 19.55, \quad a_2 = 10.72, \quad b_2 = 80.84,$$

and this results in prior expectations and variances for η_1 and η_2 of $\mathrm{E}_0(\eta_1) = -2.201$, $\mathrm{Var}_0(\eta_1) = 0.5225$, $\mathrm{E}_0(\eta_2) = -2.062$ and $\mathrm{Var}_0(\eta_2) = 0.1102$. We also need to define a covariance between η_1 and η_2. We shall use a correlation of $\rho_0(\eta_1, \eta_2) = 0.3$, and so

$$\begin{aligned} \mathrm{Cov}_0(\eta_1, \eta_2) &= \rho_0(\eta_1, \eta_2)\sqrt{\mathrm{Var}_0(\eta_1)\mathrm{Var}_0(\eta_2)} \\ &= 0.07199. \end{aligned}$$

We are now in a position to update our beliefs. Let us first consider group 1 in which we observed 44 deaths out of 560 patients. Thus we have that

$$\begin{aligned} A_1 &= a_1 + 44 = 46.59 \\ B_1 &= b_1 + 560 - 44 = 535.55 \end{aligned}$$

and the updated expectation and variance of η_1 are $\mathrm{E}_1(\eta_1) = -2.452$, $\mathrm{Var}_1(\eta_1) = 0.02356$. We now perform the Bayes linear kinematic update to obtain the expectation and variance of η_2 given x_1. These turn out to be

$$\mathrm{E}_1(\eta_2; x_1) = -2.096, \quad \mathrm{Var}_1(\eta_2; x_1) = 0.1007.$$

In group 2 we observed 62 patients dying out of 540. Thus we can update a_2 and b_2 to $A_2 = 72.72$ and $B_2 = 558.84$ and the expectation and variance of η_2 become

$$\begin{aligned} \mathrm{E}_1(\eta_2) &= \psi(72.72) - \psi(558.84) \\ &= -2.045, \\ \mathrm{Var}_1(\eta_2) &= \psi_1(72.72) + \psi_1(558.84) \\ &= 0.0156. \end{aligned}$$

When we use these values in the Bayes linear kinematic updating equations we see that

$$\mathrm{E}_1(\eta_1; x_2) = -2.190, \quad \mathrm{Var}_1(\eta_1; x_2) = 0.482.$$

From the values of a_1, b_1, a_2 and b_2 at each stage we can calculate the expectations and variances of θ_1 and θ_2. The expectations are given in Table 3.2. The variances can also be given in tabular form, as in Table 3.3. From the variance table we can see that we reduce our uncertainty about θ_1 and θ_2 at each step.

We can combine all of the expectations and variances of η_1 and η_2 at each stage,

θ_i	$\mathrm{E}_0(\theta_i)$	$\mathrm{E}_1(\theta_i)$	$\mathrm{E}_1(\theta_i; x_j)$	$\frac{x_i}{n_i}$
1	0.1170	0.0800	0.1168	0.0786
2	0.1171	0.1151	0.1132	0.1148

Table 3.2: Expectations of θ_1 and θ_2 at each stage

θ_i	$\mathrm{Var}_0(\theta_i)$	$\mathrm{Var}_1(\theta_i)$	$\mathrm{Var}_1(\theta_i; x_j)$
1	0.00446	0.000126	0.00412
2	0.00111	0.000161	0.00157

Table 3.3: Variances of θ_1 and θ_2 at each stage

using Equations 3.18 and 3.19 to give the unique commutative Bayes linear kinematic expectations and variances. These are $\mathrm{E}_{(2)}(\eta_1; x_1, x_2) = -2.450$, $\mathrm{Var}_{(2)}(\eta_1; x_1, x_2) = 0.02348$, $\mathrm{E}_{(2)}(\eta_2; x_1, x_2) = -2.051$ and $\mathrm{Var}_{(2)}(\eta_2; x_1, x_2) = 0.01543$. We can now convert back to moments involving our success probabilities. To do this we find our values for a_i^* and b_i^*;

$$a_1^* = 46.77, \quad b_1^* = 536.85, \quad a_2^* = 73.64, \quad b_2^* = 569.06.$$

We can then calculate our posterior expectations and variances for θ_1 and θ_2 and these turn out to be

$$\begin{aligned} \mathrm{E}_{(2)}(\theta_1; x_1, x_2) &= 0.08014 \qquad \mathrm{E}_{(2)}(\theta_2; x_1, x_2) = 0.1146 \\ \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) &= 0.000126 \quad \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) = 0.000158. \end{aligned}$$

3.8.3 Mode and log-curvature

As we commented earlier, the quantity about which we learn on receipt of information need not be θ_i. It need not even be a direct function of θ_i as in the previous section. We could simply define a quantity which is updated when data are observed and this update is used as a guide to updating θ_i. This is known as a guide relationship (West *et al.*, 1985).

For example, considering the logistic transformation, we could define the quantity μ_i as

$$\mu_i = \log\left(\frac{\theta_i}{1-\theta_i}\right).$$

Rather than updating beliefs about the mean and variance of μ_i directly we could use a related quantity, η_i, with mean and variance given by the mode of μ_i and the curvature

at the mode of the log density of μ_i. Thus our guide relationship (denoted $\approx$) is

$$\eta_i \approx \mu_i = \log\left(\frac{\theta_i}{1-\theta_i}\right).$$

So, we are proposing quantities η_i associated with θ_i in such a way that the mean of η_i is equal to the mode of $\log(\theta_i/[1-\theta_i])$ and the variance is the curvature of the log-density at this quantity. Observing x_i causes us to change our beliefs about η_i in the way this would imply.

The density of θ_i is

$$f_{\theta_i}(\theta_i) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\theta_i^{a_i-1}(1-\theta_i)^{b_i-1},$$

and, in terms of μ_i, θ_i is

$$\theta_i = \frac{e^{\mu_i}}{1+e^{\mu_i}}.$$

To obtain the density of μ_i, first we must differentiate the density of θ_i to obtain the Jacobian. The derivative is

$$\frac{d\theta_i}{d\mu_i} = \frac{(1+e^{\mu_i})e^{\mu_i} - e^{2\mu_i}}{(1+e^{\mu_i})^2} = \left(\frac{e^{\mu_i}}{1+e^{\mu_i}}\right)\left(\frac{1}{1+e^{\mu_i}}\right) = \theta_i(1-\theta_i)$$

and so $d\theta_i = \theta_i(1-\theta_i)d\mu_i$. Hence the density of μ_i is

$$f_{\mu_i}(\mu_i) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\theta^{a_i}(1-\theta_i)^{b_i}.$$

If we take logs,

$$l_i(\mu_i) = \log\{f_{\mu_i}(\mu_i)\} = (\text{const}) \ + a_i\log(\theta_i) + b_i\log(1-\theta_i).$$

Differentiating this a single time and then setting the derivative equal to zero gives us the mode of μ_i. The derivative is

$$\frac{d}{d\mu_i}[l_i(\mu_i)] = \left(\frac{a_i}{\theta_i} - \frac{b_i}{1-\theta_i}\right)\theta_i(1-\theta_i) = a_i(1-\theta_i) - b_i\theta_i. \tag{3.20}$$

We now set Equation 3.20 equal to zero to find the mode. Let m_i be the mode of μ_i. Then let

$$m_i^* = \frac{e^{m_i}}{1+e^{m_i}}.$$

This gives $a_i(1-m_i^*) = b_i m_i^*$ so $m_i^* = \frac{a_i}{a_i+b_i}$ and the mode of μ_i is

$$m_i = \log\left(\frac{a_i}{b_i}\right).$$

To find the curvature we must differentiate Equation 3.20 a further time. The second derivative is given by

$$\frac{d^2}{d\mu_i^2}[l_i(\mu_i)] = -(a_i+b_i)\theta_i(1-\theta_i).$$

At the mode $\theta_i = a_i/(a_i+b_i)$ so, substituting this into the above equation,

$$\left[\frac{d^2 l_i(\mu_i)}{d\mu_i^2}\right]_{m_i} = -\frac{a_i b_i}{a_i+b_i}.$$

Therefore our required prior variance for η_i is

$$\mathrm{Var}_0(\eta_i) = -\left[\frac{d^2 l_i(\mu_i)}{d\mu_i^2}\right]_{m_i}^{-1} = \frac{1}{a_i} + \frac{1}{b_i}.$$

Hence, our two prior moments can be expressed solely in terms of the prior beta parameters. They are

$$\mathrm{E}_0(\eta_i) = \log\left(\frac{a_i}{b_i}\right), \quad \mathrm{Var}_0(\eta_i) = \frac{1}{a_i} + \frac{1}{b_i}. \tag{3.21}$$

Observing x_1 successes out of n_1 trials in group 1 and x_2 successes out of n_2 trials in group 2 will lead to posterior expectations and variances of

$$\mathrm{E}_1(\eta_i) = \log\left(\frac{A_i}{B_i}\right), \quad \mathrm{Var}_1(\eta_i) = \frac{1}{A_i} + \frac{1}{B_i},$$

where $A_i = a_i + x_i$ and $B_i = b_i + n_i - x_i$. These changes in belief can then be propagated through to the other group via Bayes linear kinematics using Equation 3.15.

The uniqueness condition, Equation 3.17, is satisfied as long as the variance decreases from prior to posterior. Clearly the variance will decrease if we increase either a_i or b_i which shall happen if we observe anything. Therefore our sufficient condition for a unique commutative solution shall always be satisfied and a unique commutative Bayes linear kinematic solution does exist. It is given in Equations 3.18 and 3.19.

Having found the adjusted expectations and variances $\mathrm{E}_{(2)}(\eta_i; x_i, x_j) = \bar{m}_i$ and $\mathrm{Var}_{(2)}(\eta_i; x_i, x_j) = v_i$ for $i = 1, 2$, we wish to convert back to quantities involving θ_1 and θ_2. If we assume the relationship in Equation 3.21 still holds then the posterior

parameter values a_i^* and b_i^* are

$$a_i^* = \frac{1 + e^{\bar{m}_i}}{v_i}, \quad b_i^* = \frac{1 + e^{\bar{m}_i}}{v_i e^{\bar{m}_i}}. \tag{3.22}$$

This gives Bayes linear kinematic adjusted beta distributions for θ_1 and θ_2 of $\theta_i; x_i, x_j \sim$beta(a_i^*, b_i^*). Thus we can find, via the standard formulae for beta random variables, the mean and variance of θ_1 and θ_2.

3.8.4 Example: Heart attack data

We can apply this approach to the heart attack data. Given our prior specifications for a_i and b_i, $i = 1, 2$, the prior moments of η_1 and η_2 are

$$\begin{aligned} \mathrm{E}_0(\eta_1) = -2.021, \quad \mathrm{E}_0(\eta_2) = -2.020, \\ \mathrm{Var}_0(\eta_1) = 0.4373, \quad \mathrm{Var}_0(\eta_2) = 0.1057. \end{aligned}$$

This gives a prior covariance between η_1 and η_2, for a prior correlation of 0.3, of $\mathrm{Cov}_0(\eta_1, \eta_2) = 0.06448$. Having observed $X_1 = 44$ and $X_2 = 62$ we make the fully Bayesian conjugate updates within each group which translates to η_1 and η_2 as

$$\begin{aligned} \mathrm{E}_1(\eta_1) = -2.442, \quad \mathrm{E}_1(\eta_2) = -2.039, \\ \mathrm{Var}_1(\eta_1) = 0.02333, \quad \mathrm{Var}_1(\eta_2) = 0.01554. \end{aligned}$$

We then propagate theses changes using Bayes linear kinematics and calculate the unique commutative Bayes linear kinematic solution. This is

$$\begin{aligned} \mathrm{E}_{(2)}(\eta_1; x_1, x_2) = -2.441, \quad \mathrm{E}_{(2)}(\eta_2; x_1, x_2) = -2.049, \\ \mathrm{Var}_{(2)}(\eta_1; x_1, x_2) = 0.02323, \quad \mathrm{Var}_{(2)}(\eta_2; x_1, x_2) = 0.01533. \end{aligned}$$

We can now convert back to quantities involving θ_1 and θ_2. First we find the posterior parameter values using Equation 3.22. They are $a_1^* = 46.80, a_2^* = 73.64, b_1^* = 537.34$ and $b_2^* = 571.32$. We can use these to find the posterior moments of θ_1 and θ_2. They are

$$\begin{aligned} \mathrm{E}_{(2)}(\theta_1; x_1, x_2) = 0.08012, \quad \mathrm{E}_{(2)}(\theta_2; x_1, x_2) = 0.1142, \\ \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) = 0.0001260, \quad \mathrm{Var}_{(2)}(\theta_2; x_1, x_2) = 0.0001566. \end{aligned}$$

These values are very similar to those found using the mean and variance of η_i directly.

3.9 Complementary log-log transformation

An alternative in the beta-binomial case to the logistic transformation discussed in the previous section is the complementary log-log transformation. When taking this transformation the expectations and variances of the transformed quantities are not straightforward to calculate and so we propose using our guide relationship again. Thus the mean and variance of η_i, $i = 1, 2$, are found as the mode of μ_i and the curvature at the mode of the log-density of μ_i, where

$$\mu_i = g(\theta_i) = \log[-\log(1-\theta_i)]. \tag{3.23}$$

If $\theta_i \sim$beta(a_i, b_i), with density

$$f_{\theta_i}(\theta_i) = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}\theta_i^{a_i-1}(1-\theta_i)^{b_i-1}$$

and $\mu_i = \log[-\log(1-\theta_i)]$ then $\theta_i = 1-\exp[-e^{\mu_i}]$. Differentiating this with respect to μ_i gives

$$\frac{d\theta_i}{d\mu_i} = \exp[-e^{\mu_i}]e^{\mu_i} = \exp[\mu_i - e^{\mu_i}] = -\log(1-\theta_i)(1-\theta_i),$$

and so $d\theta_i = -\log(1-\theta_i)(1-\theta_i)d\mu_i = e^{\mu_i}\exp[e^{\mu_i}]d\mu_i$. Hence the density of μ_i is

$$f_{\mu_i}(\mu_i) = f_{\theta_i}(\theta_i)\frac{d\theta_i}{d\mu_i} = \frac{\Gamma(a_i+b_i)}{\Gamma(a_i)\Gamma(b_i)}e^{\mu_i}\exp[-e^{\mu_i}]\theta_i^{a_i-1}(1-\theta_i)^{b_i-1}.$$

Taking logs gives the log-density of μ_i,

$$l_i(\mu_i) = \log\{f_{\mu_i}(\mu_i)\} = k_i + \mu_i - e^{\mu_i} + (a_i-1)\log(\theta_i) + (b_i-1)\log(1-\theta_i)$$

where k_i is a constant. To find the mode, m_i, of μ_i, we differentiate the log-density and then set the derivative equal to zero. Thus the mode is the solution m_i satisfying

$$\left(\frac{dl_i(\mu_i)}{d\mu_i}\right)_{m_i} = 1 - e^{m_i} + \left[\frac{(a_i-1)}{\theta_{m,i}} - \frac{(b_i-1)}{1-\theta_{m,i}}\right]e^{m_i}\exp[-e^{m_i}] = 0, \tag{3.24}$$

where $\theta_{m,i} = 1-\exp[-e^{m_i}]$, and is found numerically, for example by Newton's method. The second derivative is

$$\begin{aligned}\frac{d^2l_i(\mu_i)}{d\mu_i^2} &= -e^{\mu_i} - \left[\frac{(a_i-1)}{\theta_i^2} + \frac{(b_i-1)}{(1-\theta_i)^2}\right]e^{2\mu_i}\exp[-2e^{\mu_i}] \\ &\quad + \left[\frac{(a_i-1)}{\theta_i} - \frac{(b_i-1)}{1-\theta_i}\right]e^{\mu_i}(1-e^{\mu_i})\exp[-e^{\mu_i}].\end{aligned} \tag{3.25}$$

The mean and variance of η_i can then be found as

$$\mathrm{E}_0(\eta_i) = m_i, \quad \mathrm{Var}_0(\eta_i) = -\left[\frac{d^2 l_i(\mu_i)}{d\mu_i^2}\right]^{-1}_{m_i}.$$

Having made the conjugate updates, the same procedure can be applied but using $A_i = a_i + x_i$ and $B_i = b_i + n_i - x_i$ in place of a_i and b_i in the density and subsequent derivatives. Defining a Bayes linear structure for η_1, η_2, allows the updates to be propagated to η_j, $j \neq i$ via Equation 3.15.

From Equation 3.17 there is a unique commutative solution to the problem using Bayes linear kinematics if

$$\mathrm{Var}_1(\eta_i) < \mathrm{Var}_0(\eta_i)$$

for $i = 1$ or 2 or both. An analytic proof that this condition always holds is not yet available. However this has been investigated numerically. It is only necessary to consider the effect of a single observation $x_i = 1$ with $n_i = 1$. This is because this is equivalent to the observation $x_i = 0$ with $n_i = 1$ with a_i and b_i exchanged and any observation with larger n_i has the cumulative effect of a sequence of observations with $n_i = 1$. The increase in the precision of η_i given an observation $x_i = 1$ with $n_i = 1$ was investigated over a rectangular grid of values of (a_i, b_i) with $-1 \leq \log(a_i) \leq 12$ and $-1 \leq \log(b_i) \leq 12$ in steps of 0.1 and every value was positive.

Following the numerical investigation we have empirical evidence suggesting that a unique commutative Bayes linear kinematic solution will exist, at least over a very large range of a_i and b_i. It is given by Equations 3.18 and 3.19. Note that, once an adjusted mean and precision for η_i are found, Equations 3.24 and 3.25 provide simultaneous linear equations in a_i^* and b_i^*, the new values of a_i and b_i, which are easily solved.

3.9.1 Example: Heart attack data

Let us now apply this complementary log-log model to the heart attack data. Taking the usual values for the prior beta parameters gives prior moments for η_1 and η_2 of

$$\begin{aligned} \mathrm{E}_0(\eta_1) &= -2.060, \quad \mathrm{E}_0(\eta_2) = -2.078, \\ \mathrm{Var}_0(\eta_1) &= 0.3864, \quad \mathrm{Var}_0(\eta_2) = 0.09339. \end{aligned}$$

We wish to specify a prior correlation between η_1 and η_2 of 0.3. This leads to a prior covariance of $\mathrm{Cov}_0(\eta_1, \eta_2) = 0.3 \times \sqrt{0.3864 \times 0.09339} = 0.05699$. Having observed 44 success out of 560 trials in group 1 and 62 successes out of 540 trials in group 2 the

expectations and variances of η_1 and η_2 are updated to

$$\begin{aligned} \mathrm{E}_1(\eta_1) &= -2.483, \quad \mathrm{E}_1(\eta_2) = -2.100, \\ \mathrm{Var}_1(\eta_1) &= 0.02148, \quad \mathrm{Var}_1(\eta_2) = 0.01377. \end{aligned}$$

We then apply Bayes linear kinematics and can find the unique commutative Bayes linear kinematic solution. This is

$$\begin{aligned} \mathrm{E}_{(2)}(\eta_1; x_1, x_2) &= -2.482, \quad \mathrm{E}_{(2)}(\eta_2; x_1, x_2) = -2.110, \\ \mathrm{Var}_{(2)}(\eta_1; x_1, x_2) &= 0.02138, \quad \mathrm{Var}_{(2)}(\eta_2; x_1, x_2) = 0.01358. \end{aligned}$$

Working out the posterior beta parameter values, based upon the assumption of beta marginals holding, requires Newton's method. Generally convergence is effectively achieved within 10 iterations. The posterior parameter values are $a_1^* = 46.81, b_1^* = 537.40, a_2^* = 73.71$ and $b_2^* = 572.27$. This leads, through the usual formulae for the beta distribution, to the posterior means and variances for θ_1 and θ_2 of

$$\begin{aligned} \mathrm{E}_{(2)}(\theta_1; x_1, x_2) &= 0.0801, \quad \mathrm{E}_{(2)}(\theta_2; x_1, x_2) = 0.114, \\ \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) &= 0.0001260, \quad \mathrm{Var}_{(2)}(\theta_2; x_1, x_2) = 0.0001562. \end{aligned}$$

These results are very similar to those achieved in the logistic transformation model.

3.10 The probit transformation

The third common link function for binomial parameters is the probit function. This is

$$\eta_i = \Phi^{-1}(\theta_i),$$

where $\Phi^{-1}()$ denotes the inverse cumulative distribution function of the standard Normal distribution.

With our priors for the binomial parameters being $\theta_i \sim$beta(a_i, b_i) neither the direct approach to specifying means and variance or the approach utilising the guide relationship will give equations for the mean and variance of η_i which are solvable equations of a_i and b_i as they did in the logit and complementary log-log transformations.

This is because the cumulative distribution function for a Normal random variable cannot be written down in a simple closed form and so quantities for it can only be calculated numerically.

Thus we shall not consider the probit transformation further.

3.11 Pseudo mean and variance

We have been considering the beta-binomial case and looking at transforming θ_i to $\eta_i = g(\theta_i)$ where $g(\cdot)$ is some suitable function, such as logit. Suppose that we do not actually specify this function $g(\cdot)$ at all but simply say that

$$\hat{\mathrm{E}}_0(\eta_i) = \frac{a_i}{a_i + b_i} \quad \text{and} \quad \hat{\mathrm{Var}}_0(\eta_i) = \frac{1}{a_i + b_i}, \tag{3.26}$$

where $\hat{\mathrm{E}}(\cdot)$ and $\hat{\mathrm{Var}}(\cdot)$ are pseudo expectations and variances respectively. More generally,

$$\hat{\mathrm{E}}_0(\eta_i) = g_1\left(\frac{a_i}{a_i + b_i}\right) \quad \text{and} \quad \hat{\mathrm{Var}}_0(\eta_i) = g_2\left(\frac{1}{a_i + b_i}\right), \tag{3.27}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are suitable monotonic functions. The advantage of this is that the variance does not depend on the mean. In fact, if we make n_i observations in group i we will replace

$$\hat{\mathrm{Var}}_0(\eta_i) = \frac{1}{a_i + b_i} \quad \text{with} \quad \hat{\mathrm{Var}}_1(\eta_i) = \frac{1}{a_i + b_i + n_i},$$

so, at least if we use the identity function for $g_2(\cdot)$, as in Equation 3.26, the variance changes in a very simple and obvious way as we observe data.

The updates of the expectations are also simple and, having observed x_i successes in n_i trials, would be

$$\hat{\mathrm{E}}_1(\eta_i) = \frac{a_i + x_i}{a_i + b_i + n_i}.$$

The disadvantage of Equation 3.26 is that the mean has to be restricted to $(0, 1)$. This is awkward in a Bayes linear structure. It would be better to work on $(-\infty, \infty)$. So let us use a suitable transformation $g_1(\cdot)$. For example

$$g_1(y) = \log\left(\frac{y}{1-y}\right).$$

In this case the pseudo-expectation of η_i is

$$\hat{\mathrm{E}}_0(\eta_i) = g_1\left(\frac{a_i}{a_i + b_i}\right) = \log\left(\frac{a_i/(a_i + b_i)}{1 - a_i/(a_i + b_i)}\right) = \log\left(\frac{a_i}{b_i}\right),$$

as in Equation 3.21 where the variance, however, is $\dfrac{1}{a_i} + \dfrac{1}{b_i}$. Now if we observe x_i successes in group i the expectation and variance of η_i become

$$\hat{\mathrm{E}}_1(\eta_i) = \log\left(\frac{a_i + x_i}{b_i + n_i - x_i}\right), \quad \hat{\mathrm{Var}}_1(\eta_i) = \frac{1}{a_i + b_i + n_i}. \tag{3.28}$$

We can propagate these changes in belief about η_i through to η_j, $j \neq i$, via Equation 3.15. In order to convert our changes in belief about η_i back to quantities involving θ_i we shall need to be able to calculate a_i and b_i in terms of the mean m_i and variance v_i of η_i. Thus

$$a_i = \frac{e^{m_i}}{v_i[1+e^{m_i}]}, \quad b_i = \frac{1}{v_i[1+e^{m_i}]}. \tag{3.29}$$

Having found a_i and b_i the expectation and variance of θ_i can be calculated from the standard formulae for a beta distribution, Equation 3.11. We wish to know when a unique commutative solution exists for this model. Our uniqueness condition, Equation 3.17, in this case is

$$\frac{1}{a_i+b_i+n_i} < \frac{1}{a_i+b_i}.$$

Clearly this condition shall always hold and so a commutative Bayes linear kinematic update shall always exist. This solution is given by Equations 3.18 and 3.19.

3.11.1 Example: Heart attack data

We make prior specifications, as in previous sections, of $a_1 = 2.59$, $b_1 = 19.55$, $a_2 = 10.72$, and $b_2 = 80.84$. This results in prior pseudo-expectations and variances for η_1 and η_2 of

$$\hat{\mathrm{E}}_0(\eta_1) = -2.021, \quad \hat{\mathrm{E}}_0(\eta_2) = -2.020,$$
$$\hat{\mathrm{Var}}_0(\eta_1) = 0.04517, \quad \hat{\mathrm{Var}}_0(\eta_2) = 0.01092.$$

Having observed $x_1 = 44$ successes in group 1 and $x_2 = 62$ successes in group 2 the expectations and variances of η_1 and η_2 become

$$\hat{\mathrm{E}}_1(\eta_1) = -2.442, \quad \hat{\mathrm{E}}_1(\eta_2) = -2.039,$$
$$\hat{\mathrm{Var}}_1(\eta_1) = 0.001718, \quad \hat{\mathrm{Var}}_1(\eta_2) = 0.001583.$$

We can then propagate these changes in belief using Bayes linear kinematics via Equation 3.15. We know that a commutative solution exists as the uniqueness condition shall always hold. Thus we can use Equations 3.18 and 3.19 to find this solution. It is

$$\hat{\mathrm{E}}_{(2)}(\eta_1; x_1, x_2) = -2.441, \quad \hat{\mathrm{E}}_{(2)}(\eta_2; x_1, x_2) = -2.049,$$
$$\hat{\mathrm{Var}}_{(2)}(\eta_1; x_1, x_2) = 0.00171, \quad \hat{\mathrm{Var}}_{(2)}(\eta_2; x_1, x_2) = 0.00156.$$

The posterior values for a_i and b_i $(i = 1, 2)$ are $a_1^* = 46.77$, $b_1^* = 537.21$, $a_2^* = 73.10$ and $b_2^* = 567.13$. This leads to posterior moments for θ_1 and θ_2 of

$$\begin{aligned} \mathrm{E}_{(2)}(\theta_1; x_1, x_2) &= 0.0801, \qquad \mathrm{E}_{(2)}(\theta_2; x_1, x_2) = 0.114, \\ \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) &= 0.000126, \quad \mathrm{Var}_{(2)}(\theta_2; x_1, x_2) = 0.000158. \end{aligned}$$

3.12 Discussion

We have now seen four transformations for the beta-binomial model. In each case the posterior moments for the parameters in the heart attack example were very similar. In this section we shall compare the four methods more comprehensively and, in particular, provide some justification for the pseudo-moments method using a numerical investigation.

Clearly , if we just think about a single θ, then there is no problem with the pseudo-moment method as there is a 1-1 transformation between (a, b) and the mean and variance of θ. Interest therefore centres on the relationship between beliefs about θ_1 and θ_2.

We shall investigate how the mean and variance of θ_2 change as we make observations on group 1. We shall do this for some different example values of (a_1, a_2, b_1, b_2) and consider observations of the form np successes and $n(1-p)$ failures for some p.

A plot of the posterior expectations and variances of θ_2 under the four different models is given in Figure 3.4.

Here n is given on the x-axes and the posterior expectations and variances on the y-axes. The parameter values chosen were $(a_1, a_2, b_1, b_2) = (2, 2, 3, 3)$ for the top two plots, $(a_1, a_2, b_1, b_2) = (5, 5, 3, 3)$ for the middle two plots and $(a_1, a_2, b_1, b_2) = (10, 5, 5, 3)$ for the two bottom plots. The parameter p was 0.8 in all three cases. The black line shows the result of the conjugate update to θ_1. The other lines show the effect on beliefs about θ_2 under four different methods as follows.

- red: direct mean and variance under the logistic transformation,
- green: mode and curvature for the logistic transformation,
- dark blue: pseudo-moment method, and
- light blue: complementary log-log transformation.

We see that under all combinations of parameter values the posterior variances of θ_2 behave similarly in all models. There is more variation in the expectations. Generally

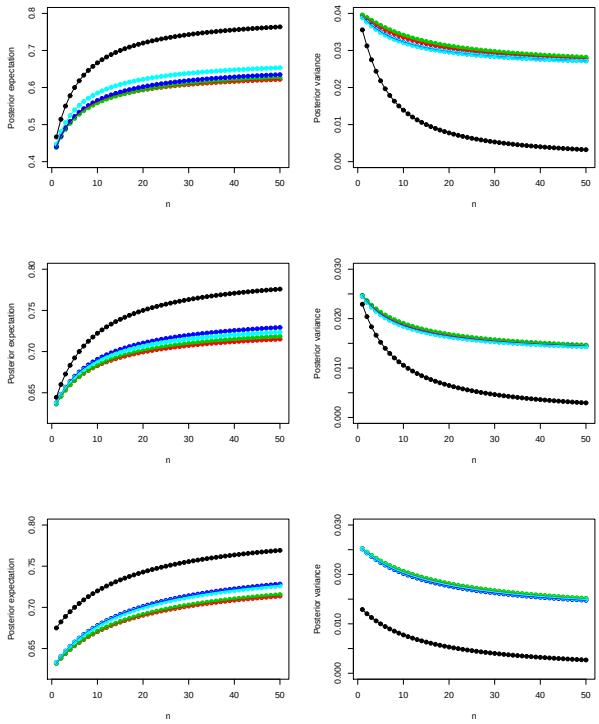

Figure 3.4: A plot of posterior expectations and variances of θ_2 after observations on group 1. Also included is the posterior expectation or variance of θ_1 in black.

the two methods employing the logistic transformation behave similarly. The pseudo-moment method is very close to these using the initial parameter values then behaves similarly to the complementary log-log model for the other two sets of parameter values. Overall there do not seem to be too many large differences between any of the methods.

3.13 The log transformation

3.13.1 Expectation and variance

In the case of Poisson random variables just one link function is commonly used. This is the natural logarithm and takes the form

$$\eta_i = \log(\theta_i),$$

for $i = 1, 2$. Clearly since $\theta_i \in [0, \infty) \Rightarrow \eta_i \in (-\infty, \infty)$ and so η_i is unbounded. We link η_1 and η_2 into a Bayes linear structure. The prior and posterior expectations and variances of θ_1, θ_2 having made within-group updates are given in Equations 3.11 and

3.12.

We can find the prior expectations and variances of each η_i. They are

$$\mathrm{E}_0(\eta_i) = \psi(a_i) - \log b_i, \quad \mathrm{Var}_0(\eta_i) = \psi_1(a_i),$$

where $\psi(\cdot)$ is the digamma function and $\psi_1(\cdot)$ is the trigamma function.

Proof. In order to find the expectation and variance of η_i in terms of the parameters of the marginal gamma distributions consider the gamma function

$$\Gamma(y) = \int_0^\infty z^{y-1}e^{-z}dz.$$

Differentiating with respect to y gives

$$\frac{d}{dy}\Gamma(y) = \int_0^\infty \log(z) \times z^{y-1}e^{-z}dz.$$

Each subsequent derivative simply multiplies the right hand side by a further $log(z)$ inside the integration. Therefore, since $b_i\theta_i \sim$gamma$(a_i, 1)$, if $z = b_i\theta_i$ then

$$\frac{1}{\Gamma(a_i)}\frac{d^n}{da_i^n}\Gamma(a_i) = \mathrm{E}_0[(\log b_i\theta_i)^n].$$

Thus the expectation of each η_i can be found as

$$\begin{aligned}
\mathrm{E}_0[\eta_i] &= \mathrm{E}_0[\log\theta_i] \\
&= \mathrm{E}_0[\log b_i\theta_i] - \log b_i \\
&= \frac{1}{\Gamma(a_i)}\frac{d}{da_i}\Gamma(a_i) - \log b_i \\
&= \psi(a_i) - \log b_i,
\end{aligned}$$

where $\psi(x) = \frac{d}{dx}\log[\Gamma(x)]$ is the digamma function. The variance is then found from

$$\begin{aligned}
\frac{d^2}{da_i^2}\Gamma(a_i) &= \frac{d}{da_i}\Gamma(a_i)\psi(a_i) \\
&= \Gamma(a_i)\psi_1(a_i) + \Gamma(a_i)\psi(a_i)^2,
\end{aligned}$$

where $\psi_1(x) = \frac{d}{dx}\psi(x)$ is the trigamma function. Thus

$$\begin{aligned}
\mathrm{Var}_0(\eta_i) &= \mathrm{Var}_0(\log b_i\theta_i) \\
&= \mathrm{E}_0[(\log b_i\theta_i)^2] - \mathrm{E}_0[\log b_i\theta_i]^2 \\
&= \psi_1(a_i) + \psi(a_i)^2 - \psi(a_i)^2 \\
&= \psi_1(a_i).
\end{aligned}$$

□

The expectation and variance of η_i having observed x_i successes in group i are

$$\mathrm{E}_0(\eta_i) = \psi(A_i) - \log B_i, \quad \mathrm{Var}_0(\eta_i) = \psi_1(A_i),$$

which are the same form as in the prior but with a_i and b_i replaced by $A_i = a_i + x_i$ and $B_i = b_i + 1$.

We propagate these changes in belief through to the other group using Equations 3.15 and 3.16. This gives $\mathrm{E}_1(\eta_1; x_2)$, $\mathrm{Var}_1(\eta_1; x_2)$, $\mathrm{E}_1(\eta_2; x_1)$ and $\mathrm{Var}_1(\eta_2; x_1)$.

We consider the sufficient condition for a unique commutative solution using Bayes linear kinematics. In this case it is

$$\psi_1(a_i + x_i) < \psi_1(a_i),$$

for some i. Thus, as long as we make a non-zero observation in either group $A_i > a_i$ and $\psi_1(A_i) < \psi_1(a_i)$ as the trigamma function is monotonically decreasing on $\mathbb{R}^+$. Thus, as long as we make a non-zero observation there will always be a unique commutative solution.

If $x_1 = x_2 = 0$ then clearly the sufficient condition does not hold. However, if we refer back to the conditions from Theorem 5 of Goldstein & Shaw (2004) as given in Section 3.4.2, then, using conditions (ii) and (iii), there is a unique commutative solution if

$$\begin{aligned}
&\mathrm{Var}_1^{-1}(\eta_1; x_1) + \mathrm{Var}_1^{-1}(\eta_1; x_2) - \mathrm{Var}_0^{-1}(\eta_1) > 0, \\
\text{or} \quad &\mathrm{Var}_1^{-1}(\eta_2; x_1) + \mathrm{Var}_1^{-1}(\eta_2; x_2) - \mathrm{Var}_0^{-1}(\eta_2) > 0.
\end{aligned}$$

If $x_1 = x_2 = 0$ then $\mathrm{Var}_1(\eta_1; x_1) = \mathrm{Var}_0(\eta_1)$. Thus, taking condition (ii),

$$\mathrm{Var}_1^{-1}(\eta_1; x_1) + \mathrm{Var}_1^{-1}(\eta_1; x_2) - \mathrm{Var}_0^{-1}(\eta_1) = \mathrm{Var}_1^{-1}(\eta_1; x_2) > 0.$$

We can show condition (iii) holds using the same reasoning. Therefore, in the case of $x_1 = x_2 = 0$, there is still a unique, commutative Bayes linear kinematic solution.

This solution is given in Equations 3.18 and 3.19 and provides posterior expectations $\mathrm{E}_{(2)}(\eta_i; x_1, x_2)$ and variances $\mathrm{Var}_{(2)}(\eta_i; x_1, x_2)$.

Assuming θ_1 and θ_2 still have marginal gamma distributions (see Section 3.7.2), we find the parameter values of these distributions by solving

$$\mathrm{E}_{(2)}(\eta_i; x_1, x_2) = \psi(a_i^*) - \log(b_i^*), \quad \mathrm{Var}_{(2)}(\eta_i; x_1, x_2) = \log(b_i^*),$$

for a_i^* and b_i^*. Then $\theta_i; x_1, x_2 \sim$gamma(a_i^*, b_i^*) and the posterior moments of θ_i are

$$\mathrm{E}_{(2)}(\theta_i; x_1, x_2) = \frac{a_i^*}{b_i^*}, \quad \mathrm{Var}_{(2)}(\theta_i; x_1, x_2) = \frac{a_i^*}{b_i^{*2}}.$$

3.13.2 Example: piston ring failures

We wish to provide a Bayes linear Bayes solution to the problem of related numbers of piston ring failures presented in Section 2.6.4. For comparability we use the same prior specifications for the prior gamma distributions of θ_1 and θ_2 used in the copula methodology. That is $a_1 = a_2 = 30$ and $b_1 = b_2 = 1$. We specify a prior correlation between η_1 and η_2 of 0.25. This leads to prior specifications for each η_i of

$$\mathrm{E}_0(\eta_1) = \mathrm{E}_0(\eta_2) = 3.384, \quad \mathrm{Var}_0(\eta_1) = \mathrm{Var}_0(\eta_2) = 0.0339.$$

When we observe $x_1 = 44$ piston ring failures in group 1 and $x_2 = 33$ failures in group 2 then the resulting posterior expectations and variances of η_1, η_2 are

$$\begin{aligned} \mathrm{E}_1(\eta_1) = 3.631, &\quad \mathrm{E}_1(\eta_2) = 3.442, \\ \mathrm{Var}_1(\eta_1) = 0.0132, &\quad \mathrm{Var}_1(\eta_2) = 0.0160. \end{aligned}$$

Thus there has been a fairly large reduction in the uncertainty on observation of the data. We can propagate these changes through to the other group using Bayes linear kinematics. A unique commutative solution exists as we have observed some failures. It is

$$\begin{aligned} \mathrm{E}_{(2)}(\eta_1; x_1, x_2) = 3.633, &\quad \mathrm{E}_{(2)}(\eta_2; x_1, x_2) = 3.471, \\ \mathrm{Var}_{(2)}(\eta_1; x_1, x_2) = 0.0131, &\quad \mathrm{Var}_{(2)}(\eta_2; x_1, x_2) = 0.0157. \end{aligned}$$

We use these values to calculate the posterior parameter values for θ_1 and θ_2 under the continued assumption of gamma marginal distributions. They are $a_1^* = 77.01, b_1^* = 2.02, a_2^* = 64.17$ and $b_2^* = 1.98$. From these we calculate the posterior expectations and

variances of the Poisson parameters.

$$\begin{aligned} \mathrm{E}_{(2)}(\theta_1; x_1, x_2) &= 38.0918, \quad \mathrm{E}_{(2)}(\theta_2; x_1, x_2) = 32.4102, \\ \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) &= 18.8423, \quad \mathrm{Var}_{(2)}(\theta_2; x_1, x_2) = 16.3699. \end{aligned}$$

3.13.3 Mode and curvature

Just as with the beta-binomial model we can utilise a guide relationship in the gamma-Poisson setup. This time we shall define our guide relationship as

$$\eta_i \approx \mu_i = \log(\theta_i).$$

Thus, rather than using the mean and variance of μ_i directly, as we did in the previous section, we use a related quantity, η_i, with mean and variance given by the mode of μ_i and curvature at the mode of the log-density of μ_i as we did in the beta-binomial model.

The density of θ_i is

$$f_{\theta_i}(\theta_i) = \frac{b_i^{a_i}\theta_i^{a_i-1}e^{-b_i\theta_i}}{\Gamma(a_i)},$$

and, in terms of μ_i, $\theta_i = e^{\mu_i}$. To obtain the density of μ_i, first we must differentiate the density of θ_i to find the Jacobian. The derivative is

$$\frac{d\theta_i}{d\mu_i} = e^{\mu_i} = \theta_i,$$

and so $d\theta_i = \theta_i d\mu_i$. Hence the density of μ_i is

$$f_{\mu_i}(\mu_i) = \frac{b_i^{a_i}\theta_i^{a_i}e^{-b_i\theta_i}}{\Gamma(a_i)}.$$

If we take logs,

$$l_i(\mu_i) = a_i \log b_i + a_i \log \theta_i - b_i\theta_i - \log \Gamma(a_i).$$

Differentiating this and setting the derivative equal to zero gives us the mode of μ_i. The derivative is

$$\frac{d}{d\mu_i}[l_i(\mu_i)] = \left(\frac{a_i}{\theta_i} - b_i\right)\theta_i = a_i - b_i\theta_i. \tag{3.30}$$

Setting this equal to zero we find the mode. Let m_i be the mode of μ_i. Then let

$$m_i^* = e^{m_i}.$$

This gives $a_i - b_i m_i^* = 0$ so $m_i^* = a_i/b_i$, and the mode of μ_i is

$$m_i = \log\left(\frac{a_i}{b_i}\right).$$

To find the curvature we must differentiate Equation 3.30 a further time. The second derivative is given by

$$\frac{d^2}{d\mu_i^2}[l_i(\mu_i)] = -b_i\theta_i.$$

At the mode $\theta_i = a_i/b_i$ so, substituting this into the above equation,

$$\left[\frac{d^2 l_i(\mu_i)}{d\mu_i^2}\right]_{m_i} = -b_i\frac{a_i}{b_i} = -a_i.$$

Therefore the required prior variance is

$$\mathrm{Var}_0(\eta_i) = -\left[\frac{d^2 l_i(\mu_i)}{d\mu_i^2}\right]_{m_i}^{-1} = \frac{1}{a_i}.$$

Hence, our two prior moments can be expressed solely in terms of the prior gamma parameters. They are

$$\mathrm{E}_0(\eta_i) = \log\left(\frac{a_i}{b_i}\right), \quad \mathrm{Var}_0(\eta_i) = \frac{1}{a_i}.$$

Having made the conjugate updates, the moments take the same form but using $A_i = a_i + x_i$ and $B_i = b_i + 1$ in place of a_i and b_i. Defining a Bayes linear structure for η_1, η_2, allows the updates to be propagated to η_j, $j \neq i$ via Equation 3.15.

From Equation 3.17 there is a unique commutative solution to the problem using Bayes linear kinematics if

$$\mathrm{Var}_1(\eta_i) < \mathrm{Var}_0(\eta_i)$$

for $i = 1$ or 2 or both. This condition will clearly hold whenever we make a non-zero observation in either of the groups. Thus a unique commutative solution will virtually always exist. In fact we showed in the previous section that a commutative solution shall exist even if both observations are zero.

Having ascertained that a unique commutative Bayes linear kinematic solution exists it is given by Equations 3.18 and 3.19. Note that, once an adjusted mean and precision for η_i are found, the parameter values of the posterior gamma distributions are found, from posterior mean $\bar{m}_i$ and variance v_i, as

$$a_i^* = \frac{1}{v_i}, \quad b_i^* = \frac{1}{v_i}e^{-\bar{m}_i}.$$

These can then be used to find posterior means and variances of θ_1, θ_2.

3.13.4 Example: piston ring failures

If we take the same prior values as in the previous example for the gamma distribution parameters and prior correlation then the resulting prior expectations and variances for η_1, η_2 are

$$\mathrm{E}_0(\eta_1) = \mathrm{E}_0(\eta_2) = 3.401, \quad \mathrm{Var}_0(\eta_1) = \mathrm{Var}_0(\eta_2) = 0.033.$$

When we observe 46 failures in compressor 1 and 33 failures in compressor 2 then, having made full Bayesian updates within each group, the expectations and variances become

$$\begin{aligned} \mathrm{E}_1(\eta_1) = 3.638, &\quad \mathrm{E}_1(\eta_2) = 3.500, \\ \mathrm{Var}_1(\eta_1) = 0.0132, &\quad \mathrm{Var}_1(\eta_2) = 0.0159. \end{aligned}$$

We can now use Bayes linear kinematics to update our beliefs in both groups as a result of these mean and variance changes. This gives $\mathrm{E}_1(\eta_i; x_j)$ and $\mathrm{Var}_1(\eta_i; x_j)$ where $j \neq i$ for each i. A commutative solution is available, using the sufficient condition, as we observe some piston ring failures. It is

$$\begin{aligned} \mathrm{E}_{(2)}(\eta_1; x_1, x_2) = 3.639, &\quad \mathrm{E}_{(2)}(\eta_2; x_1, x_2) = 3.478, \\ \mathrm{Var}_{(2)}(\eta_1; x_1, x_2) = 0.0130, &\quad \mathrm{Var}_{(2)}(\eta_2; x_1, x_2) = 0.0156. \end{aligned}$$

Solving for the posterior parameter values gives $a_1^* = 77.02, b_1^* = 2.02, a_2^* = 64.18$ and $b_2^* = 1.98$. Converting back to the expected numbers of piston ring failures results in posterior expectations and variances of

$$\begin{aligned} \mathrm{E}_{(2)}(\theta_1; x_1, x_2) = 38.0683, &\quad \mathrm{E}_{(2)}(\theta_2; x_1, x_2) = 32.3884, \\ \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) = 18.8170, &\quad \mathrm{Var}_{(2)}(\theta_2; x_1, x_2) = 16.3450. \end{aligned}$$

3.14 Direct updating

Of course, in spite of the justifications of transforming θ_1 and θ_2 given in Section 3.7, we can apply Bayes linear kinematics directly upon θ_1 and θ_2. We have the usual setup. That is

$$\theta_i \sim \mathrm{beta}(a_i, b_i), \quad \text{or} \quad \theta_i \sim \mathrm{gamma}(a_i, b_i),$$

where $X_i \mid \theta_i \sim$bin(n_i, θ_i) and $X_i \mid \theta_i \sim$Po(θ_i) respectively. The prior expectations and variances of θ_i for $i = 1, 2$ are given in Equation 3.11 in the beta-binomial case and Equation 3.13 in the Poisson-gamma case. Having observed x_i successes in group i, the expectations and variances of θ_i are given by Equations 3.12 and 3.14. As the update is conjugate we have

$$\theta_i \mid x_i \sim \text{beta}(a_i + x_i, b_i + n_i - x_i) \quad \text{or} \quad \theta_i \mid x_i \sim \text{gamma}(a_i + x_i, b_i + 1).$$

We propagate these changes through to θ_j, $j \neq i$, using Bayes linear kinematics:

$$\mathrm{E}_1(\theta_j; x_i) = \mathrm{E}_0(\theta_j) + \frac{\mathrm{Cov}_0(\theta_i, \theta_j)}{\mathrm{Var}_0(\theta_i)}\{\mathrm{E}_1(\theta_i) - \mathrm{E}_0(\theta_i)\} \tag{3.31}$$

$$\mathrm{Var}_1(\theta_j; x_i) = \mathrm{Var}_0(\theta_j) - \frac{\mathrm{Cov}_0(\theta_i, \theta_j)^2}{\mathrm{Var}_0(\theta_i)}\left\{1 - \frac{\mathrm{Var}_1(\theta_i)}{\mathrm{Var}_0(\theta_i)}\right\}. \tag{3.32}$$

We now wish to know whether a unique commutative Bayes linear kinematic solution exists. Our sufficient condition for a commutative solution to exist is, in this case, given by

$$\mathrm{Var}_1(\theta_i) < \mathrm{Var}_0(\theta_i), \tag{3.33}$$

for at least one of $i = 1$ or $i = 2$. Unlike when we have taken transformations this condition shall not always hold in this direct updating situation. This is because, as we saw in Section 3.7, variances can increase from prior to posterior for both beta and gamma distributed random variables. Thus when we update directly we must check that there is a commutative update each time individually. If such a solution does exist it is given by

$$\begin{aligned}\mathrm{E}_{(2)}(\theta_i; x_i, x_j) = \mathrm{Var}_{(2)}(\theta_i; x_i, x_j)[&\mathrm{Var}_1^{-1}(\theta_i)\mathrm{E}_1(\theta_i) \\ &+ \mathrm{Var}_1^{-1}(\theta_j; x_i)\mathrm{E}_1(\theta_j; x_i) - \mathrm{Var}_0^{-1}(\theta_i)\mathrm{E}_0(\theta_i)],\end{aligned}$$

and

$$\mathrm{Var}_{(2)}(\theta_i; x_i, x_j) = \left[\mathrm{Var}_1^{-1}(\theta_i) + \mathrm{Var}_1^{-1}(\theta_i; x_j) - \mathrm{Var}_0^{-1}(\theta_i)\right]^{-1}.$$

3.14.1 Example: Heart attack data

We shall now apply this direct Bayes linear kinematic modelling approach to the heart attack data. First we must make some prior specifications. Those we wish to make, taken directly from the copulas model, are

$$\theta_1 \sim \text{beta}(2.59, 19.55), \quad \theta_2 \sim \text{beta}(10.72, 80.84)$$

which lead to prior expectations of θ_1 and θ_2 of

$$\mathrm{E}_0(\theta_1) = 0.117, \quad \mathrm{E}_0(\theta_2) = 0.117,$$

with variances of

$$\mathrm{Var}_0(\theta_1) = 0.0045, \quad \mathrm{Var}_0(\theta_2) = 0.0011,$$

and a covariance between them (given a correlation of 0.3) of

$$\begin{aligned} \mathrm{Cov}(\theta_1, \theta_2) &= 0.3 \times \sqrt{0.0011 \times 0.0045} \\ &= 0.0006675. \end{aligned}$$

Once again we shall update these prior specifications with 44 successes (deaths) out of 560 binomial trials in group 1 and 62 successes out of 540 binomial trials in group 2. This gives the expectations and variances, found from $\theta_1 \mid X_1 = 44 \sim$beta$(46.59, 535.55)$ and $\theta_2 \mid X_2 = 62 \sim$beta$(72.72, 558.84)$, as

$$\begin{aligned} \mathrm{E}_1(\theta_1) = 0.0800, \qquad \mathrm{E}_1(\theta_2) = 0.115 \\ \mathrm{Var}_1(\theta_1) = 0.000126, \quad \mathrm{Var}_1(\theta_2) = 0.000161. \end{aligned}$$

These changes are propagated by Bayes linear kinematics using Equation 3.31. We can see that there will be a commutative Bayes linear kinematic solution in this case as $\mathrm{Var}_1(\theta_1) < \mathrm{Var}_0(\theta_1)$ and $\mathrm{Var}_1(\theta_2) < \mathrm{Var}_0(\theta_2)$ and so the sufficient condition, Equation 3.33 holds. This solution is then given by

$$\begin{aligned} \mathrm{E}_{(2)}(\theta_1; x_1, x_2) = 0.08002, \qquad \mathrm{E}_{(2)}(\theta_2; x_1, x_2) = 0.1143, \\ \mathrm{Var}_{(2)}(\theta_1; x_1, x_2) = 0.000126, \quad \mathrm{Var}_{(2)}(\theta_2; x_1, x_2) = 0.000159. \end{aligned}$$

3.15 Conclusions

In this chapter we have investigated Bayes linear approaches to the problem of correlated sets of Bernoulli trials or binomial or Poisson parameters. The first model considered Bernoulli trials and utilised a second-order exchangeable structure within each group. The result was that the sums of all of the successful Bernoulli trials were sufficient statistics for the trials themselves and reduced the computational burden associated with the updates significantly. However, the adjusted expectations and variances resulting from this model did not satisfy the mean-variance relationship of Bernoulli random variables.

We then investigated models which used Bayes linear kinematics for both the binomial

and Poisson cases. We found that within-group updating was conjugate and that changes in belief within a group could be propagated through to other groups using Bayes linear kinematic updates. We preferred to transform the binomial probabilities or Poisson parameters first onto an unrestricted scale as we felt that Bayes linear updates would then be more effective and commutative solutions easier to find. Several transformations were proposed in both cases. For the heart attack example Table 3.4 gives a comparison of the results under the different transformations.

Method	$\mathrm{E}_{(2)}(\theta_1; x_1, x_2)$	$\mathrm{E}_{(2)}(\theta_2; x_1, x_2)$	$\mathrm{Var}_{(2)}(\theta_1; x_1, x_2)$	$\mathrm{Var}_{(2)}(\theta_2; x_1, x_2)$
logit mean	0.08014	0.1146	1.26×10^{-4}	1.58×10^{-4}
logit mode	0.08012	0.1142	1.26×10^{-4}	1.57×10^{-4}
log-log	0.08013	0.1141	1.26×10^{-4}	1.56×10^{-4}
pseudo	0.08010	0.1142	1.26×10^{-4}	1.58×10^{-4}
direct	0.08002	0.1143	1.26×10^{-4}	1.59×10^{-4}

Table 3.4: A comparison of the different transformations in the beta-binomial case

The first thing we notice when we consider Table 3.4 is that the results for all of the methods are very similar. This could be due to the large number of observations in both groups and the relatively weak association assumed between them. The main difference would appear to be that the adjusted expectation for θ_1 under the direct method is lower than those of all of the other methods, which are virtually identical. This could indicate that the transformations are having an effect.

The direct method involves linear updates which are unaffected by the proximity of the lower bound at $\theta_1 = 0$. The methods using transformations show the effect of approaching this lower bound.

We can also compare the posterior expectations and variances for the Poisson example using each of the transformations. These are given in Table 3.5.

Method	$\mathrm{E}_{(2)}(\theta_1; x_1, x_2)$	$\mathrm{E}_{(2)}(\theta_2; x_1, x_2)$	$\mathrm{Var}_{(2)}(\theta_1; x_1, x_2)$	$\mathrm{Var}_{(2)}(\theta_2; x_1, x_2)$
log mean	38.0918	32.4102	18.8423	16.3699
log mode	38.0683	32.3884	18.8170	16.3450
direct	38.0880	32.5433	18.6388	15.5584

Table 3.5: A comparison of the different transformations in the gamma-Poisson case

Again the most telling thing from the table is the similarity of the results from all of the methods. Both of the variances are lower for the direct case than for the other methods (with transformations). The effect of transformations in this case would appear to be on the variances rather than the means.

We also investigated the similarities between the different methods for the beta-binomial model numerically in Section 3.12. We saw that while there were not large differences between any of the methods the two logistic transformations produced very similar results and the pseudo-moment method and complementary log-log transformation were similar.

Chapter 4

Correlated binomials in many dimensions

4.1 Introduction

In this chapter we investigate the problem of correlated binomial probabilities in more than two dimensions. To do this we utilise the Bayes linear kinematic approach introduced in the previous chapter. Again we transform the binomial parameters using a suitable link function before linking them in a Bayes linear structure. Of course there are several possible transformations for the binomial probabilities and these were considered in the previous chapter. Here we consider the logit link function. We find that when using this method there will always be a unique commutative Bayes linear kinematic solution to the problem.

We investigate an example involving data on health and smoking. For a problem with a much larger number of correlated binomials prior specification becomes an important issue. We approach this task in two stages; initially specifying the parameters of the marginal beta distributions via eliciting quantiles and then specifying a coherent covariance structure between the transformed quantities.

4.2 Elicitation of prior information

4.2.1 The beta-binomial model

Specific elicitation techniques have been developed for a range of different models. One such case is the beta-binomial model in which a number of successes X is to be observed

out of a total number of n trials. Then X can be given a binomial distribution,

$$X \mid \theta \sim \text{bin}(n, \theta)$$

and the true probability of success, θ, can be represented using a conjugate beta prior distribution with parameters a and b,

$$\theta \sim \text{beta}(a, b).$$

Chaloner & Duncan (1983) proposed a method of eliciting the parameters of the beta distribution in this model. They termed the method the predictive mode (PM) method. The PM method uses an iterative scheme based on the predictive distribution of the binomial random variable X.

The expert first estimates their prior mode m for the predictive distribution for a chosen number of n trials. The probability that the binomial variable takes some value x is

$$\begin{aligned} f(x) &= \int_0^1 f(x \mid \theta) f(\theta) d\theta \\ &= \frac{\Gamma(n+1)}{\Gamma(x+1)\Gamma(n-x+1)} \frac{\Gamma(a+b)}{\Gamma(a+b+n)} \frac{\Gamma(a+x)}{\Gamma(a)} \frac{\Gamma(b+n-x)}{\Gamma(b)}. \end{aligned} \tag{4.1}$$

Next the ratios of this probability at $m-1$ and $m+1$ to the probability at the mode m can be calculated in terms of a and b. These are

$$\begin{aligned} r_{+1} &= \frac{f(m+1)}{f(m)} = \frac{(n-m)(m-a)}{(m+1)(n-m+b-1)}, \\ r_{-1} &= \frac{f(m-1)}{f(m)} = \frac{m(n-m+b)}{(n-m+1)(m+a-1)}. \end{aligned}$$

Once r_{+1} and r_{-1} are specified then a and b can be found. The mode is fixed from this. The spread of the distribution is estimated using an iteration scheme involving 50% prediction intervals. Once the expert is satisfied with the prediction interval the final values of a and b are calculated.

Gavaskar (1988) developed an alternative approach in which, as in the Chaloner & Duncan (1983) method, the mode m of the predictive distribution is estimated by the expert given n trials. The expert is then asked for a new mode m_i for n_i trials when presented with a fictitious sample of t_i trials in which there were s_i successes $i = 1, \ldots, k$.

The hyperparameters are chosen using the modes of the predictive distribution. The beta-binomial distribution may have one or two modes. They are given, for the beta-

binomial distribution in Equation 4.1, by all of the integers in

$$\left[\frac{n(a-1)-(b-1)}{a+b-2}, \frac{(n+1)(a-1)}{a+b-2}\right],$$

an interval of length 1 (Gavaskar, 1988). Thus the modes are the closest integers to

$$\frac{(n+1)(a-1)}{a+b-2} - \frac{1}{2}.$$

Then a and b can be found to minimise

$$D = \sum_{i=1}^{k} \left(m_i - \left[\frac{(n_i+1)(a+s_i-1)}{a+b+t_i-2} - \frac{1}{2} \right] \right)^2.$$

Garthwaite *et al.* (2005) comment on both of these methods and indicate that they are essentially applications of the four main elicitation techniques for beta parameters, though usually methods estimate θ directly in some way. The four main methods of elicitation in the beta-binomial case, based upon Winkler (1967), are

- the quantile method,
- the hypothetical future sample method,
- the equivalent prior sample method,
- and the probability density function method.

The quantile method consists of;

1. The expert specifies his/her prior estimate of m, the median of θ.
2. He/she then estimates $q_1, \ldots, q_k$, k different quantiles of θ. Usually k is small (often $k = 2$).
3. These estimates are compared to beta(a, b) for various a and b and suitable values are chosen.

The steps of the hypothetical future samples method are;

1. The expert specifies θ^*, a 'prior' estimate for θ.
2. He/she is then presented with hypothetical sample data of $s = (s_1, \ldots, s_k)$ success from $t = (t_1, \ldots, t_k)$ trials.
3. For each pair (s_i, t_i) he/she estimates a 'posterior' $\theta_i^* \mid s_i$.

4. Each $\theta_i^* \mid s_i$ is combined with θ^* and a_i and b_i are calculated.

5. The a_i and b_i's are averaged to find a and b.

The method of Gavaskar (1988) is in essence a hypothetical future samples method. The equivalent sample method is implemented as;

1. The expert expresses his/her prior beliefs about θ in the form of a hypothetical sample, specifying an estimated number of successes s_0 out of a hypothetical number of trials t_0.

2. The prior distribution is then $\theta \sim$beta$(s_0, t_0 - s_0)$.

Finally, the probability density method, of which the iteration scheme of Chaloner & Duncan (1983) is a variant, takes the form

1. The expert specifies $\hat{\theta}$, his/her most likely value of θ.

2. Estimates of $\hat{\theta}_{-1/2}$ and $\hat{\theta}_{+1/2}$ are also elicited, the values of θ judged to be half as likely as $\hat{\theta}$.

3. The values of the hyperparameters, a and b are found from these via the beta probability density function.

All of these methods have been found to elicit a beta prior distribution which is 'over confident'. That is, the variance is underestimated by each of the four methods (Winkler, 1967; Schaefer & Borcherding, 1973; Garthwaite *et al.*, 2005). In their review of elicitation methods in the beta-binomial case Garthwaite *et al.* (2005) conclude that the quantile method, of the four, minimises this over confidence.

4.3 Multiple updates using Bayes linear kinematics

Consider the situation in which there are p collections of random quantities $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_p$ where

$$\boldsymbol{U}_k = (U_{k1}, \ldots, U_{kn_k})'$$

for $k = 1, \ldots, p$. Suppose that a full second order prior specification has been made for $\boldsymbol{U} = \boldsymbol{U}_1 \cup \ldots \cup \boldsymbol{U}_p$ of the form $S_0(\boldsymbol{U}) = [\mathrm{E}_0(\boldsymbol{U}), \mathrm{Var}_0(\boldsymbol{U})]$ and that data information I_k is received which causes the beliefs about $\boldsymbol{U}_k$ to be updated to $S_1(\boldsymbol{U}_k;\ I_k) = [\mathrm{E}_1(\boldsymbol{U}_k), \mathrm{Var}_1(\boldsymbol{U}_k)]$. Then, as in Equations 3.6 and 3.7, the Bayes linear kinematic update for $\boldsymbol{U}$ is

$$\mathrm{E}_1(\boldsymbol{U};\ I_k) = \mathrm{E}_0(\boldsymbol{U}) + \mathrm{Cov}_0(\boldsymbol{U}, \boldsymbol{U}_k)\mathrm{Var}_0^{-1}(\boldsymbol{U}_k)[\mathrm{E}_1(\boldsymbol{U}_k) - \mathrm{E}_0(\boldsymbol{U}_k)], \tag{4.2}$$

$$\mathrm{Var}_1(\boldsymbol{U};\ I_k) = \mathrm{Var}_0(\boldsymbol{U};\ I_k) + \mathrm{Cov}_0(\boldsymbol{U}, \boldsymbol{U}_k)\mathrm{Var}_0^{-1}(\boldsymbol{U}_k) \\ \mathrm{Var}_1(\boldsymbol{U}_k)\mathrm{Var}_0^{-1}(\boldsymbol{U}_k)\mathrm{Cov}_0(\boldsymbol{U}_k, \boldsymbol{U}). \quad (4.3)$$

Now suppose that data are observed and beliefs updated once for each of $k = 1, \ldots, p$. A Bayes linear kinematic update can be made for $\boldsymbol{U}$ each time.

As we saw in Chapter 3, successive Bayes linear kinematic updates are not necessarily commutative. However, Goldstein & Shaw (2004) give conditions under which the requirement of commutativity leads to a unique Bayes linear kinematic update. In the analyses in this chapter each $\boldsymbol{U}_k$ is always a scalar U_k and a sufficient condition for a unique commutative update is

$$\mathrm{Var}_0^{-1}(U_k)\mathrm{Var}_1(U_k) < 1, \quad (4.4)$$

for all $k = 1, \ldots, p$. This solution, when it exists, is given by

$$\mathrm{E}_p(\boldsymbol{U}; \boldsymbol{I}) = \mathrm{Var}_p(\boldsymbol{U}; \boldsymbol{I}) \left[\sum_{k=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{U};\ I_k)\mathrm{E}_1(\boldsymbol{U};\ I_k) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{U})\mathrm{E}_0(\boldsymbol{U}) \right] \quad (4.5)$$

$$\mathrm{Var}_p(\boldsymbol{U}; \boldsymbol{I}) = \left[\sum_{k=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{U};\ I_k) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{U}) \right]^{-1}, \quad (4.6)$$

where $\boldsymbol{I} = (I_1, \ldots, I_p)'$.

4.4 Bayes linear kinematics for correlated binomials

Suppose we have p groups, which could be different machines or time periods, with $p > 2$. In each group we shall observe a number of successes out of a total number of trials so that X_i for $i = 1, \ldots, p$, has a binomial distribution

$$X_i \mid \theta_i \sim \mathrm{bin}(n_i, \theta_i),$$

where θ_i is the unknown probability of observing a success and n_i is the known number of trials. We assume that, given θ_i and θ_j, X_i and X_j $(i \neq j)$ are uncorrelated. Further suppose that each θ_i is given a conjugate beta prior distribution. Thus

$$\theta_i \sim \mathrm{beta}(a_i, b_i),$$

for some parameter values a_i and b_i. The prior expectation and variance of θ_i are

$$\mathrm{E}_0(\theta_i) = \frac{a_i}{a_i + b_i}, \quad \mathrm{Var}_0(\theta_i) = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)}.$$

Having observed $X_i = x_i$ successes out of n_i trials in group i the distribution of θ_i is updated to $\theta_i \mid x_i \sim$beta$(a_i + x_i, b_i + n_i - x_i)$ giving a posterior mean and variance of

$$\mathrm{E}_1(\theta_i) = \frac{a_i + x_i}{a_i + b_i + n_i}, \quad \mathrm{Var}_1(\theta_i) = \frac{(a_i + x_i)(b_i + n_i - x_i)}{(a_i + b_i + n_i)^2(a_i + b_i + n_i + 1)}.$$

We wish to propagate these changes in belief within a group to the other $p - 1$ groups. To do this first we shall transform the θ_i's. Clearly we have, as we had in chapter 3, many options of which transformation to take in the case of the beta-binomial model. We shall consider the logistic transformation

$$\eta_i = \log\left(\frac{\theta_i}{1 - \theta_i}\right)$$

so that for $\theta_i \in [0, 1]$ we have $\eta_i \in (-\infty, \infty)$. Thus we perform Bayes linear updating on the η's, a more suitable scale for linear fitting to take place. We discussed this in Section 3.7. To do this we must initially make a full second-order prior specification for $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$. That is we specify $S_0(\boldsymbol{\eta}) = [\mathrm{E}_0(\boldsymbol{\eta}), \mathrm{Var}_0(\boldsymbol{\eta})]$, where $\mathrm{E}_0(\boldsymbol{\eta})$ is a vector of prior expectations and $\mathrm{Var}_0(\boldsymbol{\eta})$ is a prior variance matrix.

To do this we need to find the expectations and variances of η_i. We showed how to do this in Section 3.8.1 and, in terms of the parameters of the marginal beta distributions, they are

$$\begin{aligned} \mathrm{E}_0(\eta_i) &= \psi(a_i) - \psi(b_i) \\ \mathrm{Var}_0(\eta_i) &= \psi_1(a_i) + \psi_1(b_i), \end{aligned}$$

where $\psi(y)$ is the digamma function and $\psi_1(y)$ is the trigamma function. Having observed x_i we can update within each group to obtain $\mathrm{E}(\eta_i \mid x_i)$ and $\mathrm{Var}(\eta_i \mid x_i)$ using full Bayesian conjugate updating. Then

$$\begin{aligned} \mathrm{E}(\eta_i \mid x_i) &= \psi(a_i + x_i) - \psi(b_i + n_i - x_i), \\ \mathrm{Var}(\eta_i \mid x_i) &= \psi_1(a_i + y_i) + \psi_1(b_i + n_i - y_i). \end{aligned}$$

These changes are propagated through to the other groups using Bayes linear kinematics, Equations 4.2 and 4.3. This gives

$$\mathrm{E}_1(\boldsymbol{\eta}; x_i) = \mathrm{E}_0(\boldsymbol{\eta}) + \mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)\mathrm{Var}_0^{-1}(\eta_i)\left[\mathrm{E}(\eta_i \mid x_i) - \mathrm{E}_0(\eta_i)\right]$$

and

$$\begin{aligned}\mathrm{Var}_1(\boldsymbol{\eta}; x_i) = \mathrm{Var}_0(\boldsymbol{\eta}) &- \mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Cov}_0(\eta_i, \boldsymbol{\eta}) \\ &+ \mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Var}(\eta_i \mid x_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Cov}_0(\eta_i, \boldsymbol{\eta}).\end{aligned}$$

From the sufficient condition in Equation 4.4 we have a unique commutative solution if

$$\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Var}(\eta_i \mid x_i) < 1$$

for all i. We showed in Section 3.8.1 that a variance decrease will always result from observing data using this transformation and so the uniqueness condition shall always be satisfied. The Bayes linear kinematic commutative solution, having observed $\boldsymbol{x} = (x_1, \ldots, x_p)'$, is

$$\begin{aligned}\mathrm{E}_p(\boldsymbol{\eta}; \boldsymbol{x}) &= \mathrm{Var}_p(\boldsymbol{\eta}; \boldsymbol{x}) \left[\sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta}; x_i)\mathrm{E}_1(\boldsymbol{\eta}; x_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\mathrm{E}_0(\boldsymbol{\eta}) \right] \\ \mathrm{Var}_p(\boldsymbol{\eta}; \boldsymbol{x}) &= \left[\sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta}; x_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta}) \right]^{-1}.\end{aligned}$$

We can also calculate adjusted quantities in terms of the θ_i's under the assumption that they still follow a beta distribution. Initially we solve the following two equations for posterior parameter values a_i^* and b_i^*.

$$\begin{aligned}\mathrm{E}_p(\eta_i; \boldsymbol{x}) &= \psi(a_i^*) - \psi(b_i^*), \\ \mathrm{Var}_p(\eta_i; \boldsymbol{x}) &= \psi_1(a_i^*) + \psi_1(b_i^*).\end{aligned}$$

This gives a distribution for θ_i of $\theta_i; \boldsymbol{x} \sim \mathrm{beta}(a_i^*, b_i^*)$ and adjusted mean and variance of

$$\mathrm{E}_p(\theta_i; \boldsymbol{x}) = \frac{a_i^*}{a_i^* + b_i^*}, \quad \mathrm{Var}_p(\theta_i; \boldsymbol{x}) = \frac{a_i^* b_i^*}{(a_i^* + b_i^*)^2(a_i^* + b_i^* + 1)}.$$

4.5 Illustrative Example: Smoking and health

We are given data on males in Canada (Table 4.1). The data are concerned with the effects of smoking on health. They were first published by Best & Walker (1964) and reproduced in Hand *et al.* (1994). The subjects are split into four groups depending on their smoking habits;

- non-smokers

- cigarette smokers
- other (e.g. pipe, cigar, etc.)
- cigarette and other.

After 6 years the number of deaths of subjects in each group is recorded. The subjects are also grouped into five year age bands. There are 9 different age bands, beginning at age 40, and so in total we have 36 groups of patients. A plot of the data is given

Age	Non-smoker	Died	Cigarette	Died	Other	Died	Both	Died
40-44	656	18	3410	124	145	2	4531	149
45-49	359	22	2239	140	104	4	3030	169
50-54	249	19	1851	187	98	3	2267	193
55-59	632	55	3270	514	372	38	4682	576
60-64	1067	117	3791	778	846	113	6052	1001
65-69	897	170	2421	689	949	173	3880	901
70-74	668	179	1195	432	824	212	2033	613
75-80	361	120	436	214	667	243	871	337
>80	274	120	113	63	537	253	345	189

Table 4.1: Death rates amongst subjects classified by smoking habits and age

in Figure 4.1. If we denote the number of deaths in category i and age band j as X_{ij}

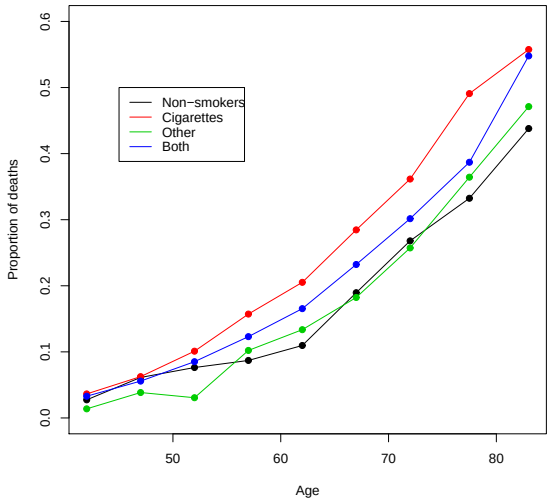

Figure 4.1: A plot of death proportions in the 4 groups

then

$$X_{ij} \mid \theta_{ij} \sim \text{bin}(n_{ij}, \theta_{ij})$$

for $i = 1, \ldots, 4$ and $j = 1, \ldots, 9$, where θ_{ij} is the probability of death for a subject in category i of age j. The total number of subjects in group ij is n_{ij}. All of the X_{ij}'s are independent conditional on the θ_{ij}'s. We can give each θ_{ij} a conjugate beta prior distribution

$$\theta_{ij} \sim \text{beta}(a_{ij}, b_{ij}).$$

It seems reasonable to assume that θ_{ij} is not independent of θ_{kl} for $i \neq k$ or $j \neq l$ as learning about the death rate with one combination of smoking habits and age would affect our beliefs about the death rates elsewhere. A parametric regression might be used in this situation but in our analysis we prefer not to assume a particular functional form for the relationship of death rate to age.

4.5.1 Elicitation of prior means and variances

Following the advice of Garthwaite *et al.* (2005) we shall use the quantile method in order to find the prior parameters, a and b, of the marginal beta distributions for each group. If we elicit values z_1 and z_2 which correspond to the $(100 \times q_1)\%$ and $(100 \times q_2)\%$ quantiles of the beta distribution then there are exact values of a and b which correspond to these assessments.

The method we use to solve for a and b utilises two simple interval halving algorithms. The first calculates the value of b for a given a. It proceeds as follows:

1. Specify a range in which b is known to lie; $b \in (b_\alpha, b_\gamma)$, $b_\alpha < b_\gamma$.
2. Specify a value for a and elicit z_1, the $(100 \times q_1)\%$ quantile of the distribution of θ.
3. Find $F_\theta(z_1; a, b_\alpha)$ and $F_\theta(z_1; a, b_\gamma)$, the values of the distribution function of θ at z_1 for b_α and b_γ.
4. Calculate

$$\begin{aligned} d_\alpha^{(b)} &= F_\theta\left(z_1; a, b_\alpha\right) - q_1 \\ d_\gamma^{(b)} &= F_\theta\left(z_1; a, b_\gamma\right) - q_1, \end{aligned}$$

   the difference between the distribution function at z_1 found using a with b_α and b_γ and its elicited value.
5. Find $b_1 = \frac{b_\alpha + b_\gamma}{2}$ and use this to calculate $d_1^{(b)}$.
6. If $d_1^{(b)}$ has the same sign as $d_\alpha^{(b)}$ then replace b_α with b_1, if not replace b_γ with b_1.

7. Repeat steps 3-6 until the value of b is found.

The second algorithm calculates both a and b, using the previous algorithm in order to do this. It is

1. Specify a range in which a is known to lie; $a \in (a_\alpha, a_\gamma)$, $a_\alpha < a_\gamma$ and then elicit the $(100 \times q_2)\%$ quantile, z_2, of the distribution of θ.
2. For a_α and a_γ find b_α and b_γ using the algorithm above.
3. Calculate $F_\theta(z_2; a_\alpha, b_\alpha)$ and $F_\theta(z_2; a_\gamma, b_\gamma)$, the values of the distribution function of θ at z_2 using (a_α, b_α) and (a_γ, b_γ).
4. Find

$$\begin{aligned} d_\alpha^{(a)} &= F_\theta\left(z_2; a_\alpha, b_\alpha\right) - q_2 \\ d_\gamma^{(a)} &= F_\theta\left(z_2; a_\gamma, b_\gamma\right) - q_2, \end{aligned}$$

   the difference between the distribution function at z_2 using (a_α, b_α) and (a_γ, b_γ) and its elicited value.
5. Calculate $a_1 = \frac{a_\alpha + a_\gamma}{2}$ and use this to calculate $d_1^{(a)}$.
6. If $d_1^{(a)}$ has the same sign as $d_\alpha^{(a)}$ then replace a_α with a_1, if not replace a_γ with a_1.
7. Repeat steps 2-6 until the value of a is found.

When implementing these algorithms it is found that 5-15 iterations of each is generally sufficient to calculate the values of a and b accurately to at least four decimal places.

Of course, alternative methods could be used to calculate the parameter values. Examples include the Newton-Raphson method, the secant method and Broydon's method. Convergence would generally be quicker than when using simple interval halving but, in many such alternative methods, convergence to the correct answer is not guaranteed. Also, when using the Newton-Raphson and Broydon methods amongst others, it is necessary to calculate derivatives for the functions of interest. Such derivatives would be far from straighforward to find in our problem.

We feel that, since the interval halving method converges quickly in the problem considered here, a more complex method is not necessary.

Returning to the data, for each i and j, we shall specify a measure of location and a measure of spread. In terms of location the prior median for θ_{ij}, m_{ij}, shall be elicited.

That is, m_{ij}, such that

$$\int_0^{m_{ij}} f_0(\theta_{ij})d\theta_{ij} = 0.5, \tag{4.7}$$

where $f_0(\theta_{ij})$ is the prior marginal beta density of θ_{ij}.

As our measure of spread we shall elicit two quantiles of $f_0(\theta_{ij})$. If we are to elicit the $100 \times q_1\%$ and $100 \times q_2\%$ quantiles with $q_1 < q_2$, and their values are given by $y_{ij}^{(-)}$ and $y_{ij}^{(+)}$ respectively, then we have the relation

$$\int_{y_{ij}^{(-)}}^{y_{ij}^{(+)}} f_0(\theta_{ij})d\theta_{ij} = q_2 - q_1. \tag{4.8}$$

Equations 4.7 and 4.8 are solved using the interval halving method to give suitable values of a_{ij} and b_{ij}. On the advice of Garthwaite & O'Hagan (2000) we shall choose to elicit tertiles and so $q_1 = 0.33$ and $q_2 = 0.67$.

To elicit these quantities questions can be put to the expert in terms of the average proportion of deaths that would be observed in that group over a large number of experiments. As 3 quantiles are being elicited to calculate 2 values (a_{ij} and b_{ij}) there is no exact solution in general. However we can find exact values of the beta parameters for each combination of 2 of the 3 quantiles;

$$\begin{aligned}
\left(y_{ij}^{(-)}, m_{ij}\right) &\Rightarrow (a_{1ij}, b_{1ij}) \\
\left(m_{ij}, y_{ij}^{(+)}\right) &\Rightarrow (a_{2ij}, b_{2ij}) \\
\left(y_{ij}^{(-)}, y_{ij}^{(+)}\right) &\Rightarrow (a_{3ij}, b_{3ij}).
\end{aligned}$$

We can then use our transformations again. We find prior means and variances from each set of parameter values above on the unrestricted scale via

$$\bar{m}_{kij} = \psi(a_{kij}) - \psi(b_{kij}), \quad v_{kij} = \psi_1(a_{kij}) + \psi_1(b_{kij}),$$

for $k = 1, 2, 3$. We can then calculate the prior mean and variance of η_{ij} as a weighted average of the expectations and variances from each pair of estimates as

$$\begin{aligned}
\bar{m}_{ij} &= w_{1ij}\bar{m}_{1ij} + w_{2ij}\bar{m}_{2ij} + w_{3ij}\bar{m}_{3ij} \\
v_{ij} &= \frac{1}{w_{ij}}\left(w_{1ij}^2 v_{1ij} + w_{2ij}^2 v_{2ij} + w_{3ij}^2 v_{3ij}\right),
\end{aligned}$$

for some weights $w_{1ij}, w_{2ij}, w_{3ij}$ where $w_{ij} = w_{1ij}^2 + w_{2ij}^2 + w_{3ij}^2$. This gives two equations in a_{ij} and b_{ij} which can be solved. The weights can be chosen to represent the relative confidence of the expert in their quantile specifications. Often a sensible choice would

appear to be $w_{1ij} = w_{2ij} = w_{3ij} = 1/3$.

The prior parameter values are then found from these. To see whether this is a sensible method for specifying prior distributions let us consider two groups, the first being cigarette smokers aged 45-49. Suppose that for this group the three quantiles elicited were $(y^{(-)}, m, y^{(+)}) = (0.048, 0.065, 0.084)$. Then, using the above method with $w_1 = w_2 = w_3 = 1/3$ gives Figure 4.2.

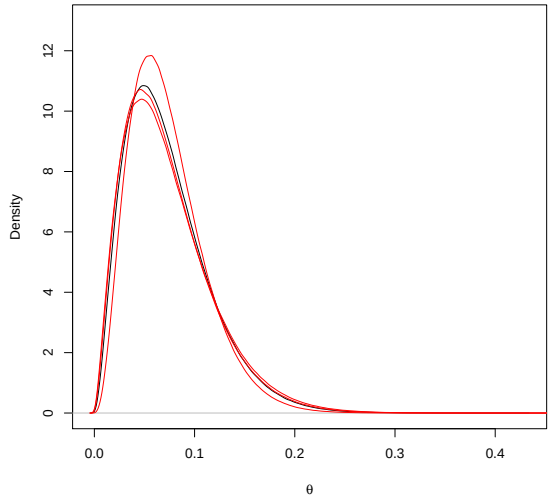

Figure 4.2: A plot of the prior beta distribution for cigarette smokers aged 45-49

The red curves are the beta distributions taken from the three sets of two quantiles. The overall beta prior distribution which results using the method above is given in black. We see that the black curve appears to be a sensible combination of the information from the three red curves.

The second group we shall consider is category "both, age group 55-59". Suppose the three quantiles elicited for this group were $(y^{(-)}, m, y^{(+)}) = (0.124, 0.167, 0.219)$. Then, using the same weights as before, the equivalent plot is in Figure 4.3.

The three constituent curves are very similar in this case. The overall black line still goes right through them (so much so it can hardly be seen) and so this method seems to be working well in the example.

Let us suppose that non-smokers will have lower death rates than smokers and this difference will become larger with increasing age. Further suppose that cigarettes are more harmful than other forms of smoking such as pipes. From this process the values of a and b chosen for each of the four groups as well as the lower and upper tertiles they correspond to are given in Tables 4.2 and 4.3.

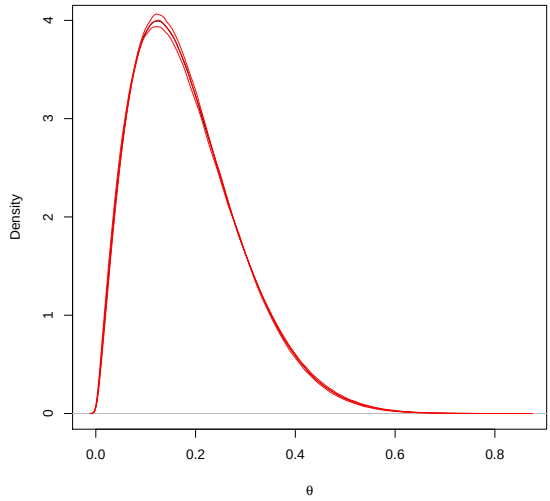

Figure 4.3: A plot of the prior beta distribution for category both aged 55-59

4.5.2 Elicitation of prior covariances

To complete the second-order prior specification covariances for $\boldsymbol{\eta}$ must be specified. This can be achieved by eliciting quantities involving the θ_{ij}'s. For each pair θ_{ij} and θ_{kl} with $(i,j) \neq (k,l)$ a prior covariance $\text{Cov}_0(\theta_{ij}, \theta_{kl})$ is required. This covariance can be elicited by asking the expert to imagine they know the 'true' value of θ_{ij} from a very large experiment and it left the median unchanged at m_{ij}. New tertiles, $y_{ij}^{(-)'}$ and $y_{ij}^{(+)'}$, are then elicited for θ_{ij} having learned θ_{kl}.

From these the parameters $a_{ij}^{'}$ and $b_{ij}^{'}$ of the beta distribution can be found as when making the prior specifications. Thus

$$\theta_{ij} \mid \theta_{kl} \sim \text{beta}(a_{ij}^{'}, b_{ij}^{'}).$$

If the expert judges that θ_{ij} and θ_{kl} are unrelated then $y_{ij}^{(-)'} = y_{ij}^{(-)}$ and $y_{ij}^{(+)'} = y_{ij}^{(+)}$, i.e., the elicited tertiles would remain unchanged as nothing has been learned about θ_{ij} by learning θ_{kl}. If the expert judges that there is a relation between θ_{ij} and θ_{kl} then $y_{ij}^{(-)'} > y_{ij}^{(-)}$ and $y_{ij}^{(+)'} < y_{ij}^{(+)}$, i.e., the elicited tertiles will have moved closer together indicating a reduction in uncertainty about θ_{ij} having observed θ_{kl}. The closer together the tertiles become the stronger the association between the two quantities.

Transforming back to η_{ij} gives $\text{Var}(\eta_{ij} \mid \eta_{kl}) = \psi_1(a_{ij}^{'}) + \psi_1(b_{ij}^{'})$ and since

$$\frac{\text{Var}(\eta_{ij} \mid \eta_{kl})}{\text{Var}_0(\eta_{ij})} = 1 - \frac{\text{Cov}_0^2(\eta_{ij}, \eta_{kl})}{\text{Var}_0(\eta_{ij})\text{Var}_0(\eta_{kl})},$$

	Non-smokers					Cigarettes				
Age	$y^{(-)}$	m	$y^{(+)}$	a	b	$y^{(-)}$	m	$y^{(+)}$	a	b
40-44	0.013	0.016	0.020	3.876	216.6	0.011	0.014	0.018	2.726	167.8
45-49	0.047	0.059	0.074	3.579	51.89	0.048	0.065	0.084	2.540	32.40
50-54	0.074	0.095	0.119	3.319	28.85	0.082	0.110	0.144	2.354	16.75
55-59	0.111	0.143	0.179	3.060	16.72	0.124	0.167	0.219	2.169	9.517
60-64	0.146	0.190	0.239	2.763	10.74	0.160	0.220	0.290	1.909	5.954
65-69	0.184	0.241	0.306	2.429	6.956	0.213	0.296	0.390	1.687	3.579
70-74	0.219	0.293	0.374	2.095	4.618	0.251	0.358	0.476	1.427	2.317
75-80	0.245	0.337	0.438	1.724	3.097	0.291	0.427	0.573	1.167	1.464
>80	0.277	0.393	0.519	1.390	1.983	0.329	0.494	0.660	1.001	1.019

Table 4.2: The prior values of the beta parameters for non-smokers and cigarette smokers

	Other					Both				
Age	$y^{(-)}$	m	$y^{(+)}$	a	b	$y^{(-)}$	m	$y^{(+)}$	a	b
40-44	0.011	0.014	0.018	2.985	187.8	0.011	0.014	0.018	2.726	167.8
45-49	0.043	0.057	0.073	2.763	40.67	0.048	0.065	0.084	2.540	32.40
50-54	0.078	0.104	0.134	2.577	19.80	0.082	0.110	0.144	2.354	16.75
55-59	0.119	0.159	0.206	2.354	11.04	0.124	0.167	0.219	2.169	9.517
60-64	0.156	0.211	0.275	2.095	6.956	0.160	0.220	0.290	1.909	5.954
65-69	0.168	0.236	0.314	1.761	4.989	0.213	0.296	0.390	1.687	3.579
70-74	0.221	0.314	0.419	1.501	2.911	0.251	0.358	0.476	1.427	2.317
75-80	0.251	0.369	0.499	1.241	1.909	0.291	0.427	0.573	1.167	1.464
>80	0.294	0.444	0.604	1.019	1.204	0.329	0.494	0.660	1.001	1.019

Table 4.3: The prior values of the beta parameters for other and both

the modulus of the prior covariance between η_{ij} and η_{kl} is given by

$$|\mathrm{Cov}_0(\eta_{ij}, \eta_{kl})| = \sqrt{\mathrm{Var}_0(\eta_{kl})[\mathrm{Var}_0(\eta_{ij}) - \mathrm{Var}(\eta_{ij} \mid \eta_{kl})]}.$$

The sign can be determined by asking whether the expert's expectation for η_{ij} would increase or decrease upon learning that η_{kl} was greater than expected.

4.5.3 Elicitation of a prior covariance structure

Clearly for the smoking and health example eliciting each covariance individually, using the method described in Section 4.5.2, would be completely impractical. It may also be difficult to avoid accidental incoherence in the resulting covariance matrix. We shall, therefore, adopt a more structural approach, taking advantage of ideas from Farrow

(2003). Thus we represent η_{ij} as

$$\eta_{ij} - \mathrm{E}_0(\eta_{ij}) = c_{1ij}U_{1j} + \ldots + c_{pij}U_{pj} + F_{ij},$$

where $\mathrm{E}_0(\eta_{ij})$ is the prior expectation of η_{ij}, U_{kj} is a common uncertainty factor (in at least two groups), c_{kij} is a coefficient to be chosen and F_{ij} is a specific uncertainty factor for group ij.

These uncertainty factors are defined so that they have zero mean and U_{kj} and U_{lj} are independent for $k \neq l$. Specific uncertainty factors are also independent of one another and all common uncertainty factors.

The specific uncertainty factor F represents uncertainty within a certain group and U represents shared uncertainty between two or more groups. We give each c_{kij} the value 1 or 0 as observing a higher proportion of deaths in one group than we expected would lead us to revise upwards our beliefs about the probability of death in other groups.

Specifically, the uncertainty factors shall take the form

$$\begin{aligned}
\eta_{1j} - \mathrm{E}_0(\eta_{1j}) &= U_{0j} + U_{3j} + F_{1j} \\
\eta_{2j} - \mathrm{E}_0(\eta_{2j}) &= U_{0j} + U_{1j} + U_{2j} + F_{2j} \\
\eta_{3j} - \mathrm{E}_0(\eta_{3j}) &= U_{0j} + U_{1j} + U_{3j} + F_{3j} \\
\eta_{4j} - \mathrm{E}_0(\eta_{4j}) &= U_{0j} + U_{1j} + U_{2j} + F_{4j},
\end{aligned}$$

where U_{0j} represents uncertainty common to all groups, U_{1j} represents uncertainty associated with smoking, U_{2j} represents uncertainty associated with cigarettes and U_{3j} represents uncertainty associated with not smoking cigarettes.

We shall now consider age. It seems reasonable that U_{kj} and $U_{kj'}$ will be related for $j \neq j'$. First we express our uncertainty factors in terms of the following uncorrelated components

$$\begin{aligned}
U_{kj} &= M_k^{(U)} + A_{kj}^{(U)} \\
F_{ij} &= M_i^{(F)} + A_{ij}^{(F)},
\end{aligned}$$

for $i = 1, \ldots, 4$, $j = 1, \ldots, 9$, and $k = 0, \ldots, 3$. Here M represents the overall uncertainty level of a factor and A shall represent the uncertainty relationship between different ages within each factor. We can link different ages within a first-order autoregression so that

$$\begin{aligned}
A_{kj}^{(U)} &= \phi_k^{(U)} A_{kj-1}^{(U)} + \epsilon_{kj}^{(U)} \\
A_{ij}^{(F)} &= \phi_i^{(F)} A_{ij-1}^{(F)} + \epsilon_{ij}^{(F)}
\end{aligned}$$

where $\mathrm{E}(\epsilon_{kj}^{(U)}) = \mathrm{E}(\epsilon_{ij}^{(F)}) = 0$, $\mathrm{E}(\epsilon_{kj}^{(U)2}) = \mathrm{E}(\epsilon_{ij}^{(F)2}) = 0$ and all of the ϵ's have zero covariances between them.

Depending on our beliefs about the relationships between uncertainties at different ages we could use a stationary or non-stationary autoregressive process.

Stationary process

Let us suppose that $\epsilon_{kj}^{(U)}$ and $\epsilon_{ij}^{(F)}$ have the variances

$$\begin{aligned} \mathrm{Var}(\epsilon_{kj}^{(U)}) &= v_{\epsilon_k}^{(U)} \\ \mathrm{Var}(\epsilon_{ij}^{(F)}) &= v_{\epsilon_i}^{(F)}. \end{aligned}$$

If we set the initial variances of $A_{kj}^{(U)}$ and $A_{ij}^{(F)}$ to be

$$\mathrm{Var}(A_{k1}^{(U)}) = \frac{v_{\epsilon_k}^{(U)}}{1-\phi_k^{2(U)}}, \quad \mathrm{Var}(A_{i1}^{(F)}) = \frac{v_{\epsilon_i}^{(F)}}{1-\phi_i^{2(F)}}$$

then the prior variances of $A_{kj}^{(U)}$ and $A_{ij}^{(F)}$ for $j = 2, \ldots, 9$ remain at these stationary values. For a proof of this see Box & Jenkins (1970).

If we denote by H_i the set of all common uncertainty factors in η_{ij} then the prior variance of η_{ij} is

$$\mathrm{Var}_0(\eta_{ij}) = V_{iM} + \sum_{q \in H_i} \frac{v_{\epsilon_q}^{(U)}}{1-\phi_q^{(U)2}} + \frac{v_{\epsilon_i}^{(F)}}{1-\phi_i^{(F)2}}$$

where $V_{iM} = \sum_{q \in H_i} \mathrm{Var}_0(M_q^{(U)}) + \mathrm{Var}_0(M_i^{(F)})$. Thus, as the model is stationary, the variance remains constant for different ages within a group. The within group covariances are

$$\mathrm{Cov}_0(\eta_{ij}, \eta_{il}) = V_{iM} + \sum_{q \in H_i} \phi_q^{(U)|l-j|} \frac{v_{\epsilon_q}^{(U)}}{1-\phi_q^{(U)2}} + \phi_i^{(F)|l-j|} \frac{v_{\epsilon_i}^{(F)}}{1-\phi_i^{(F)2}}.$$

Thus the closer together the two age groups are the higher the covariance between them. The between group covariances are

$$\mathrm{Cov}_0(\eta_{ij}, \eta_{kl}) = V_{ikM} + \sum_{q \in H_{ik}} \phi_q^{(U)|l-j|} \frac{v_{\epsilon_q}^{(U)}}{1-\phi_q^{(U)2}},$$

where $V_{ikM} = \sum_{q \in H_{ik}} \mathrm{Var}_0(M_q^{(U)})$ and $H_{ik} = H_i \cap H_k$.

Non-stationary process

We can also define a non-stationary process. This allows more flexibility in the variances and covariances between the η_{ij}'s; allowing them to increase or decrease with age within a group.

One way to achieve this is to give the ϵ's variances of

$$\begin{aligned} \mathrm{Var}(\epsilon_{kj}^{(U)}) &= h_k^{(U)2}\mathrm{Var}(\epsilon_{kj-1}^{(U)}) \\ \mathrm{Var}(\epsilon_{ij}^{(F)}) &= h_i^{(F)2}\mathrm{Var}(\epsilon_{ij-1}^{(F)}). \end{aligned}$$

We can set $A_{k1}^{(U)} = A_{i1}^{(F)} = 0$, so that all of the uncertainty associated with patients aged 40 comes through the overall uncertainty levels of the factors.

We initialise the variances of the ϵ's at the values

$$\mathrm{Var}(\epsilon_{k1}^{(U)}) = v_{\epsilon_{k1}}^{(U)} \quad \mathrm{Var}(\epsilon_{i1}^{(F)}) = v_{\epsilon_{i1}}^{(F)}.$$

In order to calculate the covariances between the θ_{ij}'s we require equations for them in terms of the covariance structure we have defined. We give the following theorem.

Theorem 4.1. *For two random variables η_{ij} and η_{kl} defined in terms of the non-stationary process above their covariance is*

$$Cov_0(\eta_{ij}, \eta_{kl}) = \sum_{q \in H} \left(Var(M_q^{(\beta)}) + g_{(q)jl}\, Var(\epsilon_{q1}^{(\beta)}) \right), \tag{4.9}$$

where H is the set of uncertainty factors common to both η_{ij} and η_{kl}, β is U or F depending on whether these are common or individual uncertainty factors and

$$g_{(q)jl} = \sum_{r=1}^{\min(j,l)-1} \phi_q^{(\beta)j+l-2(r+1)} h_q^{(\beta)2r}.$$

Proof. The quantities η_{ij} and η_{kl} take the form

$$\begin{aligned} \eta_{ij} - \mathrm{E}_0(\eta_{ij}) &= \sum_{s \in H_i} U_{sj} + F_{ij}, \\ \eta_{kl} - \mathrm{E}_0(\eta_{kl}) &= \sum_{t \in H_k} U_{tl} + F_{kl}, \end{aligned}$$

where H_i and H_k are the sets of common uncertainty factors in η_{ij} and η_{kl} respectively.

The covariance between them is

$$\begin{aligned}\mathrm{Cov}_0(\eta_{ij},\eta_{kl}) &= \mathrm{Cov}_0\left(\sum_{s\in H_i} U_{sj}+F_{ij},\sum_{t\in H_k} U_{tl}+F_{kl}\right)\\ &= \sum_{q\in H}\mathrm{Cov}_0(\beta_{qj},\beta_{ql}),\end{aligned}$$

as common uncertainty factors are independent of one another except when $s=t$ and individual uncertainty factors are independent of all other uncertainty factors.

For each $q\in H$ the uncertainty factor is defined as $\beta_{qj}=M_q^{(\beta)}+A_{qj}^{(\beta)}$ with $A_{qj}^{(\beta)}$ following an AR(1) process. Thus

$$\begin{aligned}A_{qj}^{(\beta)} &= \phi_q^{(\beta)}A_{qj-1}^{(\beta)}+\epsilon_{qj}^{(\beta)}\\ &= \phi_q^{(\beta)2}A_{qj-2}^{(\beta)}+\phi_q^{(\beta)}\epsilon_{qj-1}^{(\beta)}+\epsilon_{qj}^{(\beta)}\\ &= \sum_{v=1}^{j-1}\phi_q^{(\beta)j-v-1}\epsilon_{qv+1}^{(\beta)}+\phi_q^{(\beta)j-1}A_{q1}^{(\beta)}\\ &= \sum_{v=1}^{j-1}\phi_q^{(\beta)j-v-1}\epsilon_{qv+1}^{(\beta)},\end{aligned}$$

as $A_{q1}^{(\beta)}=0$. If we perform the same steps for $A_{ql}^{(\beta)}$ we see that

$$\mathrm{Cov}(A_{qj}^{(\beta)},A_{ql}^{(\beta)})=\mathrm{Cov}\left(\sum_{v=1}^{j-1}\phi_q^{(\beta)j-v-1}\epsilon_{qv+1}^{(\beta)},\sum_{w=1}^{l-1}\phi_q^{(\beta)l-w-1}\epsilon_{qw+1}^{(\beta)}\right).$$

Now, the ϵ's at different ages are independent of one another, and so if $j\geq l$

$$\begin{aligned}\mathrm{Cov}(A_{qj}^{(\beta)},A_{ql}^{(\beta)}) &= \sum_{r=1}^{l-1}\phi_q^{(\beta)j-r-1}\phi_q^{(\beta)l-r-1}\mathrm{Var}(\epsilon_{qr+1}^{(\beta)})\\ &= \sum_{r=1}^{l-1}\phi_q^{(\beta)j+l-2(r+1)}h_q^{(\beta)2r}\mathrm{Var}(\epsilon_{q1}^{(\beta)}),\end{aligned}$$

by recursive use of $\mathrm{Var}(\epsilon_{qr}^{(\beta)})=h_q^{(\beta)2}\mathrm{Var}(\epsilon_{qr-1}^{\beta})$. Similarly, for $l\geq j$ the covariance is of the same form but with labels l and j swapped. Thus

$$\mathrm{Cov}(A_{qj}^{(\beta)},A_{ql}^{(\beta)})=\sum_{r=1}^{\min(j,l)-1}\phi_q^{(\beta)j+l-2(r+1)}h_q^{(\beta)2r}\mathrm{Var}(\epsilon_{q1}^{(\beta)}).$$

We now return to the covariances between the common uncertainty factors.

$$\begin{aligned}\mathrm{Cov}(\beta_{qj},\beta_{ql}) &= \mathrm{Cov}(M_q^{(\beta)}+A_{qj}^{(\beta)},M_q^{(\beta)}+A_{ql}^{(\beta)})\\ &= \mathrm{Var}(M_q^{(\beta)})+\mathrm{Cov}(A_{qj}^{(\beta)},A_{ql}^{(\beta)})\\ &= \mathrm{Var}(M_q^{(\beta)})+\sum_{r=1}^{\min(j,l)-1}\phi_q^{(\beta)j+l-2(r+1)}h_q^{(\beta)2r}\mathrm{Var}(\epsilon_{q1}^{(\beta)}),\end{aligned}$$

and so

$$\begin{aligned}\mathrm{Cov}_0(\eta_{ij},\eta_{kl}) &= \sum_{q\in H}\mathrm{Cov}(\beta_{qj},\beta_{ql})\\ &= \sum_{q\in H}\left(\mathrm{Var}(M_q^{(\beta)})+g_{(q)jl}\mathrm{Var}(\epsilon_{q1}^{(\beta)})\right),\end{aligned}$$

where $g_{(q)jl}=\sum_{r=1}^{\min(j,l)-1}\phi_q^{(\beta)j+l-2(r+1)}h_q^{(\beta)2r}$. □

We can use Theorem 4.1 to calculate the covariances between η's within the same group. These are given in the following corollary.

Corollary 4.1. *For η_{ij} and η_{il} defined as above with $j\le l$, the variance of η_{ij} is*

$$Var_0(\eta_{ij})=\sum_{q\in H_i}\left(Var(M_q^{(\beta)})+c_{qj}\,Var(\epsilon_{q1}^{(\beta)})\right), \tag{4.10}$$

where $c_{qj}=\sum_{r=1}^{j-1}\phi_q^{(\beta)2(j-r-1)}h_q^{(\beta)2r}$. The covariance between η_{ij} and η_{il} is

$$Cov_0(\eta_{ij},\eta_{il})=\sum_{q\in H_i}\left(Var(M_q^{(\beta)})+d_{(q)jl}\,Var(\epsilon_{q1}^{(\beta)})\right), \tag{4.11}$$

where $d_{(q)jl}=\sum_{r=1}^{j-1}\phi_q^{(\beta)j+l-2(r+1)}h_q^{(\beta)2r}$.

Proof. (i) Equation 4.10: $\mathrm{Var}_0(\eta_{ij})=\mathrm{Cov}_0(\eta_{ij},\eta_{ij})$. By Theorem 4.1

$$\mathrm{Cov}_0(\eta_{ij},\eta_{ij})=\sum_{q\in H}\left(\mathrm{Var}(M_q^{(\beta)})+g_{(q)jj}\mathrm{Var}(\epsilon_{q1}^{(\beta)})\right).$$

Now, H is the set of uncertainty factors common to η_{ij} and η_{ij} i.e. H_i. Also,

$$g_{(q)jj}=\sum_{r=1}^{\min(j,j)-1}\phi_q^{(\beta)j+j-2r-2}h_q^{(\beta)2r}=\sum_{r=1}^{j-1}\phi_q^{(\beta)2(j-r-1)}h_q^{(\beta)2r}=c_{qj}.$$

(ii) Equation 4.11: By Theorem 4.1

$$\mathrm{Cov}_0(\eta_{ij}, \eta_{il}) = \sum_{q \in H} \left(\mathrm{Var}(M_q^{(\beta)}) + d_{(q)jl} \mathrm{Var}(\epsilon_{q1}^{(\beta)}) \right).$$

Now, η_{ij} and η_{il} share the same uncertainty factors and so $H = H_i$. Also,

$$g_{(q)jl} = \sum_{r=1}^{\min(j,l)-1} \phi_q^{(\beta)j+l-2r-2} h_q^{(\beta)2r} = \sum_{r=1}^{j-1} \phi_q^{(\beta)j+l-2(r+1)} h_q^{(\beta)2r} = d_{(q)jl},$$

as $j \leq l$.

□

Let us suppose that we believe as age increases there is more uncertainty associated with the death rates amongst patients. This belief is based on the notion that older patients are more susceptible to many different health related factors apart from smoking, for example severity of winter. Thus we shall use a non-stationary process.

Theorem 4.1 and Corollary 4.1 can be used to find all of the necessary variances and covariances for the elicitation process. For example, from the corollary we see that

$$\begin{aligned}\mathrm{Var}(\eta_{13}) = V_{1M} &+ (\phi_0^{(U)2} h_0^{(U)2} + h_0^{(U)4}) \mathrm{Var}(\epsilon_{01}^{(U)}) \\ &+ (\phi_3^{(U)2} h_3^{(U)2} + h_3^{(U)4}) \mathrm{Var}(\epsilon_{31}^{(U)}) + (\phi_1^{(F)2} h_1^{(F)2} + h_1^{(F)4}) \mathrm{Var}(\epsilon_{11}^{(F)}),\end{aligned}$$

with

$$\begin{aligned}\mathrm{Cov}(\eta_{13}, \eta_{15}) = V_{1M} &+ (\phi_0^{(U)4} h_0^{(U)2} + \phi_0^{(U)2} h_0^{(U)4}) \mathrm{Var}(\epsilon_{01}^{(U)}) \\ &+ (\phi_3^{(U)4} h_3^{(U)2} + \phi_3^{(U)2} h_3^{(U)4}) \mathrm{Var}(\epsilon_{31}^{(U)}) + (\phi_1^{(F)4} h_1^{(F)2} + \phi_1^{(F)2} h_1^{(F)4}) \mathrm{Var}(\epsilon_{11}^{(F)}),\end{aligned}$$

where $V_{1M} = \mathrm{Var}(M_0^{(U)}) + \mathrm{Var}(M_3^{(U)}) + \mathrm{Var}(M_1^{(F)})$ and

$$\begin{aligned}\mathrm{Cov}(\eta_{14}, \eta_{35}) = \mathrm{Var}(M_0^{(U)}) + \mathrm{Var}(M_3^{(U)}) &+ (\phi_0^{(U)5} h_0^{(U)2} + \phi_0^{(U)3} h_0^{(U)4} \\ &+ \phi_0^{(U)} h_0^{(U)6}) \mathrm{Var}(\epsilon_{01}^{(U)}) + (\phi_3^{(U)5} h_3^{(U)2} + \phi_3^{(U)3} h_3^{(U)4} + \phi_3^{(U)} h_3^{(U)6}) \mathrm{Var}(\epsilon_{31}^{(U)}).\end{aligned}$$

We can then use these quantities to aid us in defining a covariance structure. For the health and smoking data the chosen values of the parameters and initial variances are given in Table 4.4. These values were chosen to represent our beliefs about the strengths of the relationships between different death rates.

i	$\phi_k^{(U)}$	$\phi_i^{(F)}$	$h_k^{(U)}$	$h_i^{(F)}$	$\mathrm{Var}(\epsilon_{k1}^{(U)})$	$\mathrm{Var}(\epsilon_{i1}^{(F)})$	$\mathrm{Var}(M_k^{(U)})$	$\mathrm{Var}(M_i^{(F)})$
1	0.8	0.95	1.1	1.282	0.01	0.01	0.1	0.10
2	0.8	0.95	1.3	1.302	0.01	0.01	0.1	0.15
3	0.8	0.95	1.2	1.280	0.01	0.01	0.1	0.10
4	0.8	0.95	1.1	1.302	0.01	0.01	0.1	0.15

Table 4.4: Table of initial parameter values and variances, $k = i - 1$

Having set these values, the variance matrix for $\boldsymbol{\eta}_1 = (\eta_{11}, \ldots, \eta_{19})'$, for example, is

$$\mathrm{Var}_0(\boldsymbol{\eta}_1) = \begin{pmatrix} 0.300 & 0.300 & 0.300 & 0.300 & 0.300 & 0.300 & 0.300 & 0.300 & 0.300 \\ - & 0.341 & 0.335 & 0.330 & 0.326 & 0.323 & 0.321 & 0.318 & 0.317 \\ - & - & 0.387 & 0.376 & 0.366 & 0.359 & 0.352 & 0.347 & 0.342 \\ - & - & - & 0.446 & 0.429 & 0.415 & 0.403 & 0.393 & 0.385 \\ - & - & - & - & 0.531 & 0.507 & 0.486 & 0.469 & 0.454 \\ - & - & - & - & - & 0.658 & 0.625 & 0.596 & 0.571 \\ - & - & - & - & - & - & 0.855 & 0.808 & 0.767 \\ - & - & - & - & - & - & - & 1.167 & 1.100 \\ - & - & - & - & - & - & - & - & 1.665 \end{pmatrix}$$

Thus it can be seen that the prior correlation between η_{12} and η_{13} is 0.923 whereas the correlation between η_{12} and η_{19} is 0.420. It is felt that this represents a good balance of covariances between close and distant age groups.

Discussion

Our prior elicitation for the health and smoking example takes two stages; a marginal elicitation involving the beta distributions in order to find the prior parameters a_{ij} and b_{ij} and the specification of a prior covariance structure. Of course this second step is likely to lead to an iterative adjustment of a_{ij} and b_{ij}.

We feel, however, that such a pragmatic process is likely to lead to better assessments overall.

4.5.4 Results

We perform all updates in R following the methodology in Section 4.4. Having done this we can produce various plots.

Figure 4.4 shows the adjusted expectations of the η_{ij}'s (red) in each of the four categories. Also included in the plots are ± 2 standard deviation limits (blue) for each age

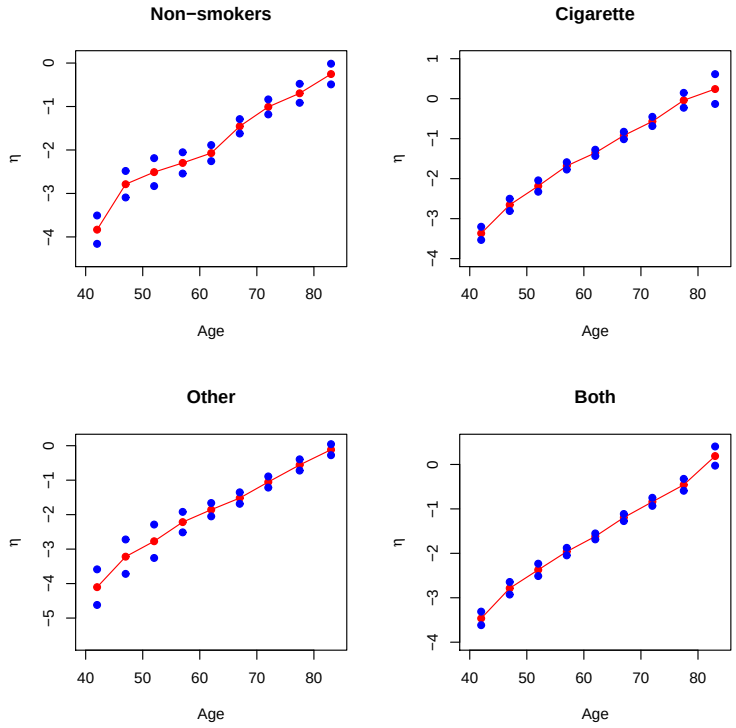

Figure 4.4: Adjusted means and confidence bounds for η_{ij} in each of the four categories of smoker.

group within each category. These were calculated as

$$\left(\mathrm{E}_{36}(\eta_{ij};\boldsymbol{x}) - 2 \times \sqrt{\mathrm{Var}_{36}(\eta_{ij};\boldsymbol{x})}, \mathrm{E}_{36}(\eta_{ij};\boldsymbol{x}) + 2 \times \sqrt{\mathrm{Var}_{36}(\eta_{ij};\boldsymbol{x})}\right).$$

From these plots we can see that we have the least uncertainty in our adjusted expectations in categories 'Cigarette' and 'Both'. This is unsurprising as these are the groups with the largest numbers of participants. In both of these categories there appears to be a fairly linear increase in the logit of the probability of death with increasing age.

In Figure 4.5 we have plotted the same quantities as in Figure 4.4 for the category Other as well as the 'data values' for this group. These were calculated as

$$\log\left(\frac{\hat{\theta}_{ij}}{1-\hat{\theta}_{ij}}\right),$$

where $\hat{\theta}_{ij} = \dfrac{x_{ij}}{n_{ij}}$ is the observed proportion of deaths in group ij. We can see from this plot that the adjusted expectations of the η_{ij}'s are very close to their observed counterparts, especially in the older age groups where we have a large quantity of data.

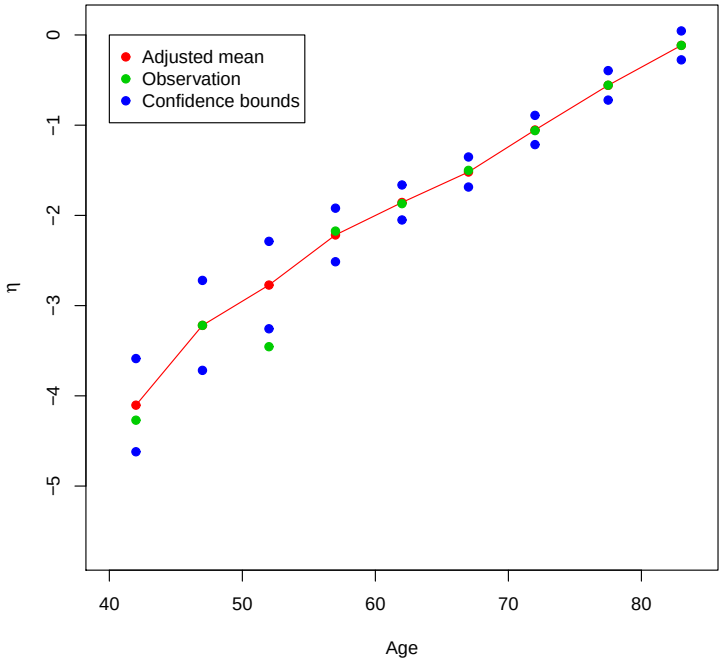

Figure 4.5: Adjusted means, ±2 standard deviation limits and observed values for η_{ij} in category Other.

At age 52 the logit of proportion of deaths is slightly lower than would be expected following the trend in the rest of the plot. This is a result of the observed value for this age category being lower than for the previous age category. The fact that the adjusted expectation is still following the general pattern of the rest of the group would appear to be a result of the high covariances we have specified between different age groups.

In Figure 4.6 we have plotted prior expectations, 'data values' and adjusted expectations for the η_{ij}'s in the category Other. This time, however, we have updated using only the data from categories 1, 2 and 4 (Non-smokers, Cigarettes and Both).

We can see that in many of the age groups the adjusted expectation of η_{ij} is closer to the corresponding data point for that age group than the prior expectation is, even though we have not included these data in the model. This would appear to indicate that the covariance structure we imposed is representative of the relationships between variables and that this model may be useful.

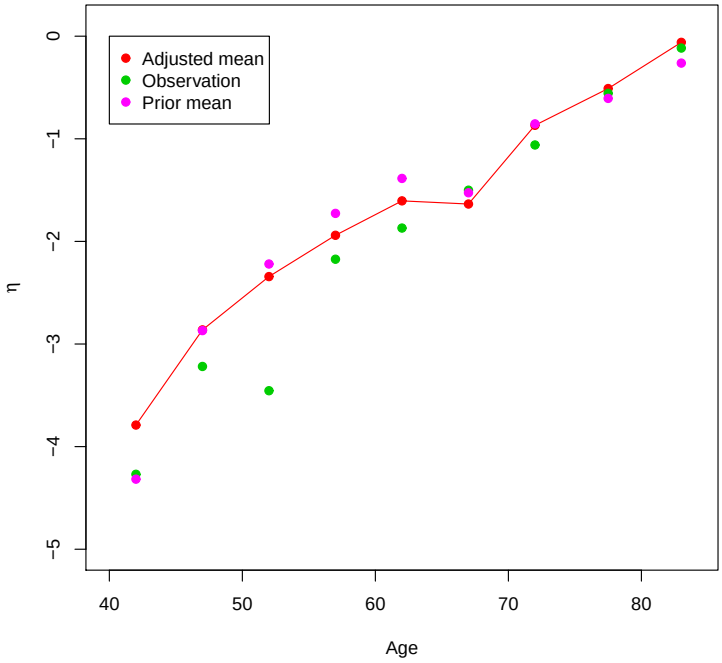

Figure 4.6: Adjusted means versus data for η_{ij} in category Other after partial update.

4.6 Conclusions

In this chapter we have investigated modelling correlated binomial probabilities in more than two dimensions. To do this we have applied Bayes linear kinematics as introduced in Chapter 3. As in that chapter we preferred to transform the binomial parameter first. The transformation we used was logit. There are good reasons for this. One is that the expectation and variance of the transformed quantity are simple to calculate. We also showed that the resulting updates shall always be commutative.

We applied our Bayes linear kinematic methodology to an example involving data on the association between health and smoking. To apply the method it was necessary to specify 36 prior expectations and $36 \times (36 + 1)/2 = 666$ variances and covariances. In terms of the marginal elicitation we preferred to 3 elicit quantiles for each of the unknown death probabilities and use these to fix the values of the parameters of the marginal beta distributions.

We then specified the covariances between the parameters by adopting a structure based on a non-stationary first-order autoregression over age for common and individual uncertainty factors between the parameters.

Chapter 5

Reliability and survival analysis

5.1 Introduction

In this chapter we consider models for count data in reliability and survival analysis. Within the sphere of reliability models we consider the analysis of failure rates and failure time distributions in the form of life tables. This work can also be found in Wilson & Farrow (2010). We utilise the methodology, developed in Chapter 3, of transforming binomial and Poisson parameters in order to perform Bayes linear kinematic updating most effectively. We consider two examples, one involving Poisson counts of failures and the other binomial counts in an analysis of failure times. In both, particular attention is paid to the elicitation of prior information and methods utilising quantiles are proposed.

The survival model we consider is a piecewise constant hazards model in which hazards for different individuals are considered proportional. Within each time period individual observations are Poisson and, if hazards are given conjugate gamma prior distributions, fully Bayesian updates can be made simply within groups. We show that these changes in belief can be propagated via Bayes linear kinematics to achieve a commutative solution. This is in contrast to a similar model in Gamerman (1991). We illustrate the approach with an example concerning Coronary Artery Bypass Graft surgery.

5.2 Bayes linear Bayes analysis for Poisson data in more than two dimensions

Let us suppose that we have Poisson counts

$$X_i \mid \theta_i \sim \text{ Poisson}(\theta_i),$$

where θ_i is the expected number of successes, $i = 1, \ldots, p$. The natural conjugate prior distribution is a gamma distribution,

$$\theta_i \sim \text{gamma}(a_i, b_i).$$

Given an observation $X_i = x_i$, the posterior distribution is gamma$(a_i + x_i,\ b_i + 1)$. More generally a known scale factor s_i, perhaps the time at risk, could be included. In this case

$$X_i \mid \theta_i \sim \text{Poisson}(s_i\theta_i),$$

and the posterior distribution is

$$\theta_i \mid x_i \sim \text{gamma}(a_i + x_i, b_i + s_i),$$

with $s_i = 1$ as a special case. The prior mean and variance are

$$\text{E}_0(\theta_i) = \frac{a_i}{b_i}, \quad \text{Var}_0(\theta_i) = \frac{a_i}{b_i^2} \tag{5.1}$$

and the posterior mean and variance are

$$\text{E}_1(\theta_i) = \frac{a_i + x_i}{b_i + s_i}, \quad \text{Var}_1(\theta_i) = \frac{a_i + x_i}{(b_i + s_i)^2}.$$

Notice that the posterior variance can be greater than the prior variance if x_i is sufficiently large.

We wish to perform Bayes linear updating. Thus it is desirable to work with variables on an unrestricted scale. In the Poisson case the transformation

$$\eta_i = \log(\theta_i)$$

is used to map from $\theta_i \in [0, \infty)$ to $\eta_i \in (-\infty, \infty)$. This is the natural link function for the Poisson distribution in generalised linear modelling. It is necessary to work with moments for both θ_i for the conjugate updates and η_i for the Bayes linear kinematic updates. The expectation and variance of θ_i are found using Equation 5.1.

In Chapter 3 we found the expectation and variance of η_i in terms of the parameters of the marginal gamma distributions associated with θ_i. They are

$$\mathrm{E}_0(\eta_i) = \psi(a_i) - \log b_i, \quad \mathrm{Var}_0(\eta_i) = \psi_1(a_i),$$

where $\psi(\cdot)$ is the digamma function and $\psi_1(\cdot)$ is the trigamma function.

Suppose that the scale factor $s_i = 1$. In generalised linear models $\log s_i$ is known as an offset. Thus here we are setting the offset equal to zero. After the conjugate updates the expectations and variances of both θ_i and η_i remain of the same form but with a_i and b_i replaced with $a_i + x_i$ and $b_i + 1$ respectively. Thus for η_i they are

$$\mathrm{E}(\eta_i \mid x_i) = \psi(a_i + x_i) - \log(b_i + 1) \tag{5.2}$$

$$\mathrm{Var}(\eta_i \mid x_i) = \psi_1(a_i + x_i). \tag{5.3}$$

Suppose that a full second order prior specification has been made for $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$ of the form $S_0(\boldsymbol{\eta}) = (\mathrm{E}_0(\boldsymbol{\eta}), \mathrm{Var}_0(\boldsymbol{\eta}))$. Observing $X_i = x_i$ leads to the Bayes linear kinematic adjusted expectation and variance for $\boldsymbol{\eta}$ of

$$\mathrm{E}_1(\boldsymbol{\eta}; x_i) = \mathrm{E}_0(\boldsymbol{\eta}) + \mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)\mathrm{Var}_0^{-1}(\eta_i)\left[\mathrm{E}(\eta_i \mid x_i) - \mathrm{E}_0(\eta_i)\right], \tag{5.4}$$

and

$$\begin{aligned}\mathrm{Var}_1(\boldsymbol{\eta}; x_i) = \mathrm{Var}_0(\boldsymbol{\eta}) &- \mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Cov}_0(\eta_i, \boldsymbol{\eta}) \\ &+ \mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Var}_0(\eta_i \mid x_i)\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Cov}_0(\eta_i, \boldsymbol{\eta}).\end{aligned} \tag{5.5}$$

which depend only on the prior specifications and fully Bayesian conjugate updates which have already been calculated. Here, for example, $\mathrm{E}_1(\boldsymbol{\eta}; x_i)$ denotes the adjusted expectation after 1 observation has been made and the observation is given after the semicolon.

Now consider whether, having observed $\boldsymbol{x} = (x_1, \ldots, x_p)'$, there is a unique commutative Bayes linear kinematic update for $\boldsymbol{\eta}$. From Goldstein & Shaw (2004) a sufficient condition for a unique commutative solution is

$$\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Var}(\eta_i \mid x_i) < 1$$

for all i. The variances are $\mathrm{Var}_0(\eta_i) = \psi_1(a_i)$ and $\mathrm{Var}(\eta_i \mid x_i) = \psi_1(a_i + x_i)$. Each x_i must be a nonnegative integer. The trigamma function is monotonically decreasing on $\mathbb{R}_+$ and $\psi_1(x) \to 0$ as $x \to \infty$ so, as long as $x_i > 0$ for each i, $\mathrm{Var}(\eta_i \mid x_i) < \mathrm{Var}_0(\eta_i)$ for all i and the uniqueness condition is met. If this is the case then the Bayes linear

kinematic unique commutative solution is

$$\mathrm{E}_p(\boldsymbol{\eta};\boldsymbol{x}) = \mathrm{Var}_p(\boldsymbol{\eta};\boldsymbol{x}) \times \left[\sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta};x_i)\mathrm{E}_1(\boldsymbol{\eta};x_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\mathrm{E}_0(\boldsymbol{\eta})\right], \quad (5.6)$$

$$\mathrm{Var}_p(\boldsymbol{\eta};\boldsymbol{x}) = \left[\sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta};x_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\right]^{-1}. \quad (5.7)$$

In fact, we can see that these equations define a commutative update even if $x_i = 0$ for some i and this case satisfies the more general conditions in Theorem 5 of Goldstein & Shaw (2004).

Suppose that we have a commutative Bayes linear kinematic solution for $\boldsymbol{\eta}$ at stage $i-1$ and it is given by $\mathrm{E}_{i-1}(\eta_i;\boldsymbol{x}_{i-1})$ and $\mathrm{Var}_{i-1}(\eta_i;\boldsymbol{x}_{i-1})$ for η_i where $\boldsymbol{x}_{i-1} = (x_1,\ldots,x_{i-1})'$. Now suppose we observe x_i. From Theorem 5 of Goldstein & Shaw (2004) we have a commutative update if

$$\mathrm{Var}^{-1}(\eta_i \mid x_i) + \mathrm{Var}_{i-1}^{-1}(\eta_i;\boldsymbol{x}_{i-1}) - \mathrm{Var}_0^{-1}(\eta_i) > 0. \quad (5.8)$$

Now, the Bayes linear kinematic commutative variance can be broken down sequentially as

$$\mathrm{Var}_{i-1}^{-1}(\eta_i;\boldsymbol{x}_{i-1}) = \mathrm{Var}_1^{-1}(\eta_i;x_{i-1}) + \mathrm{Var}_{i-2}^{-1}(\eta_i;\boldsymbol{x}_{i-2}) - \mathrm{Var}_0^{-1}(\eta_i),$$

and so Equation 5.8 becomes

$$\mathrm{Var}^{-1}(\eta_i \mid x_i) + \sum_{j=1}^{i-1} \mathrm{Var}_1^{-1}(\eta_i;x_j) - (i-1)\mathrm{Var}_0^{-1}(\eta_i) > 0. \quad (5.9)$$

Thus, if Equation 5.9 holds for all i then we have a unique commutative solution. Now consider the gamma-Poisson case. If $x_j = 0$ then $\mathrm{Var}(\eta_j \mid x_j) = \mathrm{Var}_0(\eta_j)$ and if $x_j \geq 1$ then $\mathrm{Var}(\eta_j \mid x_j) < \mathrm{Var}_0(\eta_j)$. This means that the adjusted variance of η_i, $\mathrm{Var}_1(\eta_i;x_j) \leq \mathrm{Var}_0(\eta_i)$ and so $\mathrm{Var}_1^{-1}(\eta_i;x_j) - \mathrm{Var}_0^{-1}(\eta_i) \geq 0$. Thus, from Equation 5.9, our uniqueness condition is

$$\begin{aligned}
&\mathrm{Var}^{-1}(\eta_i \mid x_i) + \textstyle\sum_{j=1}^{i-1} \mathrm{Var}_1^{-1}(\eta_i;x_j) - (i-1)\mathrm{Var}_0^{-1}(\eta_i) \\
&\qquad \geq \mathrm{Var}^{-1}(\eta_i \mid x_i) + (i-1)\mathrm{Var}_0^{-1}(\eta_i) - (i-1)\mathrm{Var}_0^{-1}(\eta_i) \\
&\qquad = \mathrm{Var}^{-1}(\eta_i \mid x_i) \\
&\qquad > 0.
\end{aligned}$$

Hence the commutative update always exists in the gamma-Poisson case.

Having found the revised expectations and variances of every η_i, the means and variances of θ_i can be found by first solving the following equations for a_i^* and b_i^*.

$$\begin{aligned} \mathrm{E}_p(\eta_i; \boldsymbol{x}) &= \psi(a_i^*) - \log(b_i^*) \\ \mathrm{Var}_p(\eta_i; \boldsymbol{x}) &= \psi_1(a_i^*). \end{aligned}$$

The revised mean and variance of θ_i are then

$$\mathrm{E}_p(\theta_i; \boldsymbol{x}) = \frac{a_i^*}{b_i^*}, \quad \mathrm{Var}_p(\theta_i; \boldsymbol{x}) = \frac{a_i^*}{b_i^{*2}}.$$

5.2.1 Comparing the gamma and log-Normal distributions

We can investigate the similarity between the log-Normal and gamma distributions as priors for Poisson parameters under our Bayes-Bayes linear-Bayes models. If the log-Normal and gamma distributions were exactly the same then our Bayes linear updates would be exactly the same as fully Bayesian updates.

If $X \sim$Po(θ) and either

$$\theta \sim \text{gamma}(a, b), \quad \theta \sim \text{log-N}(\mu, \sigma^2),$$

where μ and σ^2 are the mean and variance of $\eta = \log\theta$, the corresponding Normal distribution, then we can match the means and variances of the unrestricted parameters in each case and plot the resulting densities.

If we use the direct mean and variance approach then the mean and variance of η are

$$\begin{aligned} \mu &= \psi(a) - \log b, \\ \sigma^2 &= \psi_1(a). \end{aligned}$$

If we use the mode and curvature approach the parameters of the log-Normal distribution are

$$\begin{aligned} \mu &= \log\frac{a}{b}, \\ \sigma^2 &= \frac{1}{a}. \end{aligned}$$

We set $b = 1$ as it is simply a scaling parameter and plot the resulting densities of both distributions for $a = 2, 10, 50, 100$ in Figure 5.1 for the direct method and Figure 5.2 for the mode and curvature approach.

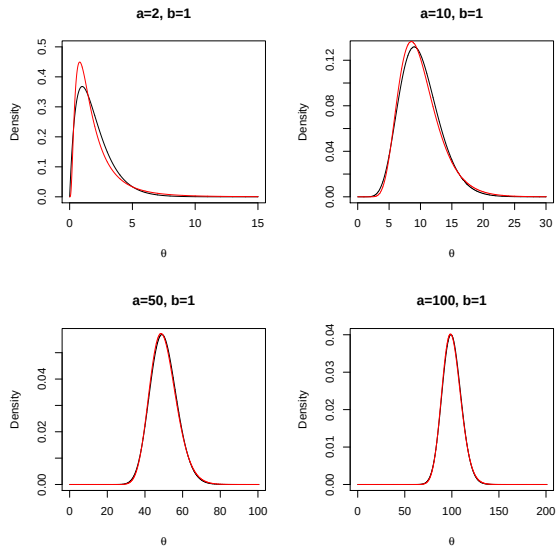

Figure 5.1: A plot comparing the gamma and log-Normal distributions using the direct method

The gamma densities are given in black and the log-Normal densities in red. We see that, in both cases, although the densities don't appear too similar for small a, as a increases the two distributions become more and more similar. It appears that it is the mode and curvature approach which most closely matches the log-Normal distribution for large a.

5.3 Bayes linear kinematics for failure rates

5.3.1 Example: failure rates of piston-rings

Data are presented in Davies & Goldsmith (1972), reproduced in Hand *et al.* (1994), on the numbers of failures of piston-rings in four steam driven compressors over a number of years. Within each compressor there are three legs: north ($i = 1$), centre ($i = 2$) and south ($i = 3$). The south leg of each compressor is adjacent to the drive.

The numbers of failures X_{ij} in each leg i of each compressor j are given in Table 5.1.

Questions of interest for these data are

1. whether the rate of piston-ring failures varies between compressors
2. whether the rate of piston-ring failures varies between legs
3. whether the pattern of the location of failures is different for different compressors.

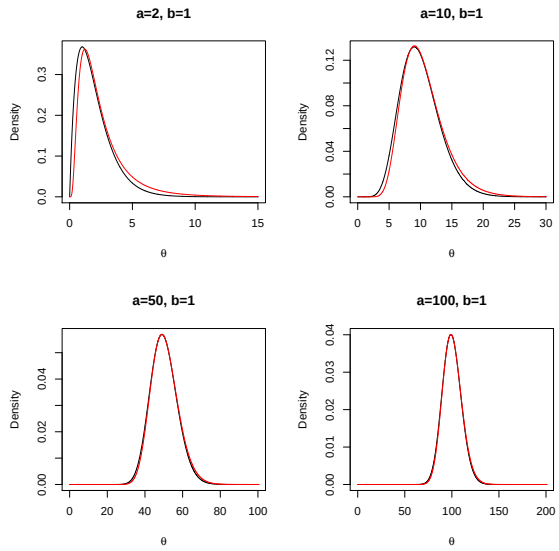

Figure 5.2: A plot comparing the gamma and log-Normal distributions using the mode and curvature approach

	Leg			
Compressor	North	Centre	South	Total
1	17	17	12	46
2	11	9	13	33
3	11	8	19	38
4	14	7	28	49
Total	53	41	72	166

Table 5.1: Piston-ring failures

We have

$$X_{ij} \mid \theta_{ij} \sim \text{Poisson}(\theta_{ij}).$$

Let $\eta_{ij} = \log(\theta_{ij})$. Giving each θ_{ij} a conjugate gamma prior distribution allows us to utilise the methodology in Section 5.2 There are twelve conjugate updates to perform, one for each element of $\boldsymbol{X} = (X_{11}, \ldots, X_{14}, X_{21}, \ldots, X_{24}, \ldots, X_{34})'$.

5.3.2 Elicitation of prior beliefs

Now let us consider the process of specification of the expert's prior beliefs. The elicitation process consists of finding the parameters a_{ij}, b_{ij} of the marginal gamma distributions and eliciting prior covariances between the η_{ij}'s.

Elicitation of prior expectations and variances

To find a_{ij} and b_{ij}, quantiles can be elicited for the gamma prior distribution of θ_{ij} and, following Garthwaite & O'Hagan (2000), the median $q_{ij}[1/2]$ and lower and upper tertiles $q_{ij}[1/3]$ and $q_{ij}[2/3]$ are chosen. To perform these elicitations, questions are put to the expert in terms of the average number of failures per unit time over a very long period.

The compressors are identically designed and are all oriented the same way. Suppose that, *a priori*, there is no reason to believe any leg of any compressor would be more prone to failures than any other. Thus the marginal elicitation process reduces to eliciting a single median $q[1/2]$ and a single lower and upper tertile $q[1/3]$ and $q[2/3]$ for the failure rate θ in any leg of any compressor. Since three elicitations ($q[1/3],\ q[1/2],\ q[2/3]$) are made to determine two parameters a, b, in general there is no exact solution.

Indeed, Pratt *et al.* (1995) propose an approximate solution. However, in the case of the gamma distribution, an exact solution can be found by considering

$$\begin{aligned} F_0(q[2/3]) = 0.67 \quad &\Rightarrow \quad q[2/3] = F_0^{-1}(0.67) \\ F_0(q[1/3]) = 0.33 \quad &\Rightarrow \quad q[1/3] = F_0^{-1}(0.33), \end{aligned}$$

where F_0 is the prior distribution function of θ and F_0^{-1} is its inverse. Whilst $q[2/3]$ and $q[1/3]$ depend on b their ratio does not. Thus a can be found from the quantity

$$\frac{q[2/3]}{q[1/3]} = \frac{F_0^{-1}(0.67)}{F_0^{-1}(0.33)}.$$

The elicited median $q[1/2]$ can then be used to find b.

An interval halving algorithm can be used to solve for a as follows;

1. Specify a range in which a is believed to lie; $a \in (a_\alpha, a_\beta)$, $a_\alpha < a_\beta$. Elicit values for $q[2/3]$ and $q[1/3]$, the upper and lower tertiles of the distribution of θ respectively.

2. Set $b = 1$ (arbitrary). Calculate

$$R_\alpha = \frac{F_{0\alpha}^{-1}(0.67)}{F_{0\alpha}^{-1}(0.33)}, \quad R_\beta = \frac{F_{0\beta}^{-1}(0.67)}{F_{0\beta}^{-1}(0.33)},$$

   the ratios of the upper tertile to the lower tertile for $\theta \sim$gamma$(a_\alpha, 1)$ and $\theta \sim$gamma$(a_\beta, 1)$ respectively.

3. Find the differences between the ratio of upper tertile to lower tertile using a_α

and a_β and its elicited value,

$$d_\alpha = R_\alpha - \frac{q[2/3]}{q[1/3]} \quad d_\beta = R_\beta - \frac{q[2/3]}{q[1/3]}$$

If $d_\alpha d_\beta > 0$ half a_α, double a_β and return to 2. If not, proceed to 4.

4. Calculate $a_1 = \frac{a_\alpha + a_\beta}{2}$ and use this to find R_1 and d_1 as above. If $d_1 d_\alpha > 0$ replace a_α with a_1. If not replace a_β with a_1.

5. Iterate through steps 2-4 until a sufficiently accurate value for a is found

To find b, given a, calculate t where

$$\Pr(\theta < t) = 0.5, \text{ for } \theta \sim \text{gamma}(a, 1).$$

Then b is such that $\frac{t}{b} = q[1/2]$ as increasing the scale parameter by b reduces the value of any quantile by b. Thus $b = \frac{t}{q[1/2]}$.

Elicitation of prior covariances

To complete the second order prior specification, covariances for $\boldsymbol{\eta}$ must be specified. This is achieved by eliciting quantities involving the θ_{ij}'s. For each pair θ_{ij} and θ_{kl} with $(i,j) \neq (k,l)$, a prior covariance $\text{Cov}_0(\eta_{ij}, \eta_{kl})$ is required. This covariance is elicited by asking the expert to suppose that the value of θ_{kl}, the population average number of piston-ring failures per unit of time over a very long period, is now known and indicating that this has left the median for θ_{ij} unchanged at $q_{ij}(1/2)$. New tertiles, $q'_{ij}(1/3)$ and $q'_{ij}(2/3)$, are then elicited for θ_{ij} having learned θ_{kl}.

From these the parameters a'_{ij}, b'_{ij} of the gamma distribution can be found as when making the marginal prior specifications. Thus

$$\theta_{ij} \mid \theta_{kl} \sim \text{gamma}(a'_{ij}, b'_{ij}).$$

If the expert judges that θ_{ij} and θ_{kl} are unrelated then $q'_{ij}(1/3) = q_{ij}(1/3)$ and $q'_{ij}(2/3) = q'_{ij}(2/3)$ i.e., the elicited tertiles would remain unchanged as nothing has been learned about θ_{ij} by learning θ_{kl}. If the expert judges that there is a relation between θ_{ij} and θ_{kl} then $q'_{ij}(1/3) > q_{ij}(1/3)$ and $q'_{ij}(2/3) < q_{ij}(2/3)$ i.e., the elicited tertiles will have moved closer together indicating a reduction in uncertainty about θ_{ij} having learned θ_{kl}. The closer together the tertiles become, the stronger the association between the two quantities.

Transforming back to η_{ij} gives

$$\mathrm{Var}(\eta_{ij} \mid \eta_{kl}) = \psi_1(a_{ij}^{'})$$

which can be combined with the prior variances to give a prior covariance via

$$\frac{\mathrm{Var}(\eta_{ij} \mid \eta_{kl})}{\mathrm{Var}_0(\eta_{ij})} = 1 - \frac{\mathrm{Cov}_0^2(\eta_{ij}, \eta_{kl})}{\mathrm{Var}_0(\eta_{ij})\mathrm{Var}_0(\eta_{kl})}.$$

Thus the modulus of the prior covariance between η_{ij} and η_{kl} is

$$|\mathrm{Cov}_0(\eta_{ij}, \eta_{kl})| = \sqrt{\mathrm{Var}_0(\eta_{kl})[\mathrm{Var}_0(\eta_{ij}) - \mathrm{Var}(\eta_{ij} \mid \eta_{kl})]}.$$

The sign of the covariance is determined by asking whether the expert's expectation for θ_{ij} would increase or decrease upon learning that θ_{kl} was greater than expected. This method is based on that which was used in the projects described in Spiropoulos (1995); Goldstein *et al.* (1993); Farrow (2003).

5.3.3 Results

Suppose that the expert settles on values of $q(1/3) = 11$ and $q(2/3) = 20$ for the lower and upper tertiles following the elicitation process. This leads to $a = 2.441$. If the expert also gives a median $q(1/2) = 15$, b is found to be 0.1411. If the four compressors are judged to be exchangable and the legs within each compressor are also regarded as exchangable (which, of course, might not be the case), the elicitation of a covariance structure can be reduced to the specification of three different covariances:

$$\mathrm{Cov}_0(\eta_{ij}, \eta_{kl}) = \begin{cases} c_1, \text{ when } i = k, j \neq l, \\ c_2, \text{ when } i \neq k, j = l, \\ c_3, \text{ when } i \neq k, j \neq l. \end{cases}$$

That is, a covariance for the same leg in different compressors, a covariance for different legs in the same compressor and a covariance for different legs in different compressors.

Table 5.2 shows an example of elicited adjusted tertiles in the above three cases and the resulting adjusted gamma parameter values and covariances and correlations of η_{ij}.

Using these prior specifications we can perform fully conjugate updates to obtain the expectations and variances in Equations 5.2 and 5.3. These updates are propagated via Bayes linear kinematics using Equations 5.4 and 5.5. A unique commutative Bayes linear kinematic solution can be found in this example as at least one piston ring failed

Case (h)	$q[1/3]'$	$q[2/3]'$	a'	b'	c_h	ρ_h
1	12	18.5	4.412	0.2722	0.356	0.704
2	11.75	18.75	3.824	0.2331	0.322	0.639
3	11.25	19.25	2.960	0.1756	0.229	0.453

Table 5.2: Elicitation of covariances c_h and correlations ρ_h in cases $h = 1, 2, 3$.

in each group, satisfying the sufficient condition. It is given in Equations 5.6 and 5.7. For this solution the adjusted values of the gamma parameters are calculated. Figure

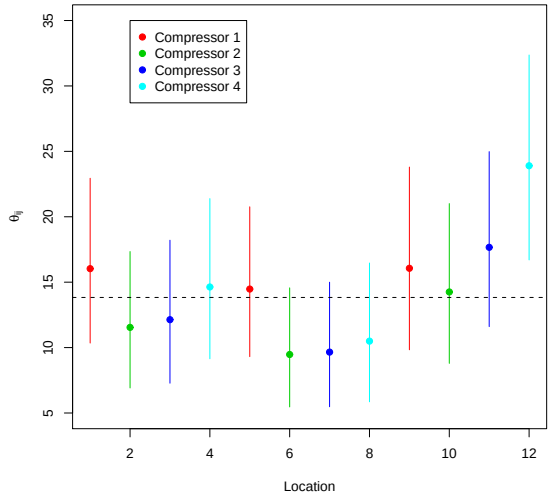

Figure 5.3: A plot of $\mathrm{E}_{12}(\theta_{ij}; \boldsymbol{X} = \boldsymbol{x})$ and 95% symmetric credible intervals

5.3 shows the adjusted expectations of the θ_{ij}'s and adjusted 95% symmetric credible intervals for each of the 12 legs. The dashed line on the plot is the observed mean number of piston-ring failures in the time period, 166/12.

The first four locations correspond to the north leg, the next four to the centre leg and the final four to the south leg. A full list of locations along with posterior moments are given in Table 5.3. It appears that location 12, the south leg of compressor 4, has an unusually high rate of piston-ring failures.

In this example, a unique commutative Bayes linear kinematic adjustment also exists if the transformation is not used as the variance of each θ_i decreases on observation of the number of failures in that group. The bracketed figures in Table 5.3 show the results. The prior specification was derived from the same elicited tertiles as in the analysis with the transformation. The results are similar but generally a little lower. It seems that the effect of the observations which are less than the prior mean may be

Location	Compressor and leg		Posterior mean		95% interval			
1	1	North	16.033	(15.541)	10.347	(9.809)	22.944	(22.569)
2	2	North	11.544	(11.537)	6.914	(6.658)	17.342	(17.733)
3	3	North	12.133	(12.007)	7.278	(7.094)	18.208	(18.191)
4	4	North	14.632	(14.349)	9.145	(8.964)	21.389	(20.979)
5	1	Centre	14.474	(13.997)	9.309	(8.365)	20.761	(21.055)
6	2	Centre	9.472	(9.347)	5.464	(4.883)	14.566	(15.236)
7	3	Centre	9.655	(9.246)	5.477	(4.918)	14.997	(14.919)
8	4	Centre	10.494	(9.390)	5.854	(5.186)	16.466	(14.821)
9	1	South	16.062	(14.961)	9.831	(9.745)	23.798	(21.278)
10	2	South	14.252	(14.072)	8.793	(8.806)	21.007	(20.553)
11	3	South	17.664	(17.082)	11.603	(11.111)	24.979	(24.315)
12	4	South	23.898	(20.927)	16.701	(14.184)	32.365	(28.961)

Table 5.3: $\mathrm{E}_{12}(\theta_{ij}; \boldsymbol{x})$ and 95% symmetric credible intervals for the 12 locations. The figures in brackets refer to an analysis without using transformations.

greater when no tansformation is used. While such an analysis without transformations is possible in this example, in general it is not.

In Figure 5.4 we have plotted the prior marginal gamma densities for all of the θ_{ij}'s in red along with the corresponding marginal densities of $\theta_{ij} \mid x_{ij}$ in blue and the posterior marginal densities of $\theta_{ij}; \boldsymbol{x}$ in green. The four north legs are in the top row, the four centre legs in the middle row and the four south legs in the bottom row of the plot.

The reduction in uncertainty when observing data from within the group is clear from the plot, as indicated in the reduced spread in the blue curve from the red curve. A larger reduction in uncertainty appears to occur when incorporating the data from the other groups emphasised by the reduction in spread between the blue curve to the green curve for most of the 12 groups.

5.4 Survival analysis for survival times and failure times

Survival analysis is concerned with modelling the amount of time taken until some event occurs. In terms of the applications in this chapter this corresponds to the amount of time until an item fails from some startpoint for that item or the amount of time after surgery until a certain symptom is observed. The time taken until the event occurs is referred to as the failure time or survival time respectively.

The failure/survival time t of a component can be regarded as the value of a random

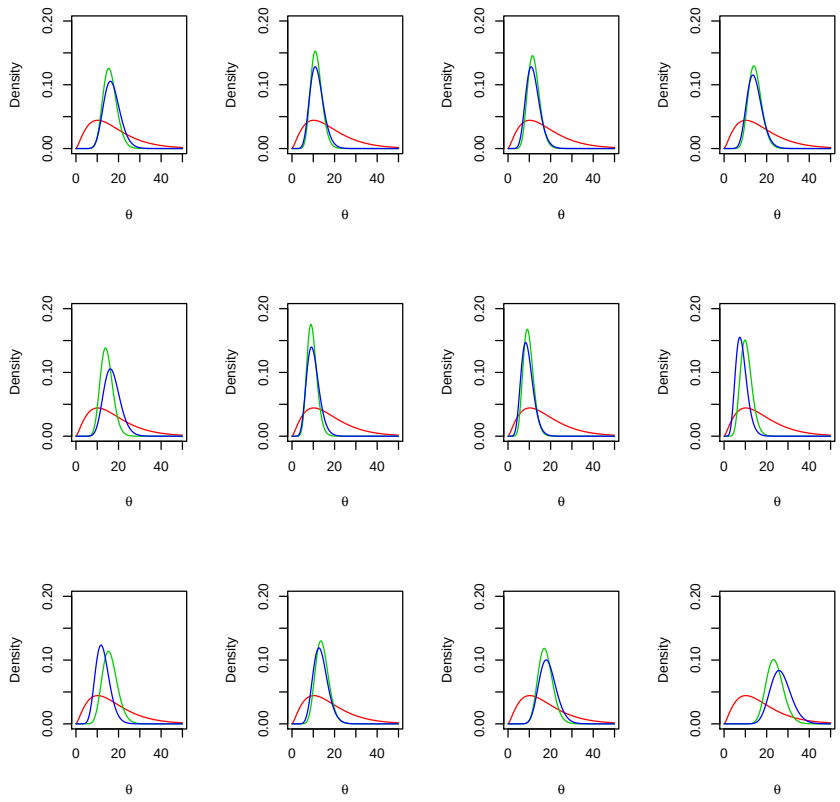

Figure 5.4: Marginal gamma distributions at different stages of the update

variable T. Associated with T is the reliability

$$R(t) = \Pr(T \geq t)$$

which is the probability that an item has not failed by time t. In survival analysis this is known as the survivor function $S(t)$. Also associated with T is a hazard function $h(t)$ which is the instantaneous rate of failure at t or the following limit;

$$h(t) = \lim_{\delta t \to 0} \frac{\Pr(\text{fail in } [t, t+\delta t) \mid \text{not failed by } t)}{\delta t}.$$

Censoring occurs when the exact failure/survival time is not known for some reason. Right censoring is the most common type in which all that is known is that the failure/survival time $t > c$ for some value c. Right censoring occurs when an item has not failed or a patient not suffered the onset of symptoms by the end of the study. Censoring can be both informative and non-informative. Non-informative censoring occurs when the failure/survival time T is independent of the mechanism which causes

an observation to be censored at c.

Information on Bayesian survival analysis can be found in Klein & Moeschberger (1997) and Ibrahim et al (2001).

5.5 Bayes linear kinematics for failure time distributions

5.5.1 Applying Bayes linear kinematics to life table data

Initially we consider the case where failure times are grouped into intervals as in a life table.

Suppose that time is split into intervals so that the i'th interval is $[\tau_i, \tau_{i+1})$ for $i = 0, \ldots, p$ with $\tau_0 = 0$ and $\tau_p = \infty$. Represent the number of failures of items in interval i by x_i and the number of items which have not failed by the start of interval i by N_i. Suppose initially that there is no censoring so an interval is recorded for the failure time of every item. Then

$$N_i = N_{i-1} - x_{i-1}.$$

The number of failures of items in each interval follows a binomial distribution

$$X_i \mid \theta_i \sim \text{bin}(N_i, \theta_i),$$

where θ_i is the unknown population probability that an item fails in the i'th interval given that it has not failed by time τ_i. Each $\theta_i \in (0, 1)$ and so can be given a marginal beta distribution, $\theta_i \sim$beta(a_i, b_i). The beta distribution is conjugate to the binomial distribution and so observation of x_i failures in interval i leads to a within interval update of

$$\theta_i \mid X_i \sim \text{beta}(A_i, B_i),$$

where $A_i = a_i + x_i$ and $B_i = b_i + N_i - x_i$. The prior expectation and variance of θ_i are given by the standard formulae for the beta distribution;

$$\text{E}_0(\theta_i) = \frac{a_i}{a_i + b_i} \quad \text{Var}_0(\theta_i) = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)},$$

with the posterior counterparts $\text{E}(\theta_i \mid x_i)$ and $\text{Var}(\theta_i \mid x_i)$ the same but using A_i and B_i. The next step is to transform to an unrestricted scale so that Bayes linear kinematic updating is most effective.

We have seen that in the beta-binomial case many different transformations are possible. For the failure-time application the complementary log-log transformation is chosen as it is more convenient for computation of the reliability $R(t) = \Pr(T \geq t)$,

where T is a lifetime. Hence

$$\mu_i = g(\theta_i) = \log[-\log(1-\theta_i)]. \tag{5.10}$$

However, as we have seen, the moments of μ_i are generally not straightforward so we return to our "guide relationship" where we take a related quantity η_i, which has mean and variance found from the mode of μ_i and the curvature at the mode of the log density of μ_i.

The prior mean and variance of η_i are

$$\mathrm{E}_0(\eta_i) = m_i, \quad \mathrm{Var}_0(\eta_i) = -\left[\frac{d^2 l_i(\mu_i)}{d\mu_i^2}\right]^{-1}_{m_i},$$

where $l_i(\mu_i)$ is the log-density of μ_i and m_i is the mode of μ_i or the solution of

$$\left(\frac{dl_i(\mu_i)}{d\mu_i}\right)_{m_i} = 1 - e^{m_i} + \left[\frac{(a_i-1)}{\theta_{m,i}} - \frac{(b_i-1)}{1-\theta_{m,i}}\right] e^{m_i}\exp[-e^{m_i}] = 0, \tag{5.11}$$

where $\theta_{m,i} = 1 - \exp[-e^{m_i}]$. The required second derivative is

$$\begin{aligned}\frac{d^2 l_i(\mu_i)}{d\mu_i^2} &= -e^{\mu_i} - \left[\frac{(a_i-1)}{\theta_i^2} + \frac{(b_i-1)}{(1-\theta_i)^2}\right] e^{2\mu_i}\exp[-2e^{\mu_i}] \\ &\quad + \left[\frac{(a_i-1)}{\theta_i} - \frac{(b_i-1)}{1-\theta_i}\right] e^{\mu_i}(1-e^{\mu_i})\exp[-e^{\mu_i}].\end{aligned} \tag{5.12}$$

This solution was derived in Section 3.9.

Having made the conjugate updates, the same procedure can be applied but using $A_i = a_i + x_i$ and $B_i = b_i + N_i - x_i$ in place of a_i and b_i in the density and subsequent derivatives. Defining a Bayes linear structure for $\eta_1, \ldots, \eta_p$, i.e., specification of $\mathrm{Cov}_0(\boldsymbol{\eta})$, allows the updates to be propagated to the other quantities in $\boldsymbol{\eta}$ via Equations 4.2 and 4.3. Note that, once an adjusted mean and precision for η_i are found, Equations 5.11 and 5.12 provide simultaneous linear equations in a_i^* and b_i^*, the new values of a_i and b_i, which are easily solved.

From Equation 4.4 a sufficient condition for a unique commutative solution to the problem using Bayes linear kinematics is

$$\mathrm{Var}_0^{-1}(\eta_i)\mathrm{Var}(\eta_i \mid x_i) < 1, \ \forall i.$$

Referring back to our numerical investigation of this condition given in Section 3.9 we conclude that, at least over a large range of parameter values, this condition shall

always hold. When a unique commutative solution does exist it is given by

$$\begin{aligned} \mathrm{E}_p(\boldsymbol{\eta};\boldsymbol{x}) &= \mathrm{Var}_p(\boldsymbol{\eta};\boldsymbol{x})\left[\sum_{i=1}^{p}\mathrm{Var}_1^{-1}(\boldsymbol{\eta};x_i)\mathrm{E}_1(\boldsymbol{\eta};x_i)-(p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\mathrm{E}_0(\boldsymbol{\eta})\right], \\ \mathrm{Var}_p(\boldsymbol{\eta};\boldsymbol{x}) &= \left[\sum_{i=1}^{p}\mathrm{Var}_1^{-1}(\boldsymbol{\eta};x_i)-(p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\right]^{-1} \end{aligned}$$

The aim is to make inference for the reliability function. This can be expressed at each of the interval boundaries in terms of the conditional probabilities of failure up to that interval. That is

$$\begin{aligned} R(\tau_i) &= \Pr(T \geq \tau_i) \\ &= \prod_{j=1}^{i}(1-\theta_j). \end{aligned}$$

As the reliability is a product of $i-1$ correlated terms, means and variances for it are not straightforward to calculate. The log of the reliability, however, is given by the sum of $i-1$ correlated terms, in particular

$$\log(R(\tau_i)) = \sum_{j=1}^{i}\log(1-\theta_j).$$

Using the complementary log-log link function gives

$$\log[R(\tau_i)] = -\sum_{j=1}^{i}\exp(\mu_j).$$

To see what might reasonably be concluded about this quantity, the guide relationship (5.10) can again be considered. After observing data it might be reasonable to suppose that the result of such a nonlinear transformation is approximated by what happens if $\boldsymbol{\eta}$ has a multivariate normal distribution and $\log(1-\theta_i) = -\exp(\eta_i)$. Thus

$$\boldsymbol{\eta} \mid \boldsymbol{x} \sim \mathrm{MVN}_p(\boldsymbol{M}, V),$$

where $\boldsymbol{M} = \mathrm{E}(\boldsymbol{\eta} \mid \boldsymbol{x})$ and $V = \mathrm{Var}(\boldsymbol{\eta} \mid \boldsymbol{x})$. Now $\exp(\boldsymbol{\mu}) \mid \boldsymbol{x}$ can be regarded as having approximately the moments of a multivariate lognormal distribution.

If $\mathrm{E}(\eta_i) = M_i$, $\mathrm{Var}(\eta_i) = V_{ii}$ and $\mathrm{Cov}(\eta_i, \eta_j) = V_{ij}$ then, following this guideline and writing $w_i = -\log[1-\theta_i] = \exp(\eta_i)$,

$$\mathrm{E}(w_i) \approx e^{M_i + V_{ii}/2}, \quad \mathrm{Var}(w_i) \approx e^{2M_i+V_{ii}}(e^{V_{ii}}-1),$$

$$\mathrm{Cov}(w_i, w_j) \approx e^{M_i+M_j+(V_{ii}+V_{jj})/2}(e^{V_{ii}+V_{jj}+2V_{ij}} - 1).$$

Returning to the log reliability the posterior expectation can now be found at each τ_i as

$$\begin{aligned} \mathrm{E}[\log(R(\tau_i))] &= -\sum_{j=1}^{i} w_j \\ &\approx -\sum_{j=1}^{i} e^{M_j+V_{jj}/2} \end{aligned} \tag{5.13}$$

and the posterior variances can also be calculated as

$$\begin{aligned} \mathrm{Var}(\log[R(\tau_i)]) &= \sum_{j=1}^{i}\sum_{k=1}^{i} \mathrm{Cov}(w_j, w_k) \\ &\approx \sum_{j=1}^{i}\sum_{k=1}^{i} e^{M_j+M_{jk}+(V_{jj}+V_{kk})/2}(e^{V_{jj}+V_{kk}+2V_{jk}} - 1). \end{aligned} \tag{5.14}$$

5.5.2 Right Censoring

If an observation is right censored during interval i, that is at time t with $\tau_i < t < \tau_{i+1}$, then its contribution to the likelihood is

$$\prod_{j=0}^{i}(1-\theta_j)^{x_j},$$

where $x_j = 1$ for $j < i$ and

$$x_i = \frac{t-\tau_i}{\tau_{i+1}-\tau_i}.$$

5.5.3 Example: Centrifuge cloths

Data are given in Lawless (1982) on the failure times of sugar centrifuge cloths. In all there are 229 cloths and all fail within 78 weeks. There is no censoring in the data. The data are presented in Table 5.4.

As for the piston-rings example, the elicitation process contains two stages: elicitation of the median and tertiles for the marginal beta distributions and elicitation of a coherent covariance structure for $\boldsymbol{\eta}$. The marginal elicitation process is very similar to that for the piston-rings example from earlier in the chapter and the smoking and health example from Section 4.5. The parameters a_i, b_i of the marginal beta distributions are found from the elicited median and tertiles.

i	Weeks	N_i	x_i	i	Weeks	N_i	x_i	i	Weeks	N_i	x_i
1	[0,2)	229	24	14	[26,28)	34	4	27	[52,54)	2	0
2	[2,4)	205	36	15	[28,30)	30	1	28	[54,56)	2	0
3	[4,6)	169	27	16	[30,32)	29	4	29	[56,58)	2	1
4	[6,8)	142	23	17	[32,34)	25	4	30	[58,60)	1	0
5	[8,10)	119	15	18	[34,36)	21	5	31	[60,62)	1	0
6	[10,12)	104	9	19	[36,38)	16	2	32	[62,64)	1	0
7	[12,14)	95	12	20	[38,40)	14	2	33	[64,66)	1	0
8	[14,16)	83	11	21	[40,42)	12	2	34	[66,68)	1	0
9	[16,18)	72	13	22	[42,44)	10	2	35	[68,70)	1	0
10	[18,20)	59	4	23	[44,46)	8	2	36	[70,72)	1	0
11	[20,22)	55	12	24	[46,48)	6	0	37	[72,74)	1	0
12	[22,24)	43	5	25	[48,50)	6	0	38	[74,76)	1	0
13	[24,26)	38	4	26	[50,52)	6	4	39	[76,78)	1	1

Table 5.4: The failure times of centrifuge cloths

The prior values for the a_i's and the b_i's resulting from the elicitation process are given in Table 5.5.

i	a_i	b_i	i	a_i	b_i	i	a_i	b_i
1	2.206	23.530	14	2.208	11.797	27	2.217	5.843
2	2.206	22.356	15	2.209	11.160	28	2.218	5.664
3	2.206	21.229	16	2.209	10.552	29	2.219	5.485
4	2.206	20.152	17	2.210	9.943	30	2.220	5.143
5	2.207	19.128	18	2.210	9.422	31	2.220	5.101
6	2.207	18.153	19	2.211	8.894	32	2.220	5.056
7	2.207	17.214	20	2.211	8.395	33	2.220	5.011
8	2.207	16.329	21	2.212	7.916	34	2.221	4.996
9	2.207	15.481	22	2.213	7.459	35	2.221	4.951
10	2.207	14.671	23	2.214	7.025	36	4.901	4.901
11	2.207	13.901	24	2.215	6.613	37	2.221	4.888
12	2.208	13.167	25	2.215	6.415	38	2.221	4.836
13	2.208	12.463	26	2.216	6.218	39	2.222	4.817

Table 5.5: Prior marginal parameter values for centrifuge cloths

Covariances between different elements of $\boldsymbol{\eta}$ can be elicited using a method similar to that used in the piston-rings example. The expert can be asked to imagine knowing the value of θ_j, from a very large experiment, and provide revised tertiles for θ_i given that the "true" value of θ_j was found to be equal to its prior median. As in the Poisson case, this leads to a calculation of the reduction in variance of η_i given knowledge of η_j and hence to the covariance of η_i and η_j.

However, with a large number of intervals it may be unappealing to consider all of the covariances individually. It may also be difficult to avoid accidental incoherence in the resulting covariance matrix. In any case it may well give more satisfactory results to adopt a more structured approach. Therefore ideas from Farrow (2003) are used to give $\mathrm{Var}_0(\boldsymbol{\eta})$ a more structured form.

For example, bearing in mind the ordering of the time intervals, uncertainties about $\boldsymbol{\eta}$ might well be represented by a stationary process. Let $F_i = \eta_i - \mathrm{E}_0(\eta_i)$ so that F_i is a zero expectation quantity which depends on time. Then $F_1, \ldots, F_p$ can be linked via a stationary process such as a first order autoregression, in which case

$$F_i = \phi F_{i-1} + \varepsilon_i, \ (i = 2, \ldots, p)$$

where $\phi < 1$, $\mathrm{E}(\varepsilon_i) = 0$, $\mathrm{E}(\varepsilon_i^2) = v_\varepsilon$ and $\mathrm{E}(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$. For stationarity the initial variance of F_1 is set at the stationary value

$$\mathrm{Var}_0(F_1) = \frac{v_\varepsilon}{1 - \phi^2} = v_F.$$

The covariances between $\eta_1, \ldots, \eta_p$ are now given by $\mathrm{Cov}_0(\eta_i, \eta_j) = \phi^{|j-i|} v_F$. Thus covariances are weaker for intervals which are further apart. If a small number of covariances are elicited directly, the parameters can then be adjusted until the expert is happy with the result. Note that using a stationary process in this way implies that all of the variances of $\eta_1, \ldots, \eta_p$ are equal and this is likely to require a process of iterative adjustment of the assessed values of a_i and b_i. It is felt, however, that such a process is likely to lead to better prior assessments overall.

For the example the values $v_F = 0.453$, $\phi = 0.97$ were adopted and therefore $v_\varepsilon = 0.0268$.

The conjugate updates take place using $A_i = a_i + x_i$ and $B_i = b_i + N_i - x_i$ in place of a_i and b_i in Equations 5.11 and 5.12 to calculate $\mathrm{E}(\eta_i \mid x_i)$ and $\mathrm{Var}(\eta_i \mid x_i)$ respectively. These are then used in Equations 4.2 and 4.3 to calculate the Bayes linear kinematic update for $\boldsymbol{\eta}$ at each stage: $\mathrm{E}_1(\boldsymbol{\eta};\ x_i)$ and $\mathrm{Var}_1(\boldsymbol{\eta};\ x_i)$. The unique commutative Bayes linear kinematic solution is then given by

$$\begin{aligned}
\mathrm{Var}_{(39)}(\boldsymbol{\eta}) &= \left(\sum_{i=1}^{39} \mathrm{Var}_1^{-1}(\boldsymbol{\eta};\ x_i) - 38\mathrm{Var}_0^{-1}(\boldsymbol{\eta}) \right)^{-1} \\
\mathrm{E}_{(39)}(\boldsymbol{\eta}) &= \mathrm{Var}_{(39)}(\boldsymbol{\eta}) \left(\sum_{i=1}^{39} \mathrm{Var}^{-1}(\boldsymbol{\eta};\ x_i) \mathrm{E}_1(\boldsymbol{\eta};\ d_i) - 38\mathrm{Var}_0^{-1}(\boldsymbol{\eta}) \mathrm{E}_0(\boldsymbol{\eta}) \right).
\end{aligned}$$

Having performed the updates, posterior parameter values are found and are given in

Table 5.6. It is clear that there has been a significant reduction in uncertainty upon observation of the data.

i	a_i^*	b_i^*	i	a_i^*	b_i^*	i	a_i^*	b_i^*
1	46.802	329.558	14	23.947	158.264	27	9.844	31.960
2	70.360	407.096	15	21.846	143.812	28	8.699	28.626
3	66.560	379.147	16	22.658	137.701	29	7.942	25.952
4	59.693	350.579	17	22.745	126.533	30	7.059	22.748
5	49.318	319.843	18	22.128	113.915	31	6.415	21.351
6	42.224	293.478	19	19.589	98.948	32	5.932	20.151
7	42.425	283.729	20	18.010	87.220	33	5.560	19.388
8	41.357	263.071	21	16.769	76.599	34	5.274	18.186
9	40.274	241.228	22	15.552	66.492	35	5.050	17.203
10	33.790	208.987	23	14.168	56.826	36	4.877	16.218
11	35.298	203.808	24	12.517	47.813	37	4.744	15.342
12	29.492	183.741	25	11.770	42.383	38	4.647	14.339
13	26.092	170.223	26	11.657	37.861	39	4.581	13.416

Table 5.6: Posterior marginal parameter values for centrifuge cloths

We can plot the prior and posterior means of the η_i's as in Figure 5.5. The posterior means are in red, the prior means are blue and the green lines represent ± 2 standard deviation intervals.

We see that our prior means may have been a bit low early on and a bit high in later intervals. There is more uncertainty at later times due to the smaller number of observations.

From the posterior means and variances we can calculate the posterior means and variances of the log reliability at each of the interval boundaries as in Equations 5.14 and 5.15. The posterior means and a credible region for the log reliability are plotted in Figure 5.6.

In order to calculate the credible region we initially transformed using logs. If we set $Z_i = -\log[R(\tau_i)]$ then the expectation and variance of $\log Z_i$ are approximately

$$\begin{aligned} m_i &= \mathrm{E}[\log Z_i] \approx \log(\mathrm{E}[Z_i]), \\ v_i &= \mathrm{Var}(\log Z_i) \approx \frac{1}{\mathrm{E}[Z_i]^2}\mathrm{Var}(Z_i). \end{aligned}$$

A sensible interval where the log reliability may lie is then

$$\left(-\exp\{m_i + 2\sqrt{v_i}\}, -\exp\{m_i - 2\sqrt{v_i}\}\right).$$

We see that the log-reliability is monotonically decreasing as it should be. The variance

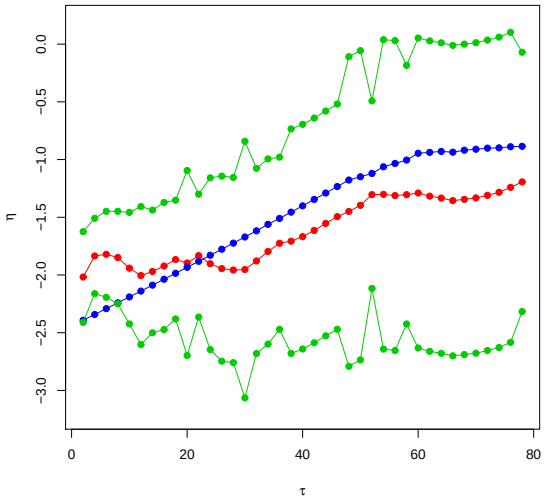

Figure 5.5: A plot of prior and posterior means and ± 2 standard deviation intervals for each η_i

is clearly increasing with increasing time. There are far fewer centrifuge cloths which have not failed by the start of later intervals and so fewer data within each interval. This could perhaps explain this increase in variance. We are also accumulating variance as we go along since this is a cumulative sum.

5.6 A dynamic Bayes linear piecewise constant hazards model for survival analysis

5.6.1 Piecewise constant hazards model

Suppose we have individuals $i = 1, \ldots, p$ and individual i has covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iq})$. Associated with each individual is a hazard function $h_i(t)$ for the random variable T at time t. If we assume a proportional hazards model (Cox, 1972) then the hazard functions of individuals are related via

$$h_i(t) = \phi_i h_0(t),$$

where ϕ_i is a constant with respect to time and $h_0(t)$ is the baseline hazard function. We can relate an individual's hazard function to their covariates by setting

$$\phi_i = \exp(\boldsymbol{x}_i' \boldsymbol{\beta}), \tag{5.15}$$

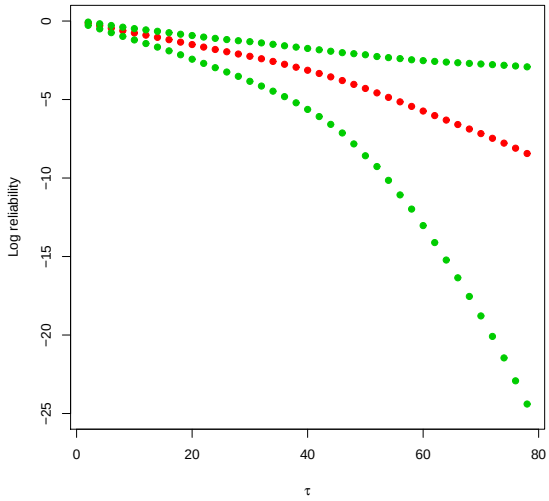

Figure 5.6: A plot of the posterior mean log-reliability as well as posterior credible regions for the centrifuge cloth example

for some parameter vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)$. In the case of no censoring we observe the values of the covariates and the death times $t_1, \ldots, t_p$ (which may not be deaths but time until some event) of each of the individuals.

In the above setup we assumed that the values of the parameters $\boldsymbol{\beta}$ remained constant over time. That is, the effect of the covariates for an individual remained constant over time. In reality this may not be the case. Therefore we shall instead consider a dynamic model

$$\phi_i(t) = \exp(\boldsymbol{x}_i^{'}\boldsymbol{\beta}(t)), \tag{5.16}$$

so that we can model changes in the effects of the covariates over time. The static model in Equation 5.15 is just a special case of this more general model.

The piecewise constant hazards model (Ibrahim *et al.*, 2001) comes about by taking some fixed time points $\tau_0, \tau_1, \ldots, \tau_r$ such that $\tau_0 = 0$ and τ_r is greater than the largest death time. This splits time up into intervals. The j'th interval is defined as $I_j = [\tau_{j-1}, \tau_j)$. Then, for $\tau_{j-1} \leq t < \tau_j$, the baseline hazard is

$$h_0(t) = \lambda_{0j},$$

and the hazard function for individual i is

$$h_i(t) = \lambda_{ij} = \phi_{ij}\lambda_{0j}.$$

That is, the hazard for each individual remains constant through each of the time intervals. The integrated hazard is then

$$\begin{aligned} H_i(t) &= \int_0^t h_i(u)du \\ &= \sum_{k=1}^{j-1} \lambda_{ik}(\tau_k - \tau_{k-1}) + \lambda_{ij}(t - \tau_{j-1}), \end{aligned}$$

From this we can calculate the survival function for individual i,

$$\begin{aligned} S_i(t) &= \exp\{-H_i(t)\} \\ &= \exp\left\{ -\left[\sum_{k=1}^{j-1} \lambda_{ik}(\tau_k - \tau_{k-1}) + \lambda_{ij}(t - \tau_{j-1}) \right] \right\}. \end{aligned}$$

We can use the hazard function and the survival function in order to find the probability density function of t.

$$\begin{aligned} f_i(t) &= h_i(t)S_i(t) \\ &= \lambda_{ij} \exp\left\{ -\left[\sum_{k=1}^{j-1} \lambda_{ik}(\tau_k - \tau_{k-1}) + \lambda_{ij}(t - \tau_{j-1}) \right] \right\} \\ &= \prod_{k=1}^{j-1} \exp\{-\lambda_{ik}(\tau_k - \tau_{k-1})\} \times \lambda_{ij} \exp\{-\lambda_{ij}(t - \tau_{j-1})\}. \end{aligned}$$

If we condition on our random variable $T \geq \tau_j$ then we obtain the conditional survival function and conditional probability density function for individual i at time t. These are

$$f_i(t \mid T \geq \tau_{j-1}) = \lambda_{ij} \exp\{-\lambda_{ij}(t - \tau_{j-1})\}, \tag{5.17}$$

and

$$S_i(t \mid T \geq \tau_{j-1}) = \exp\{-\lambda_{ij}(t - \tau_{j-1})\}. \tag{5.18}$$

Thus the conditional density takes the form of a shifted exponential distribution.

5.6.2 The Gamerman model

Gamerman (1991) proposed a dynamic piecewise constant hazards model for survival data of the form given above. This was not the first time such a model had been suggested (Kalbfleisch, 1978) but Gamerman's paper considered dependence between the hazard increments. It was based very closely on the Dynamic Generalized Linear Models of West *et al.* (1985).

His analysis took the form of an observational model with three elements;

- a conjugate prior,
- an evolving system vector, as in a dynamic linear model. The joint distribution of the system vector was not specified at times $\tau_0, \ldots, \tau_r$, just the first and second order moments,
- and a guide relationship between the parameters of the conjugate prior and the corresponding elements of the system vector.

Due to the partial specification Bayes linear updating was used to update the system vector.

More specifically, the parameters in the linear predictor didn't take a full distributional form but were given instead a second-order specification, i.e., a mean vector $\boldsymbol{m}_j$ and variance matrix C_j for each interval j. The covariances between intervals were specified using the system evolution as in a dynamic linear model.

The evolution step then used a system evolution matrix and innovation to move time on a step and give a prior mean vector $\boldsymbol{a}_{j+1}$ and variance matrix P_{j+1} for the next interval.

Gamerman utilised a guide relationship in which a quantity, η_{ij}, was introduced for each constant hazard λ_{ij}. The idea of the guide relationship was that the η_{ij} were updated within an interval and these changes in belief were propagated, through the guide relationship

$$\eta_{ij} \approx \log(\lambda_{ij}) = \boldsymbol{x}_i^{'}\boldsymbol{\beta}_j$$

to $\log(\lambda_{ij})$ and hence the parameter vector $\boldsymbol{\beta}_j$.The guide relationship, here, is denoted by $\approx$.

The updating of each η_{ij} within a time interval, was achieved via a mixture of fully Bayesian and Bayes linear methods. Each λ_{ij} was given a gamma prior distribution which was conjugate to the conditional likelihood for that individual in that interval. Thus updating of individual η_{ij}'s was straightforward. These changes in belief were then propagated through to the parameter vector $\boldsymbol{\beta}_j$ individually using Bayes linear methods. This gave the posterior mean vector $\boldsymbol{m}_j$ and variance matrix C_j within an interval.

Gamerman found, however, that the final answer depended on the order in which data were included. 'The dependence on the order that the observations are processed is of concern...The results, however, do not differ by much'. The commutative updates of Bayes linear kinematics offer a solution to this problem.

Another important point about the Gamerman model is that naturally it is a filtering procedure, with the parameter vector $\boldsymbol{\beta}_j$ only updated by information up to and including interval j. Gamerman, however, proposed a smoothing step at the end of the analysis.

5.6.3 System evolution

We are going to assume a dynamic model similar to that developed by Gamerman (1991) based upon the dynamic generalised linear models of West *et al.* (1985). In fact the relationships we exploit to make prior specifications for our parameter vector shall take the form of the system evolution in each of the above papers.

That is, we represent the parameter vector in interval I_j in terms of that in the previous interval. Specifically,

$$\boldsymbol{\beta}_j = G_j(b_j)\boldsymbol{\beta}_{j-1} + \boldsymbol{\epsilon}_j,$$

where $G_j(b_j)$ is the system evolution matrix for I_j, b_j is the length of I_j and $\boldsymbol{\epsilon}_j$ is the cumulative innovation over I_j which has mean zero and variance matrix E_j. Usually $E_j = b_j\bar{E}_j$, where $\bar{E}_j$ is the unit covariance matrix in I_j (Gamerman, 1991).

The parameter vector $\boldsymbol{\beta}_{j-1}$ is not given a full distributional form but is simply specified in terms of its mean vector and variance matrix. Thus if our prior beliefs for the parameter vector at time zero, $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0q})$, are given by

$$\boldsymbol{\beta}_0 \sim [\boldsymbol{m}_0, C_0],$$

then we can calculate the prior specification for the parameters in interval I_j as

$$\boldsymbol{\beta}_j \sim [\boldsymbol{m}_j, C_j],$$

where the mean vector is

$$\begin{aligned} \boldsymbol{m}_j &= \mathrm{E}[G_j(b_j)\boldsymbol{\beta}_{j-1} + \boldsymbol{\epsilon}_j] \\ &= G_j(b_j)\mathrm{E}[\boldsymbol{\beta}_{j-1}] + \mathrm{E}[\boldsymbol{\epsilon}_j] \\ &= G_j(b_j)\boldsymbol{m}_{j-1}, \end{aligned}$$

and the variance matrix is

$$\begin{aligned} C_j &= \mathrm{Var}(G_j(b_j)\boldsymbol{\beta}_{j-1} + \boldsymbol{\epsilon}_j) \\ &= G_j(b_j)\mathrm{Var}(\boldsymbol{\beta}_{j-1})G_j'(b_j) + \mathrm{Var}(\boldsymbol{\epsilon}_j) \\ &= G_j(b_j)C_{j-1}G_j'(b_j) + E_j, \end{aligned}$$

as $\boldsymbol{\beta}_{j-1}$ and $\boldsymbol{\epsilon}_j$ are independent. We can also calculate the covariance matrix between parameter vectors in different intervals. This is given by

$$\mathrm{Cov}(\boldsymbol{\beta}_j, \boldsymbol{\beta}_{j+l}) = C_j \prod_{m=j+1}^{j+l} G_m(b_m) = C_{j,j+l}.$$

For example the covariance matrix of $\boldsymbol{\beta}_{j-1}$ and $\boldsymbol{\beta}_j$ is

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\beta}_{j-1}, \boldsymbol{\beta}_j) &= \mathrm{Cov}(\boldsymbol{\beta}_{j-1}, G_j(b_j)\boldsymbol{\beta}_{j-1} + \boldsymbol{\epsilon}_j) \\ &= \mathrm{Var}(\boldsymbol{\beta}_{j-1})G_j(b_j) \\ &= C_{j-1}G_j(b_j). \end{aligned}$$

5.6.4 Guide relationship

We use a guide relationship proposed by West *et al.* (1985) of the form

$$\eta_{ij} \approx \log(\lambda_{ij}) = \boldsymbol{x}_i^{'}\boldsymbol{\beta}_j + \log \lambda_{0j},$$

where $\approx$ represents the guide relationship. Thus updating of the quantities η_{ij} can be seen as a guide to update the parameters in the model. The prior expectation and variance of $\log \lambda_{ij}$ are

$$\begin{aligned} f_{ij} &= \mathrm{E}_0[\log \lambda_{ij}] &&= \boldsymbol{x}_i^{'}\boldsymbol{m}_j + \log \lambda_{0j}, \\ q_{ij} &= \mathrm{Var}_0(\log \lambda_{ij}) &&= \boldsymbol{x}_i^{'}C_j\boldsymbol{x}_i. \end{aligned}$$

We shall also need covariances between each of the transformed quantities. These are

$$\begin{aligned} q_{(ik)j} &= \mathrm{Cov}_0(\log \lambda_{ij}, \log \lambda_{kj}) &&= \boldsymbol{x}_i^{'}C_j\boldsymbol{x}_k, \\ q_{i(jl)} &= \mathrm{Cov}_0(\log \lambda_{ij}, \log \lambda_{il}) &&= \boldsymbol{x}_i^{'}C_{j,l}\boldsymbol{x}_i, \\ q_{(ik)(jl)} &= \mathrm{Cov}_0(\log \lambda_{ij}, \log \lambda_{kl}) &&= \boldsymbol{x}_i^{'}C_{j,l}\boldsymbol{x}_k. \end{aligned}$$

Finally we need the covariances between the transformed quantities and the parameter values. These are

$$\begin{aligned} \boldsymbol{s}_{ij} &= \mathrm{Cov}_0(\log \lambda_{ij}, \boldsymbol{\beta}_j) &&= \boldsymbol{x}_i^{'}C_j, \\ \boldsymbol{s}_{i(jl)} &= \mathrm{Cov}_0(\log \lambda_{ij}, \boldsymbol{\beta}_l) &&= \boldsymbol{x}_i^{'}C_{j,l}. \end{aligned}$$

We wish to find a commutative Bayes linear kinematic solution to the problem. To do this we wish to update all of the parameter values for all time points for each observation. Thus Bayes linear kinematics will provide a natural smoothed solution which contrasts with the filtering method of Gamerman (1991).

We define $\boldsymbol{H} = (\boldsymbol{\eta}, \boldsymbol{\beta})^{'}$, where $\boldsymbol{\eta} = (\eta_{11}, \eta_{21}, \ldots, \eta_{p1}, \ldots, \eta_{r1}, \ldots, \eta_{p_r r})^{'}$ and

$\boldsymbol{\beta} = (\beta_{11}, \ldots, \beta_{q1}, \ldots, \beta_{1r}, \ldots, \beta_{qr})'$ to be the set of all quantities of interest.

Prior specifications for this set are given by $\mathrm{E}_0(\boldsymbol{H}) = \boldsymbol{l}$ and

$$\mathrm{Var}_0(\boldsymbol{H}) = \begin{bmatrix} \mathrm{Var}_0(\boldsymbol{\eta}) & \mathrm{Cov}_0(\boldsymbol{\eta}, \boldsymbol{\beta}) \\ \mathrm{Cov}_0(\boldsymbol{\beta}, \boldsymbol{\eta}) & \mathrm{Var}_0(\boldsymbol{\beta}) \end{bmatrix} = Z,$$

and each of the components of the matrix can be calculated in terms of the quantities found above.

5.6.5 Bayes linear kinematics

In each interval I_j we can give each of the λ_{ij} gamma prior distributions so that they are conjugate to the conditional density and survival functions given in Equations 5.17 and 5.18. If individual i is alive at time τ_{j-1} then their prior hazard for interval I_j is

$$[\lambda_{ij}]_0 \sim \mathrm{gamma}(\alpha_{ij}, \theta_{ij}).$$

The prior expectation and variance of η_{ij}, using our guide relationship, are

$$\mathrm{E}_0(\eta_{ij}) = g_1(\alpha_{ij}, \theta_{ij}) = f_{ij},$$

and

$$\mathrm{Var}_0(\eta_{ij}) = g_2(\alpha_{ij}, \theta_{ij}) = q_{ij},$$

for some functions $g_1()$ and $g_2()$. We can then solve these two equations simultaneously for α_{ij} and θ_{ij} given f_{ij} and q_{ij}. The likelihood contribution from individual i in interval I_j is then

$$(\lambda_{ij})^{\delta_{ij}} \exp\{-\lambda_{ij}(t_{ij} - \tau_{ij})\},$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if individual } i \text{ dies in } I_j, \\ 0, & \text{if individual } i \text{ survives } I_j, \end{cases}$$

and

$$t_{ij} = \begin{cases} t_i, & \text{if individual } i \text{ dies in } I_j, \\ \tau_j, & \text{if individual } i \text{ survives } I_j. \end{cases}$$

Thus the update for λ_{ij} is conjugate and so

$$[\lambda_{ij}]_1 \sim \mathrm{gamma}(\alpha_{ij} + \delta_{ij}, \theta_{ij} + t_{ij} - \tau_{j-1}).$$

This gives a posterior mean and variance for η_{ij} of

$$\mathrm{E}_1(\eta_{ij}) = g_1(\alpha_{ij} + \delta_{ij}, \theta_{ij} + t_{ij} - \tau_{j-1}) = k_{ij},$$

and

$$\mathrm{Var}_1(\eta_{ij}) = g_2(\alpha_{ij} + \delta_{ij}, \theta_{ij} + t_{ij} - \tau_{j-1}) = r_{ij}.$$

We can propagate these changes in belief about individuals through to the other individuals, other intervals and model parameters using Bayes linear kinematics. This gives an adjusted expectation and variance for $\boldsymbol{H}$ of

$$\mathrm{E}_{1(ij)}(\boldsymbol{H}) = \boldsymbol{l} + \mathrm{Cov}_0(\boldsymbol{H}, \eta_{ij})\frac{[k_{ij} - f_{ij}]}{q_{ij}},$$

and

$$\mathrm{Var}_{1(ij)}(\boldsymbol{H}) = Z - \mathrm{Cov}_0(\boldsymbol{H}, \eta_{ij})\mathrm{Cov}_0(\eta_{ij}, \boldsymbol{H})\frac{[1 - \dfrac{r_{ij}}{q_{ij}}]}{q_{ij}}.$$

We need to check whether a unique commutative solution exists. Using the uniqueness condition we have a unique solution if

$$\frac{r_{ij}}{q_{ij}} < 1, \tag{5.19}$$

for some combination of i, j. When it exists the adjusted expectation and variance are

$$\mathrm{Var}_{p^*}(\boldsymbol{H}) = \left[\sum_{j=1}^{r}\sum_{i=1}^{p_j}\mathrm{Var}^{-1}_{1(ij)}(\boldsymbol{H}) - (p^* - 1)\mathrm{Var}_0(\boldsymbol{H})\right]^{-1}$$

and

$$\mathrm{E}_{p^*}(\boldsymbol{H}) \;=\; \mathrm{Var}_{p^*}(\boldsymbol{H})\left[\sum_{j=1}^{r}\sum_{i=1}^{p_j}\mathrm{Var}^{-1}_{1(ij)}(\boldsymbol{H})\mathrm{E}_{1(ij)}(\boldsymbol{H}) - (p^* - 1)\mathrm{Var}_0(\boldsymbol{H})\mathrm{E}_0(\boldsymbol{H})\right],$$

where $p^* = p + \sum_{j=2}^{r} p_j$. From these we obtain the posterior means and variances of the parameters as the final q elements of the expectation vector and the final q diagonal elements of the variance matrix.

5.6.6 Right Censoring

We can introduce non-informative right censoring into this model in the following way. Let us suppose that, for individual i, rather than observing a death time we observe a time of censoring denoted t_i^*. Then the likelihood contribution for individual i in

interval I_j becomes

$$(\lambda_{ij})^{\delta_{ij}} \exp\{-\lambda_{ij}(t_{ij} - \tau_{j-1})\},$$

where

$$\delta_{ij} = \begin{cases} 1, \text{ if individual } i \text{ dies in } I_j, \\ 0, \text{ if individual } i \text{ survives or is censored in } I_j, \end{cases}$$

and

$$t_{ij} = \begin{cases} t_i, \text{ if individual } i \text{ dies in } I_j, \\ \tau_j, \text{ if individual } i \text{ survives } I_j, \\ t_i^*, \text{if individual } i \text{ is censored in } I_j. \end{cases}$$

5.6.7 Expectation and variance of η_{ij}

As λ_{ij} and η_{ij} are linked using a guide relationship we have some freedom with regards to how we specify the mean and variance of η_{ij}. We considered two possibilities in Section 3.13; direct calculation from the mean and variance of $\log \lambda_{ij}$ and use of the mode and the curvature at the mode of the log density of $\log \lambda_{ij}$.

In this chapter we favour the mode and curvature approach as, with many updates to perform, it is preferable not to have to use a numerical method with each update in calculation of the gamma distribution parameters.

Thus we shall set

$$\mu_{ij} = \log(\lambda_{ij}).$$

Thus our guide relationship is between μ_{ij} and η_{ij}. Rather than use the mean and variance of μ_{ij} directly we shall say that η_{ij} is such that it has mean given by the mode and variance given by the curvature at the mode of the log density of μ_{ij}.

Thus the prior mean and variance of η_{ij} are

$$f_{ij} = m_{ij}, \quad q_{ij} = -\left[\frac{d^2 l_{ij}(\mu_{ij})}{d\mu_{ij}^2}\right]^{-1}_{m_{ij}},$$

where m_{ij} is the mode and $l_{ij}(\mu_{ij})$ is the log-density of μ_{ij}. We found these quantities in Section 3.13 giving the prior expectation and variance of η_{ij} as

$$f_{ij} = \log\left(\frac{\alpha_{ij}}{\theta_{ij}}\right), \quad q_{ij} = \frac{1}{\alpha_{ij}}.$$

Given these specifications we can solve for α_{ij} and θ_{ij} directly. We see that

$$\alpha_{ij} = \frac{1}{q_{ij}}, \quad \theta_{ij} = \frac{1}{q_{ij}} e^{-f_{ij}}.$$

Having observed the outcome for individual i in interval I_j the posterior mean and variance are

$$k_{ij} = \log\left(\frac{\alpha_{ij} + \delta_{ij}}{\theta_{ij} + t_{ij} - \tau_{j-1}}\right), \quad r_{ij} = \frac{1}{\alpha_{ij} + \delta_{ij}}.$$

Thus, from the uniqueness condition, we have a unique Bayes linear kinematic update if

$$\frac{1}{\alpha_{ij} + \delta_{ij}} < \frac{1}{\alpha_{ij}}.$$

That is, if we observe at least one death as in the direct case.

5.6.8 Example: CABG data

We shall concern ourselves with 2 studies into morbidity and risk factors after coronary artery bypass graft (CABG) surgery. Both studies took place at the Freeman hospital in Newcastle, the first between January 1980 and June 1987 and the second between June 1987 and December 1992 (Posner *et al.*, 1995, 1996).

All of the participants in the studies were male. The aim was to look for associations between risk factors and outcomes. Three different surgical techniques were used for the CABG surgery and they were;

- Venous graft
- Single mammary artery graft
- Bilateral mammary artery graft

Data collection was carried out either by questionnaires given to the patients by their GP or by visits to the patients. 575 patients are considered.

The model

Although in the original dataset there were 7 response variables we shall consider just one. This is the time in months after surgery until the onset of angina and shall be denoted by the random variable T.

We shall use a selection of what were found (Posner *et al.*, 1995, 1996) to be the most useful covariates in the model.

- x_1: age in years at the time of operation.
- x_2, x_3: change in activity since operation. This is a factor with three levels. No change shall be regarded as the baseline level and then x_2 shall represent increase and x_3 decrease.
- x_4: Total kg of tobacco smoked before the operation.
- x_5: Total kg of tobacco smoked since the operation.

As well as these 5 covariates we are also interested in operation type.

Elicitation of prior information

We wish to elicit prior information. In particular, for the CABG example, this means eliciting the prior means, variances and covariances for the parameter vector at time zero $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01}, \ldots, \boldsymbol{\beta}_{05})$. Note that we include an extra parameter in the vector, β_{00}. This will take the place of the baseline hazard ($x_{i0} = 1$ always) in all time intervals. The other quantities necessary in the model specification are the system evolution matrices $G_j(b_j)$ and the variance matrices for ϵ_j, E_j.

In order to specify the prior parameters we utilise the proportional hazards assumption. If individuals i and k have hazards $h_i(t)$ and $h_k(t)$ respectively then their ratio is

$$\frac{h_i(t)}{h_k(t)} = \exp\{\beta_1(x_{i1} - x_{k1})\} \exp\{\beta_2(x_{i2} - x_{k2})\} \cdots \exp\{\beta_5(x_{i5} - x_{k5})\}.$$

Thus, if all explanatory variables between individuals i and k were equal except for variable 2, the ratio would only depend on β_2. So we can specify quantities for the parameters by eliciting information about ratios of hazards between individuals.

To do this, on the advice of Revie *et al.* (2010), we shall use the Pearson and Tukey method (Pearson & Tukey, 1965; Keefer & Bodily, 1983). This proceeds by eliciting 5%, 50% and 95% quantiles, denoted $Q_Z(0.05), Q_Z(0.5)$ and $Q_Z(0.95)$, for some variable Z. The mean and variance are then given by

$$\begin{aligned} \mathrm{E}(Z) &= 0.63Q_Z(0.5) + 0.185[Q_Z(0.05) + Q_Z(0.95)], \\ \mathrm{Var}(Z) &= \left(\frac{Q_Z(0.95) - Q_Z(0.05)}{3.29 - 0.1C}\right), \end{aligned}$$

where

$$C = (Q_Z(0.95) + Q_Z(0.05) - 2Q_Z(0.5)) \left(\frac{3.25}{Q_Z(0.95) + Q_Z(0.05)}\right)^2.$$

	$Q_{\beta_{0l}}(0.05)$	$Q_{\beta_{0l}}(0.5)$	$Q_{\beta_{0l}}(0.95)$	$\mathrm{E}_0(\beta_{0l})$	$\mathrm{Var}_0(\beta_{0l})$
β_{02}	-0.69	-0.22	-0.01	-0.268	0.0312
β_{03}	0.049	0.182	0.693	0.252	0.0630
β_{04}	0.0018	0.0041	0.041	0.0055	1.79×10^{-7}
β_{05}	0.018	0.069	0.139	0.073	0.00239

Table 5.7: Elicited quantiles and resulting expectations and variances for the parameters

If we consider β_{01} then this is associated with the variable for age of individual. Ages in the investigation range from 29 to 78 and so we can consider that lower age would imply lower risk of quick relapse of angina. We elicit the mean and variance of β_{01} by supposing that patient i is 10 years older than patient k. Then the ratio of hazard functions for the two patients is

$$\frac{h_i(t)}{h_k(t)} = \exp\{10\beta_{01}\},$$

as long as individuals i and k have identical remaining covariates. We elicit $5\%, 50\%$ and 95% quantiles for the increase in risk associated with this increase in age. These are 1.2, 2 and 4 respectively. Taking logs and dividing by 10 gives prior quantiles for β_{01} of 0.018, 0.069 and 0.139. Using the Pearson and Tukey formulae this gives a prior mean and variance for β_{01} of

$$\mathrm{E}_0(\beta_{01}) = 0.073, \quad \mathrm{Var}_0(\beta_{01}) = 0.00239.$$

We can perform this process for $\beta_{02}, \ldots, \beta_{05}$. The results are given in Table 5.7. In terms of the system evolution matrices we shall set

$$G_j(b_j) = I_6,$$

for all j. That is, the prior parameter values for different intervals are linked by a simple random walk (Gamerman, 1991). Initially we shall choose E_j as

$$E_j = \frac{1}{b_j} I_6,$$

so that it too has a simple form. Each of our time intervals shall be a year in length up to a maximum time of 12 years (the length of the study). Thus we have specified a decreasing innovation variance as we believe that over the long term the effects of the covariates will settle to some equilibrium value.

5.6.9 Results

Having updated the 5 parameters of interest for each of the 3 different methods of surgery using the data we can plot the effects of the 5 covariates over time. All calculations used the mode and log curvature model for expectations and variances. The posterior means for the effect of increased and decreased activity for bilateral mammary artery graft surgery is seen in Figure 5.7.

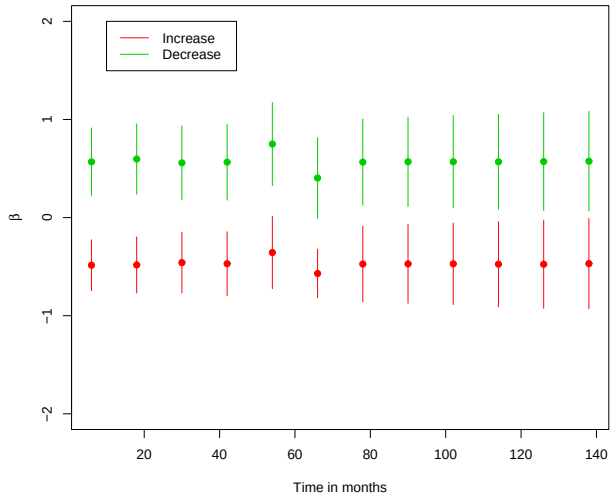

Figure 5.7: The effect of increased activity and decreased activity on the onset of angina after bilateral mammary artery graft surgery

The effect of increased activity is shown in red and decreased activity in green. The parameter means are plotted at the mid-point of each interval but would remain constant within that interval (year). Also plotted are posterior ± 2 standard deviation intervals.

The effect of the two covariates, compared to the baseline of no change in activity since surgery, is marked. Almost all of the posterior means for increased activity are negative, suggesting that this may decrease the hazard for the recurrence of angina. Decreased activity seems to have the opposite effect.

It would appear that our dynamic model could be appropriate as the effects of both covariates appears to change a little over time. As a result of the variance structure assumed the parameter values settle down as time increases.

We wish to know how much of an effect the prior means for the effects of these parameters are having on the posterior means. We can compare the posterior means for increased and decreased activity to their posterior means if they were given prior means

	Posterior means	
Interval	$\mathrm{E}_0(\beta_{jl}) = 0.252$	$\mathrm{E}_0(\beta_{jl}) = 0.000$
1	0.5677157	0.3069310
2	0.5964486	0.3282760
3	0.5577411	0.3060185
4	0.5648023	0.3144519
5	0.7496526	0.4647165
6	0.4031519	0.1732290
7	0.5654184	0.3092419
8	0.5689720	0.3120885
9	0.5691358	0.3121256
10	0.5684579	0.3117424
11	0.5707288	0.3133976
12	0.5744478	0.3166212

Table 5.8: Posterior means for the effect of decreased activity under different prior means

of zero under bilateral artery graft surgery. The posterior means for decreased activity in both cases are given in Table 5.8.

We see that the posterior means are larger when a positive prior expectation is specified. However, even with a prior expectation of zero all of the posterior means for the effect of decreased activity are positive. This suggests that decreased activity increases the hazard.

5.7 Conclusions

In this chapter two applications of Bayes linear kinematics have been investigated for reliability analysis, the first being the modelling of related Poisson distributions and the second in the analysis of life table data. In both cases taking transformations which mapped parameters onto an unrestricted scale allowed for more effective Bayes linear kinematic updates to be made by working on a scale in which linear fitting is more appropriate. Further, they allowed general comments to be made about when a unique commutative Bayes linear kinematic solution exists.

In the life table model a complementary log-log transformation was used as this allowed for a fairly straightforward calculation of the reliability function. Of course with the binomial distribution several transformations are possible and these have been investigated in earlier chapters. We showed, in both applications, that Bayes linear kinematics offers an alternative approach to fully Bayesian methods in which all calculations are tractable and computationally intensive numerical methods are not necessary.

In both reliability applications the Bayes linear kinematic approach makes careful assessment of genuine beliefs about relationships between quantities a practical proposition without the imposition of artificial distributional assumptions. Additional assumptions or approximations are required to interpret the results in terms of observable quantities or their untransformed moments but these are comparable to approximations which are traditionally used, for example, for confidence intervals for parameters of lifetime distributions.

The application to survival analysis considered was the piecewise constant hazards model with dependent increments. In particular, a method similar in flavour to that of Gamerman (1991) was developed which, by utilising Bayes linear kinematics, allowed development of a commutative solution in a situation in which one had previously not been available.

Chapter 6

The design of experiments

6.1 Introduction

In this chapter we consider the application of the Bayes linear kinematic approach to correlated counts within the context of experimental design. We show that the approach is not prey to the major obstacle in Bayesian design of experiments, the computational burden of having to perform large numbers of maximisations and integrations using numerical techniques (usually MCMC) or simulations.

We consider two applications within experimental design; usability testing and bioassay. We provide solutions in both cases by maximising expected utility. Within the usability testing application there are two probabilities of interest and, if a single test is performed, just one observation is made. Therefore a Bayes linear kinematic solution to the problem is possible without considering the issue of commutativity.

For the bioassay application we provide a solution which considers the sample size and design point problems simultaneously. A new utility function, the Bayes linear kinematic utility, is proposed. The application is illustrated with an example concerned with the effects of eutrophication on fish.

6.2 Bayesian experimental design

The choice of the design of an experiment can be viewed as a decision problem. There are trade offs between the costs of performing the experiment and the benefits derived from it. The benefits can be thought of in terms of gains in knowledge. There are two ways to do this. One is to think directly in terms of the value of the additional knowledge gained and the other is to consider the payoff associated with some decision to be made after the experiment which is known as the terminal decision. Benefits must

then be balanced against the costs, for example financial and ethical, of performing the experiment.

The optimal design of an experiment would be the best possible choice of design, found in a decision theoretic way. For guidance on statistical decision theory see Raiffa & Schlaifer (1961) and Smith (1992) or, for multi-attribute utilities, Keeney & Raiffa (1993).

Design of experiments problems can be represented using influence diagrams (Smith, 1992). An influence diagram presents a schematic representation of the decision problem in which rectangular nodes are used to represent decisions, round nodes, known as chance nodes, represent uncertain outcomes and a diamond node represents the resulting payoff. Dependence is represented using directed arcs so that $a \rightarrow b$ means b is dependent on a. In particular, arcs into a chance node mean that the conditional distribution specified for the variable at that node involves conditioning on the values of the variables at the direct predecessors. In the case of a decision node, the arcs into it show the information available when the decision is to be made.

As Farrow & Goldstein (2006) comment, in Bayesian experimental design there is also the issue of prior knowledge which is hard to quantify for complicated design problems. The typical form of a Bayesian experimental design problem can be illustrated using the influence diagram in Figure 6.1.

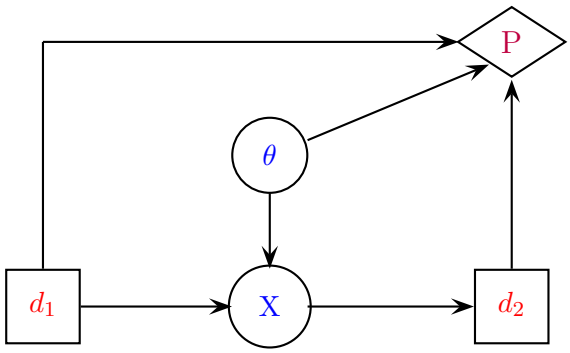

Figure 6.1: Influence diagram showing the process of Bayesian experimental design

In the figure $d_1 \in D_1$ represents the initial decision to be made, that of the choice of the design of the experiment. Performing the experiment will result in data, $x \in X$, being obtained. The data will depend on the choice of experiment, d_1, and some underlying parameters $\theta \in \Theta$. For example, if the experiment was being conducted to test the effectiveness of certain types of fertiliser then θ could be the actual effectiveness of each of the fertilisers. It is never possible to observe θ.

Having performed the experiment and observed the outcome, there is a further decision to be made, the terminal decision d_2. In the fertilisers example this might be which fertiliser to release onto the market. The terminal decision having been made, a payoff P will ensue. Preferences among different probability distributions for P are described by a utility function $U(P)$. This utility function expresses the decision maker's preferences over uncertain outcomes. The objective of a Bayesian experimental design problem is to maximise the prior expectation of the utility function $\mathrm{E}\{U(P)\}$.

Lindley (1972) considered the optimisation problem and gave a description of it, repeated in Valks (2005), as consisting of two parts;

- a prescription of the experiment to be performed (d_1), and
- a decision rule prescribing the optimal terminal decision d_2 for every outcome x of the chosen d_1.

Chaloner & Verdinelli (1995) discuss the procedure for a general utility function of the form $U(d_1, d_2, \theta, x)$. For any initial decision d_1, the expected utility of the best decision is given by

$$U(d_1) = \int_{\mathcal{X}} \max_{d_2 \in D_2} \int_{\Theta} U(d_1, d_2, \theta, x) f(\theta|x, d_1) f(x|d_1) d\theta dx,$$

where f represents the relevant probability density function. The Bayesian solution to the experimental design problem is then provided by the design d_1^* which maximises this equation, i.e.,

$$U(d_1^*) = \max_{d_1 \in D_1} \int_{\mathcal{X}} \max_{d_2 \in D_2} \int_{\Theta} U(d_1, d_2, \theta, x) f(\theta|x, d_1) f(x|d_1) d\theta dx. \tag{6.1}$$

This approach was first introduced by Lindley (1972).

We see that the theory of Bayesian experimental design is fairly straightforward. However, a full Bayesian analysis tends to be computationally difficult. Müller (1999) comments that except in special cases neither the maximisation nor the the integration can be solved analytically and approximation or simulation based methods or both are needed.

With the relatively recent advances in the area of numerical integration more complicated problems can now be solved, usually via MCMC methods. Examples of where such design problems are solved in a Bayesian context include Lindley (1997), Gittens & Pezeshk (2000) and Farrow & Goldstein (2006).

There have also been several papers utilising simulation methods in the recent literature. Müller (1999) reviews four different simulation strategies; prior simulation,

smoothing of Monte Carlo simulations, a Markov chain Monte Carlo simulation strategy and a simulated annealing type approach. Kuo *et al.* (1999) use simulation to find the optimal design in a quantal bioassay context. They utilise a nonparametric Bayesian approach to do this and assume a Dirichlet process prior.

Walker (2003) also proposes a nonparametric approach to the problem of sample size determination, utilising simulation methods to do so. Müller (2004) considers the problem of finding the optimal sample size in the context of multiple testing. His illustration is the choice of the number of microarray experiments but his approach is more widely applicable to situations where marginal and posterior distributions are efficient to sample from.

M'Lan *et al.* (2008) consider finding the optimal sample size in a Bayesian context for binomial proportions. They give an overview of the area and discuss several simulation and numerical approaches including Monte Carlo simulation and curve fitting techniques.

Thus we see that solution of this problem is still typically extremely computationally intensive and tends to require many complex integrals to be computed numerically or using simulation.

6.2.1 Utility

In order to solve a decision problem such as those found in experimental design we need to specify a utility function. But what is a utility function? We shall answer this question following the explanations in Smith (1992), Wilkinson (1998) and French & Insua (2000).

Suppose that when we make a decision we receive a reward. We wish to make the decision which gets us the 'best' reward. For example the rewards could be to be given tickets to a Sunderland football match or a hot air balloon ride (a tough choice!). Any person given the decision to make would choose the reward that they prefer out of the two.

Not all decisions lead to certain rewards, however. It could be the case that we have the decision of whether to take a reward of £50 with certainty or receive either £0 or £100, each with probability a half. This second type of reward is known as a gamble or lottery. More formally, a gamble

$$G = \alpha_1 R_1 + \alpha_2 R_2 + \ldots + \alpha_k R_k$$

returns reward R_i with probability α_i. Gambles are also known as distributions of

rewards.

We now impose some (reasonable) conditions on gambles (or rewards) in order to formulate utility. The first gives a rule for preference orderings.

If we have 2 gambles G_1 and G_2 then you either prefer G_1 to G_2 denoted $G_1 \overset{*}{>} G_2$, prefer G_2 to G_1 ($G_1 \overset{*}{<} G_2$) or find G_1 and G_2 equally preferable ($G_1 \overset{*}{=} G_2$).

So in the choice between gambles above you would either prefer to take £50 with certainty, prefer £0 or £100 each with probability a half or be indifferent between the two.

For a coherent individual preferences over gambles must also be transitive. That is, for gambles G_1, G_2 and G_3, if

$$\begin{array}{llll} G_1 \overset{*}{<} G_2 & \text{and} & G_2 \overset{*}{<} G_3 & \text{then} \quad G_1 \overset{*}{<} G_3 \\ G_1 \overset{*}{<} G_2 & \text{and} & G_2 \overset{*}{=} G_3 & \text{then} \quad G_1 \overset{*}{<} G_3 \\ G_1 \overset{*}{=} G_2 & \text{and} & G_2 \overset{*}{<} G_3 & \text{then} \quad G_1 \overset{*}{<} G_3 \\ G_1 \overset{*}{=} G_2 & \text{and} & G_2 \overset{*}{=} G_3 & \text{then} \quad G_1 \overset{*}{=} G_3 \end{array}$$

It follows from this that for gambles $G_1, \ldots, G_r$ there is a *preference ordering* given by

$$G_1 \overset{*}{\leq} G_2 \overset{*}{\leq} \ldots \overset{*}{\leq} G_r,$$

where $G_i \overset{*}{\leq} G_j$ means that either gamble G_j is preferred to G_i or the two gambles are equally preferred.

Gambles must also satisfy an independence condition. That is, for all $0 < \alpha < 1$, where α is a probability,

$$G_1 \overset{*}{<} G_2 \Leftrightarrow \alpha G_1 + (1-\alpha)G_3 \overset{*}{<} \alpha G_2 + (1-\alpha)G_3.$$

The final condition concerns the continuity of gambles. It says that if $G_1 \overset{*}{<} G_2 \overset{*}{<} G_3$ then there exist probabilities α and β such that

$$G_2 \overset{*}{<} \alpha G_1 + (1-\alpha)G_3 \quad \text{and} \quad G_2 \overset{*}{>} \beta G_1 + (1-\beta)G_3.$$

If your preferences satisfy these conditions then there exists a utility function U such that

$$U(G_i) < U(G_j) \quad \text{whenever} \quad G_i \overset{*}{<} G_j.$$

See, for example, Smith (1992). An alternative but equivalent definition (French & Insua, 2000) is as follows.

Definition. A *utility function* U on gambles G assigns a real number $U(G)$ to each

gamble subject to the following 2 conditions;

- If $G_1 \overset{*}{<} G_2$ then $U(G_1) < U(G_2)$ and if $G_1 \overset{*}{=} G_2$ then $U(G_1) = U(G_2)$.
- For any $\alpha \in [0, 1]$ and gambles G_1 and G_2

$$U(\alpha G_1 + (1 - \alpha)G_2) = \alpha U(G_1) + (1 - \alpha)U(G_2).$$

Thus utility can be thought of as a measure of our attitude towards gambles. The larger the utility, the stronger our preference is for the gamble.

If we consider the utility function of a gamble $U(G) = \alpha_1 U(R_1) + \ldots + \alpha_k U(R_k)$ then it is clear that

$$U(G) = \mathrm{E}[U(G)]. \tag{6.2}$$

This is a very important result as it indicates that the utility of a gamble is equal to its expectation in the case of a single attribute. Thus we can solve a decision problem for a decision maker's optimal decision by finding the maximum expected utility as this corresponds to their highest utility. Another important property of utility is that it is unique up to linear transformations. However, it can be shown (DeGroot, 1970), that the utility function resulting from such a linear transformation is equivalent to the original utility function.

In fact, it is easily seen that preferences implied by U are the same as those implied by

$$U^* = a + bU,$$

for constants a, b where $b > 0$. So U and U^* are strategically equivalent.

Often utility functions are defined in terms of monetary values (e.g. profits to be made or costs to be incurred). A risk neutral individual's utility would increase linearly with money. A risk averse individual would have a concave utility function and a risk prone individual's utility function would be convex. It follows from this that

$$\begin{aligned} \text{Risk aversion} &\Rightarrow U''(g) < 0, \quad \forall g, \\ \text{Risk neutrality} &\Rightarrow U''(g) = 0, \quad \forall g, \\ \text{Risk proneness} &\Rightarrow U''(g) > 0, \quad \forall g, \end{aligned}$$

Risk aversion can also be measured (French & Insua (2000)) by

$$r(g) = -\frac{U''(g)}{U'(g)},$$

where $r(g) > 0$ for a risk averse individual, $r(g) = 0$ for a risk neutral individual and

$r(g) < 0$ for a risk prone individual.

6.2.2 Multi-attribute utility

So far we have considered utility functions in terms of gambles on just a single type of reward, for example money. But what if you have multiple attributes in your decision problem?

For illustration consider a medical experiment. One attribute in the problem could be the financial cost of performing the experiment. There are also ethical costs to take into account, which could include harm or discomfort to experimental subjects, and benefits, perhaps in terms of gain in knowledge or information. Thus there would be three attributes to consider in the utility function. How do we construct a utility function which combines each of these attributes?

We have the following definitions from Keeney & Raiffa (1993).

Definition: Attributes $\boldsymbol{A}_1 = (A_{11}, \ldots, A_{1m})$ and $\boldsymbol{A}_2 = (A_{21}, \ldots, A_{2l})$ are *utility independent* if conditional preferences over lotteries on $\boldsymbol{A}_1$ given $\boldsymbol{A}_2 = \boldsymbol{a}_2$ do not depend on the value of $\boldsymbol{a}_2$.

Definition: Attributes $\boldsymbol{A} = (A_1, \ldots, A_n)$ are *mutually utility independent* if every subset of $\boldsymbol{A}$ is utility independent of its complement.

If all $\boldsymbol{A}$ can be assumed to be mutually utility independent then we have the result (Keeney & Raiffa, 1993) that the overall utility function has to take one of two forms;

$$\begin{aligned} \text{Additive} \qquad & U(\boldsymbol{A}) = \textstyle\sum_{i=1}^{n} c_i U_i(A_i), \\ \text{Multiplicative} \quad & (1 + kU(\boldsymbol{A})) = \textstyle\prod_{i=1}^{n} (1 + kc_i U_i(A_i), \end{aligned}$$

where $U_i(A_i)$ is the marginal utility for A_i. The additive form is the special case of the multiplicative form with $k = 0$ since we can rearrange the multiplicative form into

$$U(\boldsymbol{A}) = \sum_{i=1}^{n} c_i U_i(A_i) + k \sum_{i \neq j} \sum c_i c_j U_i(A_i) U_j(A_j) + k^2 \sum_{i \neq j} \sum_{i \neq k} \sum c_i c_j c_k U_i(A_i) U_j(A_j) U_k(A_k) + \ldots$$

Equation 6.2 still holds in the case of more than one attribute. However, we can not necessarily combine marginal utilities for attributes by combining their expectations over lotteries. So $U(G) = \mathrm{E}(U(G))$ and $U(\boldsymbol{A}) = U_1(A_1)U_2(A_2)$, assuming $\boldsymbol{A} = (A_1, A_2)$, but

$$U(G) = \mathrm{E}(U(G)) \neq \mathrm{E}(U_1(G))\mathrm{E}(U_2(G)).$$

However, if F_1 and F_2 are two probability distributions over $\boldsymbol{A}$ then probability distribution F_1 is at least as desirable as F_2 (Keeney & Raiffa, 1976) if and only if

$$E_{F_1}[U(\boldsymbol{A})] \geq \mathrm{E}_{F_2}[U(\boldsymbol{A})],$$

where E_{F_i} is the expectation with respect to F_i. Thus, in the multi-attribute case, maximising expected utility is still an appropriate procedure.

In order to combine utility functions like this they must first be on a common scale. The scale generally chosen is $[0, 1]$ with a utility of one representing the best possible outcome and zero the worst. It is important that neither marginal utility changes sign. If there are just two attributes, A_1 and A_2, the two cases reduce to a single binary form for the utility function

$$U(A_1, A_2) = c_1 U_1(A_1) + c_2 U_2(A_2) + c_3 U_1(A_1) U_2(A_2),$$

where $0 < c_i < 1$, $-c_i \leq c_3 < 1 - c_i$ for $i = 1, 2$ and $c_1 + c_2 + c_3 = 1$. For an additive utility function $c_3 = 0$. We can classify attributes A_1 and A_2 using c_3. If

$$\begin{aligned} c_3 > 0 \quad &\Rightarrow \quad \text{attributes are complementary,} \\ c_3 < 0 \quad &\Rightarrow \quad \text{attributes are substitutes,} \\ c_3 = 0 \quad &\Rightarrow \quad \text{attributes are preference unrelated.} \end{aligned}$$

We can think about how we might specify c_1, c_2 and c_3 (Keeney & Raiffa, 1976). We define $\overline{a}_i$ as the best possible value of attribute i and $\underline{a}_i$ as the worst possible value. Then c_1 is the probability under which you are indifferent between $(\overline{a}_1, \underline{a}_2)$ and a gamble on $(\underline{a}_1, \underline{a}_2)$ and $(\overline{a}_1, \overline{a}_2)$, i.e.,

$$(\overline{a}_1, \underline{a}_2) \stackrel{*}{=} (1 - c_1)(\underline{a}_1, \underline{a}_2) + c_1(\overline{a}_1, \overline{a}_2).$$

Similarly c_2 is the probability such that

$$(\underline{a}_1, \overline{a}_2) \stackrel{*}{=} (1 - c_2)(\underline{a}_1, \underline{a}_2) + c_2(\overline{a}_1, \overline{a}_2).$$

We can then find c_3 as $1 - c_1 - c_2$.

In order to construct complex multi-attribute utility functions it can be useful to consider utility hierarchies (Keeney & Raiffa, 1976, 1993). We can represent such a hierarchy in graphical form. The overall utility is separated into the marginal utilities of its individual attributes, each of which is represented by a node. Arrows from each of the attributes into the overall utility node indicate that this is the 'child' node for

each of the 'parent' attribute nodes. Each of the attributes can then be separated into sub-attributes as necessary. The sub-attributes are then the parent nodes of the child node for the corresponding attribute.

If, for each child node, the parent nodes are all mutually utility independent, we call the resulting hierarchy a mutually utility independent hierarchy (Farrow & Goldstein, 2006). We can construct a utility function given such a hierarchy in the following way.

- For each parent set of sub-attributes at the lowest level of the hierarchy construct an additive or a multiplicative utility function for the child.
- Repeat this step for each node at the next level up in the hierarchy and continue this process until the overall utility is obtained.

6.3 Usability experiment

Let us now concern ourselves with usability testing (Dumas & Redish, 1999) prior to software going on sale or the launch of a new website. One important aspect of this is to see whether the product is 'user-friendly'. This is generally done by taking a sample of users and asking each to perform a number of tasks. From the results of these tasks usability problems with the software are identified and the software can either be launched as it is or rewritten. To find the optimal design for such an experiment a Bayesian decision theoretic approach can be adopted. This has been considered by Valks (2005) in the context of finding the optimal sample size and this is what we shall concern ourselves with here. Valks, however, considered a fully Bayesian approach to the problem.

The decision tree relating to such an approach for a single task takes the form of Figure 6.2. For simplicity, only one branch for n and one branch for x are shown.

In the user experiment there are two decisions to be made. The first is that of the sample size; how many users to include in the experiment. Having performed the experiment there is then the decision of whether to launch or rewrite. This will be based on how many of the users successfully completed the task in the experiment. This is the terminal decision. Zero users is a special case of the sample size in which the optimal decision is deemed to be not to perform a usability experiment but simply to launch or rewrite the software immediately.

The user problem outlined above can be solved in a Bayesian context using expected utility theory. To do this we maximise the expected utility beginning with the terminal decision and then fold the decision tree backwards to the optimal sample size determination. We shall now outline the solution to this problem in the case of a single

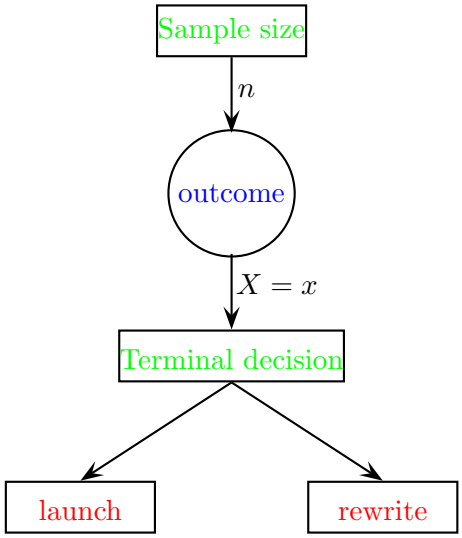

Figure 6.2: Decision tree for the user problem

task.

6.3.1 The Bayes linear kinematic solution

If n users are asked to perform a single task then the number of users who successfully complete that task, X, follows a binomial distribution

$$X \mid \theta_l \sim \text{bin}(n, \theta_l),$$

where θ_l is the probability that a user successfully completes the task if the product were simply to be launched. We make an exchangeability assumption here; that users in the experiment and users who buy the product in the future are exchangeable. We are also interested in a second probability, θ_r, that of a user successfully completing the task after rewrite. The two probabilities can be given conjugate prior beta distributions

$$\theta_l \sim \text{beta}(a_l, b_l), \quad \theta_r \mid a_r, b_r \sim \text{beta}(a_r, b_r).$$

Clearly θ_l and θ_r will not in general be independent in our prior beliefs. If the decision maker learned that the true probability of a user completing the task after launch was higher than their expectation then this would likely lead to a revision upwards of the expected probability of success after rewrite. To incorporate this dependence we shall first transform the probabilities using the canonical link function for binomial data, the

logit link,

$$\eta_l = \log\left(\frac{\theta_l}{1-\theta_l}\right), \quad \eta_r = \log\left(\frac{\theta_r}{1-\theta_r}\right),$$

so that η_l and η_r are on an unrestricted scale. We then link η_l and η_r in a Bayes linear structure. Prior beliefs in terms of θ_l and θ_r can be converted to those in terms of η_l and η_r, as seen in Chapter 3, via

$$\begin{aligned} \mathrm{E}_0(\theta_k) = \frac{a_k}{a_k+b_k} &\Rightarrow \mathrm{E}_0(\eta_k) = h_1(a_k, b_k), \\ \mathrm{Var}_0(\theta_k) = \frac{a_k b_k}{(a_k+b_k)^2(a_k+b_k+1)} &\Rightarrow \mathrm{Var}_0(\eta_k) = h_2(a_k, b_k), \end{aligned}$$

for $k = l, r$ where h_1 and h_2 are functions to be specified. To make Bayes linear updates a prior covariance between η_l and η_r, $\mathrm{Cov}_0(\eta_l, \eta_r)$ must also be given.

Now, when $X = x$ successes are observed θ_l is updated to

$$\theta_l \mid X = x \sim \mathrm{beta}(a_l^*, b_l^*),$$

where $a_l^* = a_l + x$ and $b_l^* = b_l + n - x$ as the beta distribution is conjugate to the binomial distribution. This leads to posterior moments for η_l of

$$\begin{aligned} \mathrm{E}(\eta_l \mid X = x) &= h_1(a_l + x, b_l + n - x), \\ \mathrm{Var}(\eta_l \mid X = x) &= h_2(a_l + x, b_l + n - x). \end{aligned}$$

These updates can be propagated through to η_r via the Bayes linear kinematic updating equations;

$$\begin{aligned} \mathrm{E}(\eta_r \mid X = x) &= \mathrm{E}_0(\eta_r) + \frac{\mathrm{Cov}_0(\eta_r, \eta_l)}{\mathrm{Var}_0(\eta_l)} \left[\mathrm{E}(\eta_l \mid X = x) - \mathrm{E}_0(\eta_l)\right], \\ \mathrm{Var}(\eta_r \mid X = x) &= \mathrm{Var}_0(\eta_r) - \frac{\mathrm{Cov}_0^2(\eta_r, \eta_l)}{\mathrm{Var}_0(\eta_l)} \left[1 - \frac{\mathrm{Var}(\eta_l \mid X = x)}{\mathrm{Var}_0(\eta_l)}\right]. \end{aligned}$$

With a single observation only a single update is made and so it is not necessary to consider any commutativity or uniqueness criteria here. The parameters of the posterior beta distribution for θ_r can then be found by solving the following equations for a_r^* and b_r^*;

$$\mathrm{E}(\eta_r \mid X = x) = h_1(a_r^*, b_r^*), \quad \mathrm{Var}(\eta_r \mid X = x) = h_2(a_r^*, b_r^*).$$

Thus the posterior beta distribution for θ_r is $\theta_r \mid X = x \sim$beta(a_r^*, b_r^*). We now have $f_0(\theta_l)$, $f_0(\theta_r)$, $f(\theta_r \mid X = x)$ and $f(\theta_r \mid X = x)$ which we shall need when solving the

decision problem. All are of the standard beta density form

$$f(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}.$$

Consider Y, the number of customers who complete the task after the software has been released. This also follows a binomial distribution

$$Y \sim \text{bin}(N, \theta_k \mid X = x),$$

where N is the number of customers who have bought the software and $k = l, r$ depends on whether the terminal decision was to launch or rewrite. Thus

$$f(Y \mid X = x, \theta_k) = \begin{pmatrix} N \\ y \end{pmatrix} \theta_k^y (1-\theta_k)^{N-y}.$$

In order to perform a decision analysis the final thing we shall need is a utility function. Following Valks (2005) a bivariate utility function shall be chosen. This can be represented diagramatically (Farrow & Goldstein, 2006) as in the mutually utility independent hierarchy given in Figure 6.3.

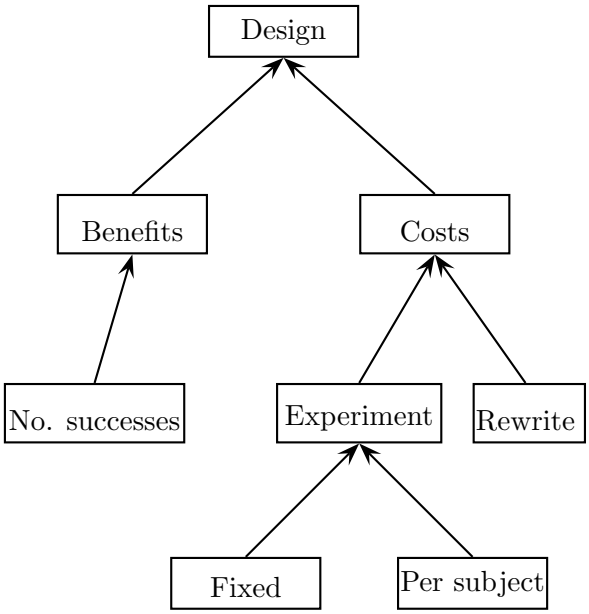

Figure 6.3: Decomposition of the utility function for the user problem

Thus we have two different utility functions to combine into our overall utility function; one for financial costs $U_f(C_k)$ which will depend on whether the software is rewritten or launched and one for benefits in terms of the number of successes in the task of future customers $U_s(Y)$. The costs are given by the cost of performing the experiment,

both fixed and per subject, and the cost of rewrite.

If we assume mutual utility independence (Keeney & Raiffa, 1993) there are two general types of bivariate utility function; additive and multiplicative. Since the former is a special case of the latter we shall consider a multiplicative utility function which takes the form

$$U = p_1 U_f(C_k) + p_2 U_s(Y) + p_3 U_f(C) U_s(Y),$$

where $p_3 = 1 - p_1 - p_2$. If $p_3 = 0$ then we have an additive utility function.

In order to formulate a solution to the decision problem let us define some quantities. Let s_n be the decision to use a sample size of n and let t_k be the terminal decision, i.e., $k = l$ for launch and $k = r$ for rewrite. Let us also define the joint prior distribution of θ_l and θ_r to be $f_0(\theta_l, \theta_r)$ and their joint posterior distribution, having performed the experiment, to be $f(\theta_l, \theta_r \mid X = x)$.

The solution to Equation 6.1 can then be found by maximising the expected utility for the terminal decision to launch or rewrite given a sample size. This is

$$\mathrm{E}[U(t_{opt} \mid X = x)] = \max\{\mathrm{E}[U(t_l \mid X = x)], \mathrm{E}[U(t_r \mid X = x)]\},$$

where

$$\mathrm{E}[U(t_k \mid X = x)] = \sum_{y=0}^{N} \int_0^1 \int_0^1 U(s_n, t_k, Y, C_k) f(Y = y \mid X = x, \theta_k) f(\theta_l, \theta_r \mid X = x) d\theta_l d\theta_r,$$

for $k = l, r$. Having done this the optimal sample size can be calculated as that corresponding to the maximum expected utility. This is

$$\mathrm{E}[U(s_{opt})] = \max\{\mathrm{E}[U(s_n)]\}, \quad n \in \mathbb{N},$$

where the expected utility for the sample size n is

$$\mathrm{E}[U(s_n)] = \sum_{x=0}^{n} f(x) \mathrm{E}[U(t_{opt} \mid X = x)]. \tag{6.3}$$

The probability that $X = x$, $f(x)$, is given by

$$f(x) = \int_0^1 f(x \mid \theta_l) f_0(\theta_l) d\theta_l = \binom{n}{x} \frac{\Gamma(a_l + b_l)}{\Gamma(a_l + b_l + n)} \frac{\Gamma(a_l + x)}{\Gamma(a_l)} \frac{\Gamma(b_l + n - x)}{\Gamma(b_l)}.$$

Now, having made the update $X = x$ the overall utility function $U(s_n, t_k, Y, C_k)$ depends only upon the chosen θ. Therefore one of the posterior beta distributions will always be sufficient to find the expected utilities. Thus the terminal decision is the

solution of

$$\mathrm{E}[U(t_{opt} \mid X = x)] = \max\{\mathrm{E}[U(t_l \mid X = x)], \mathrm{E}[U(t_r \mid X = x)]\},$$

where

$$\mathrm{E}[U(t_k \mid X = x)] = \sum_{y=0}^{N} \int_0^1 U(s_n, t_k, Y, C_k) f(Y = y \mid X = x, \theta_k) f(\theta_k \mid X = x) d\theta_k,$$

for $k = l, r$. The optimal sample size is then calculated exactly as above. We can substitute $f(Y = y \mid X = x, \theta_k)$ and $f(\theta_k \mid X = x)$ into the above expected utility to give

$$\begin{aligned}\mathrm{E}[U(t_k \mid X = x)] = \sum_{y=0}^{N} \int_0^1 \left(\begin{array}{c} N \\ y \end{array} \right) \theta_k^y (1-\theta_k)^{N-y} \\ \times \frac{\Gamma(a_k^* + b_k^*)}{\Gamma(a_k^*)\Gamma(b_k^*)} \theta_k^{a_k^*-1} (1-\theta_k)^{b_k^*-1} d\theta_k U(s_n, t_k, Y, C_k)\end{aligned}$$

$$= \sum_{y=0}^{N} \left(\begin{array}{c} N \\ y \end{array} \right) \frac{\Gamma(a_k^* + b_k^*)}{\Gamma(a_k^* + b_k^* + N)} \frac{\Gamma(a_k^* + y)}{\Gamma(a_k^*)} \frac{\Gamma(b_k^* + N - y)}{\Gamma(b_k^*)} U(s_n, t_k, Y, C_k). \tag{6.4}$$

Having performed the analysis as above the solution to the decision problem is the optimal sample size s_{opt} and the optimal terminal decision given the sample size and the data t_{opt}.

6.3.2 Expectation and Variance of η_k

Method 1: direct calculation

If we calculate the mean and variance of η_k directly we see that

$$\mathrm{E}_0(\eta_k) = h_1(a_k, b_k) = \psi(a_k) - \psi(b_k), \quad \mathrm{Var}_0(\eta_k) = h_2(a_k, b_k) = \psi_1(a_k) + \psi_1(b_k),$$

where $\psi(x) = \frac{d}{dx}\log(\Gamma(x))$ is the digamma function and $\psi_1(x) = \frac{d}{dx}\psi(x)$ is the trigamma function. A proof is given in Chapter 3.

Method 2: mode and curvature

Suppose that our transformation is now given by

$$\mu_k = \log\left(\frac{\theta_k}{1-\theta_k}\right).$$

Rather than use the mean and variance of μ_k directly we return to our guide relationship, so that η_k has mean and variance given by the mode of μ_k and the curvature at the mode of the log density of μ_k.

Hence the required mean and variance, again derived in Chapter 3, are

$$\mathrm{E}_0(\eta_k) = h_1(a_k, b_k) = \log\left(\frac{a_k}{b_k}\right), \quad \mathrm{Var}_0(\eta_k) = h_2(a_k, b_k) = \frac{1}{a_k} + \frac{1}{b_k}.$$

Clearly the variance decreases if we increase either a_k or b_k which shall happen if we observe anything. Knowledge of the mean $\mathrm{E}(\eta_k) = \bar{m}_k$ and variance $\mathrm{Var}(\eta_k) = v_k$ of η_k gives parameter values

$$a_k = \frac{1 + e^{\bar{m}_k}}{v_k}, \quad b_k = \frac{1 + e^{\bar{m}_k}}{v_k e^{\bar{m}_k}}.$$

6.3.3 Example

For comparability we shall use all of the same prior specifications as Valks (2005) Example 6.7.2, in which she represents the joint density between θ_l and θ_r using a copula function. The prior specifications she uses and hence we will use are $a_l = 3,\ b_l = 2,\ a_r = 6$, and $b_r = 2$ and these lead to prior moments of

$$\mathrm{E}_0(\theta_l) = 0.6, \quad \mathrm{Var}_0(\theta_l) = 0.040$$
$$\mathrm{E}_0(\theta_r) = 0.75, \quad \mathrm{Var}_0(\theta_r) = 0.021.$$

We require the corresponding moments of η_l and η_r in order to make a Bayes linear kinematic update. For the direct method (method 1) and using the mode and curvature at the mode (method 2) these values are given in Table 6.1.

	$\mathrm{E}_0(\eta_l)$	$\mathrm{Var}_0(\eta_l)$	$\mathrm{E}_0(\eta_r)$	$\mathrm{Var}_0(\eta_r)$	$\mathrm{Cov}_0(\eta_l, \eta_r)$
Method 1	0.50	1.04	1.28	0.83	0.377
Method 2	0.41	0.83	1.10	0.67	0.387

Table 6.1: Prior moments for η_l and η_r

The prior covariance between η_l and η_r for each of the two methods was calculated from $\mathrm{Corr}_0(\theta_l, \theta_r) = 0.6$. This was achieved (approximately) via a Taylor expansion in

2-dimensions leading to

$$\begin{aligned}\mathrm{Cov}_0(\eta_l,\eta_r) \approx g(m_l,m_r) + \frac{1}{2}[&\mathrm{Var}_0(\theta_l)\frac{\partial^2}{\partial\theta_l^2}g(m_l,m_r)\\ &+ 2\mathrm{Cov}_0(\theta_l,\theta_r)\frac{\partial^2}{\partial\theta_l\partial\theta_r}g(m_l,m_r) + \mathrm{Var}_0(\theta_r)\frac{\partial^2}{\partial\theta_r^2}g(m_l,m_r)],\end{aligned}$$

where $m_l = \mathrm{E}_0(\theta_l),\ m_r = \mathrm{E}_0(\theta_r),\ g(m_l,m_r) = [\mathrm{logit}(m_l) - \mathrm{E}_0(\eta_l)][\mathrm{logit}(m_r) - \mathrm{E}_0(\eta_r)]$ and the required derivatives are found from this.

To see this consider

$$\begin{aligned}\mathrm{Cov}(\eta_l,\eta_r) &= \mathrm{Cov}\left(\log\left[\frac{\theta_l}{1-\theta_l}\right], \log\left[\frac{\theta_r}{1-\theta_r}\right]\right)\\ &= \mathrm{E}\left[g(\theta_l,\theta_r)\right]\end{aligned}$$

where $\bar{\eta}_k$ is the prior mean of η_k and

$$g(\theta_l,\theta_r) = \left(\log\left(\frac{\theta_l}{1-\theta_l}\right) - \bar{\eta}_l\right)\left(\log\left(\frac{\theta_r}{1-\theta_r}\right) - \bar{\eta}_r\right).$$

The integrals necessary to perform this calculation are intractable and so an approximation must be found.

Suppose we have a vector $\boldsymbol{X} = (X_1,\ldots,X_n)^T$ which has a mean vector $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_n)^T$. We can find the value of an infinitely differentiable function $F(\boldsymbol{X})$ as a matrix form Taylor expansion on $\boldsymbol{X}$. To second order terms this is given by

$$f(\boldsymbol{X}) \approx f(\boldsymbol{\mu}) + \nabla f(\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu}) + \frac{1}{2!}(\boldsymbol{X}-\boldsymbol{\mu})^T\nabla^2 f(\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu}),$$

where $\nabla = \left(\frac{\partial}{\partial X_1},\ldots,\frac{\partial}{\partial X_n}\right)$ and so $\nabla f(\boldsymbol{\mu})$ is the gradiant of f at $\boldsymbol{\mu}$ and $\nabla^2 f(\boldsymbol{\mu})$ is the Hessian matrix.

We shall perform the expansion on $g(\theta_l,\theta_r)$ around the point $(\bar{\theta}_l,\bar{\theta}_r)$, the prior means of θ_l and θ_r respectively. Thus, the 2-dimensional Taylor expansion of $g(\theta_l,\theta_r)$ is given by

$$\begin{aligned}g(\theta_l,\theta_r) \approx g(\bar{\theta}_l,\bar{\theta}_r) &+ (\theta_l-\bar{\theta}_l)g_{\theta_l} + (\theta_r-\bar{\theta}_r)g_{\theta_r}\\ &+ \frac{1}{2}\left[(\theta_l-\bar{\theta}_l)^2 g_{\theta_l\theta_l} + 2(\theta_l-\bar{\theta}_l)(\theta_r-\bar{\theta}_r)g_{\theta_l\theta_r} + (\theta_r-\bar{\theta}_r)^2 g_{\theta_r\theta_r}\right],\end{aligned}$$

where $g_{\theta_l} = \frac{\partial g}{\partial\theta_l}$ is the partial derivative of g with respect to θ_l, $g_{\theta_l\theta_l} = \frac{\partial^2 g}{\partial\theta_l^2}$ is the second partial derivative and $g_{\theta_l\theta_r} = \frac{\partial^2 g}{\partial\theta_l\theta_r}$ is the mixed partial derivative. All of these derivatives are to be evaluated at $(\bar{\theta}_l,\bar{\theta}_r)$.

Taking expectations we see that

$$\begin{aligned}\mathrm{E}[g(\theta_l,\theta_r)] &\approx g(\bar{\theta}_l,\bar{\theta}_r)+(\mathrm{E}[\theta_l]-\bar{\theta}_l)g_{\theta_l}+(\mathrm{E}[\theta_r]-\bar{\theta}_r)g_{\theta_r}\\ &\quad+\frac{1}{2}\left[\mathrm{E}[(\theta_l-\bar{\theta}_l)^2]g_{\theta_l\theta_l}+2\mathrm{E}[(\theta_l-\bar{\theta}_l)(\theta_r-\bar{\theta}_r)]g_{\theta_l\theta_r}+\mathrm{E}[(\theta_r-\bar{\theta}_r)^2]g_{\theta_r\theta_r}\right]\\ &\approx g(\bar{\theta}_l,\bar{\theta}_r)+\frac{1}{2}\left[\mathrm{Var}(\theta_l)g_{\theta_l\theta_l}+2\mathrm{Cov}(\theta_l,\theta_r)g_{\theta_l\theta_r}+\mathrm{Var}(\theta_r)g_{\theta_r\theta_r}\right]\end{aligned}$$

We shall use this as our assessment of the covariance between η_l and η_r. We specified both of the variances and the covariance in terms of θ_l, θ_r earlier and so now we must calculate the derivatives above. If we set

$$f(x)=\log\left(\frac{x}{1-x}\right)-c$$

then the first derivative of f is

$$\begin{aligned}f^{(1)}(x) &= \frac{1-x}{x}\times\left(\frac{x}{(1-x)^2}+\frac{1}{1-x}\right)\\ &= \frac{1}{x(1-x)}\end{aligned}$$

The other derivatives are found similarly.

Now, if we consider the bivariate utility function from Figure 6.3, we require a benefit utility based on the number of successes of future users. Following Valks (2005) we define this as

$$U_s(y)=\frac{1-\exp(-\frac{y}{10})}{1+100\exp(-\frac{y}{10})}.$$

A discussion on the suitability of this utility and that of costs used by Valks (2005) is given in Section 6.3.4.

A plot of this for different values of $Y=y$ is given in Figure 6.4. From the plot we can see that for either a very small number or very large number of successes of future customers in the task an increase in the number of successes leads to a small increase in utility whereas if the number of successes is somewhere in the middle of the range a small increase in the number of successes leads to a much larger increase in the utility for that number of successes.

The utility for cost, again following Valks (2005), shall be

$$U_f(C_k)=1-\kappa\log\left(1+\frac{2C_k}{C_{\max}}\right),$$

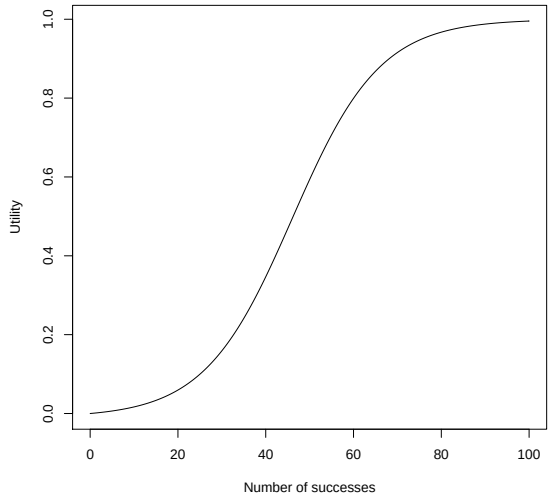

Figure 6.4: A plot of $U_s(Y)$, the utility function for the number of successes of future customers

where $\kappa = \log^{-1}(3)$ to constrain the utility, as with that of $U_s(Y)$, to $[0, 1]$. Here C_k represents the total costs incurred with $k = l, r$ for launch or rewrite and $C_{\max}$ is the maximum amount of money the company is willing to pay. This time the function is becoming steeper with decreasing cost and so the less money we spend the more keen we are to spend even less. This suggests that the decision maker is risk seeking with respect to costs. This seems unlikely. The suitability of this utility function is discussed in Section 6.3.4. A plot of this utility function for $C_{\max} = £200,000$ is given in Figure 6.5.

The two utilities can now be combined into the overall utility function as

$$U(s_n, t_k, Y, C_k) = p_1 U_f(C_k) + p_2 U_s(Y) + p_3 U_f(C_k) U_s(Y).$$

In order to carry out the analysis we take numerical values for all of these quantities.

The maximum costs the company is prepared to incur are $C_{\max} = £200,000$ as above and the cost of a rewrite of the software is $C_w = £50,000$. There is a fixed cost of performing an experiment of $C_o = £5,000$ and an additional cost per user in the experiment of $C_u = £500$. Thus the total costs if the product is launched are

$$C_l = C_0 + C_u n = 5000 + 500n,$$

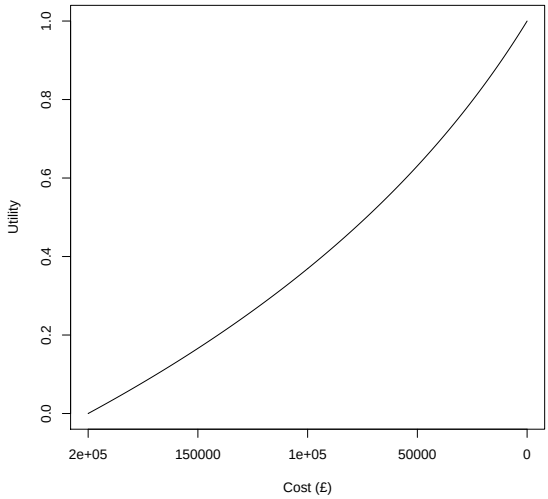

Figure 6.5: A plot of $U_f(C)$, the utility function for the costs involved

and if the product is rewritten the total costs are

$$C_r = C_o + C_u n + C_w = 55000 + 500n.$$

We shall take $p_1 = \frac{1}{6}$, $p_2 = \frac{4}{6}$ and $p_3 = \frac{1}{6}$ and, initially, $N = 100$. In order to justify these values for the trade off parameters we shall consider some utilities over different costs and benefits. If we take a total cost of £10,000, $N = 100$, a sample size of $n = 15$ and the number of future successes $y = 50$ the utility for launch is 0.6381. If the cost of the experiment is increased to £100000 then this reduces the utility. In order to achieve a utility of 0.6381 it is necessary to raise the number of future successes to $y = 60$. If we double the cost again to £200000 then in order to achieve the same utility a y of 78 is required. It is felt that this represents a sensible utility function.

The solution to the problem is found by first calculating the expected utility of launch and rewrite for each possible sample size using Equation 6.4. For each sample size the optimum value is determined and then the expected utility is averaged over the prior predictive distribution of outcomes. This is given in Equation 6.3. A plot of these quantities is given in the left hand side of Figure 6.6 for method 1 and Figure 6.7 for method 2.

The colours in these plots indicate the change in how many users it takes to be successful before the optimal decision is to launch. If two adjacent points are the same colour then the critical number of successes is the same for the two sample sizes in question. If they are different colours then the critical numbers of successes are different. It is

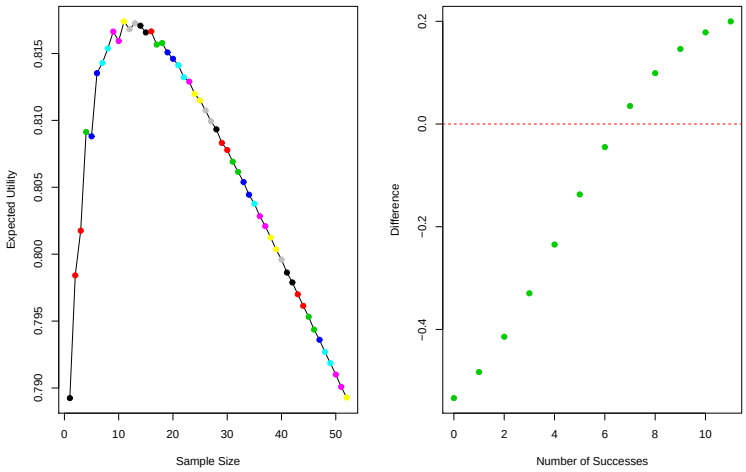

Figure 6.6: Plots of $E[U(s_n)]$ and the difference between the expected utilities of launch and rewrite for each number of observed successes given 11 users, method 1

this change in how many successes it takes for the optimal decision to be to launch which is the reason a smooth curve is not obtained.

The optimal sample size is the value which maximises this plot and is $n = 11$, giving an expected utility of 0.8174 using method 1 and at $n = 7$ giving 0.8065 using method 2. Thus the experiment should be performed with 11 users and 7 users respectively. Using these sample sizes and the fixed, per user and rewrite costs on the previous page we can calculate the overall costs incurred for the optimal solutions under the two methods. These are £10, 500 if the product is launced and £60, 500 if the product is rewritten taking 11 users (method 1) and £8500 and £58, 500 respectively taking 7 users (method 2).

Although these two optimal sample sizes do not appear too similar there is no sharp maximum in either case.

Having performed the experiment the terminal decision is then made based on the number of successes of users in the task. The expected utility of launch and rewrite given each number of possible successes $X = x$ can be calculated from Equation 6.4 and then the difference between these

$$\mathrm{E}[U(t_l \mid X = x)] - \mathrm{E}[U(t_r \mid X = x)]$$

can be plotted as in the right hand side of Figure 6.6 for method 1 and Figure 6.7 for method 2.

If this difference is positive then the optimal terminal decision is to launch and if it

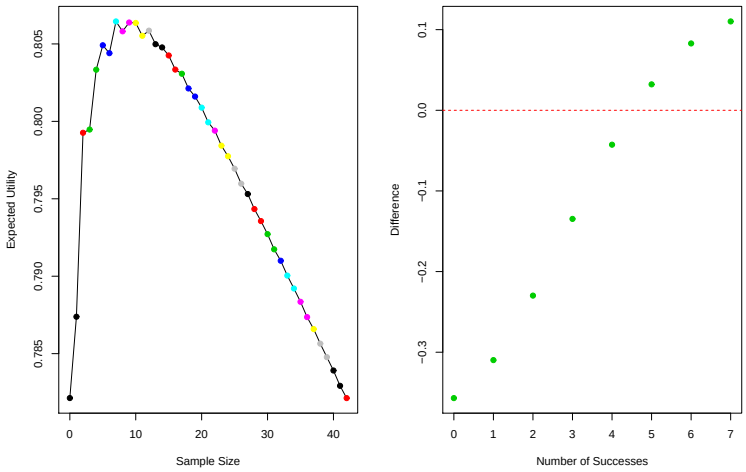

Figure 6.7: Plots of $E[U(s_n)]$ and the difference between the expected utilities of launch and rewrite for each number of observed successes given 7 users, method 2

is negative the optimal terminal decision is to rewrite. From the right hand side of Figure 6.6 we can see that if 6 or fewer users complete the task successfully we should rewrite and if 7 or more are successful we should launch using method 1. Using method 2 (Figure 6.7) we rewrite if 4 or fewer users are successful and launch if there are 5 or more successes.

Valks (2005) Chapter 6 used a joint density for θ_1, θ_2 formed using the Cook-Johnson copula, which was introduced in Chapter 2, to perform the analysis and numerical methods to compute the integrals. Thus her methodology was fully Bayesian. All of the prior specifications and the utility function used were identical to those in our analysis. As such her analysis could be thought of as the fully Bayesian 'exact' solution which we are 'approximating' with our Bayes-Bayes linear-Bayes analysis.

She found that the optimal sample size was 10, the expected utility for 10 users was 0.851 and the critical value between launch and rewrite was 7 successes. These results are consistent with ours above using the direct expectation and variance. Thus, in this example, method 1 provides a good 'approximation' to the 'exact' solution. The solution using method 2 is not particularly close to the direct method or full Bayesian solution. However, the optimal number of users using the fully Bayesian method, $n = 10$, has the third highest expected utility for method 2 and in fact the difference between it and the expected utility for the optimal sample size of 10 is just 9.78×10^{-5}.

We have plotted the adjusted expectations and variances of θ_r, $\mathrm{E}(\theta_r \mid X = x)$ and $\mathrm{Var}(\theta_r \mid X = x)$, for $n = 10$ with the prior parameter values used in the example in

Figure 6.8 for the three different methods considered as well as a model which assumes a logit-Normal prior distribution for θ_l, θ_r. The number of observed successes, x, is given on the x-axis.

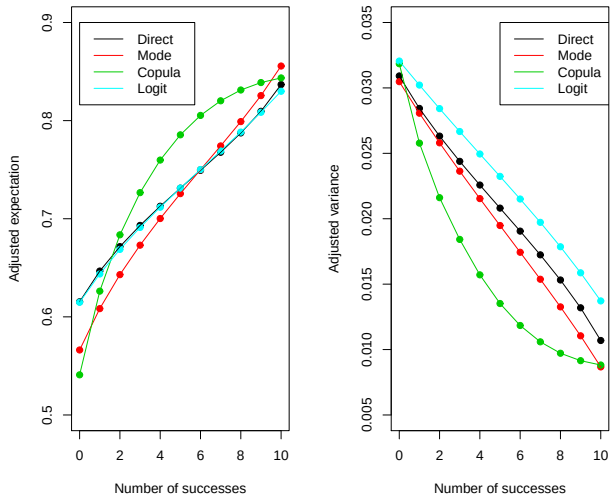

Figure 6.8: A plot of the adjusted expectations and variances of θ_r for $n = 10$

We see that the adjusted expectations and variances using methods 1 and 2 are fairly similar and are also quite similar to those of the logit-Normal model. They are both a little way from the copula based method of Valks (2005). The largest difference is between the posterior variances for the two fully Bayesian solutions.

6.3.4 Discussion

In the usability example we have followed closely the prior specifications, in terms of the parameters θ_l and θ_r and the utility functions, of Valks (2005). This has allowed us to compare her fully Bayesian approach using a copula function to our Bayes linear kinematic approach. There are certain improvements which we believe could be made to the model, however.

The first is in terms of the number of hypothetical future customers, N. We have assumed that this is fixed but how in practice would we go about choosing N in this case? In our example $N = 100$ and we could perhaps justify this by saying that the first 100 customers who buy a product shape its reputation and so are the people to

consider in the utility function. Alternatively we could specify a utility in terms of

$$A = \sum_{i=1}^{\infty} k_i s_i,$$

where k_i decreases as i increases and $s_i = 0/1$ according to whether customer i records a success. Another possibility would be

$$A = \sum_{i=1}^{\infty} k^{i-1} s_i,$$

with $0 < k < 1$. In this case

$$\mathrm{E}(A \mid \theta) = \frac{\theta}{1-k}, \quad \mathrm{Var}(A \mid \theta) = \frac{\theta(1-\theta)}{1-k^2}.$$

Both of the above approaches would have the effect of giving successes of earlier customers more weight than later customers in the utility. We could also give N some distributional form.

Consider the plot of the benefit utility given in Figure 6.4. It is convex initially, approximately linear in the middle section and concave for large numbers of successes. It would seem likely that a utility function for the number of successes would be concave. Thus a more suitable benefit utility would appear to be

$$U_s(y) = 1 - \left(\frac{y-N}{N}\right)^2, \tag{6.5}$$

for some N. Similarly, the utility for cost suggests a risk seeking individual. People tend to be risk averse when it comes to money and so a more suitable utility function could be

$$U_f(C_k) = 1 - \left(\frac{C_k}{C_{\max}}\right)^2.$$

A second possibility would be

$$U_f(C_k) = \kappa \log\left[1 + \frac{C_{\max} - C_k}{C_{\max}}\right], \tag{6.6}$$

where $\kappa^{-1} = \log 2$.

If we perform the analysis using the utility functions in Equations 6.5 and 6.6 and a fixed value of $N = 100$ we obtain an optimal sample size of $n = 10$ and an expected utility of 0.9147 for method 1. Thus, using this method, the optimal sample size remains very similar.

6.4 Bioassay experiments

Bioassay techniques are used in many different fields in order to measure the effect of varying doses of some chemical upon living things. Typically a number p of doses, $d_1, \ldots, d_p$, are chosen and at each dose d_i a number of organisms n_i are given that dose. We shall regard the doses as predetermined but it can be the case that part of the design process is to search for optimal design points. See, for example, Haines *et al.* (2003). Let us suppose that for each organism we measure whether or not a specified response is achieved. Thus for organism j taking dose i

$$X_{ij} = \begin{cases} 1, & \text{if response achieved,} \\ 0, & \text{if response not achieved.} \end{cases}$$

Then, if we regard organisms taking the same dose to be independent, we can think of the number of responses at dose i, $X_i = \sum_j X_{ij}$, as a binomial random variable

$$X_i \mid \theta_i \sim \text{bin}(n_i, \theta_i),$$

where θ_i is the probability of response for dose i. We shall also define $\boldsymbol{n} = (n_1, \ldots, n_p)'$, $\boldsymbol{X} = (X_1, \ldots, X_p)'$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$. In terms of designing a bioassay experiment interest lies in the answers to two questions;

(i) How many organisms should be included in the experiment?

(ii) What proportion of the organisms should be given each dose?

These are known as the sample size and design point problems respectively. Farrow & Goldstein (2006) Section 1 discusses previous Bayesian approaches to both problems.

6.4.1 Bayes linear kinematic solution

We shall answer these questions by solving the problem in a Bayesian context, maximising expected utility. We give each θ_i a conjugate beta prior distribution

$$\theta_i \sim \text{beta}(a_i, b_i).$$

To incorporate dependence between the θ_i's we shall first transform them to η_i's on the $(-\infty, \infty)$ scale and link the η_i's in a Bayes linear structure. Thus

$$\eta_i = g(\theta_i),$$

where $g(\theta_i)$ is one of the suitable link functions for the binomial distribution. The prior expectation and variance of θ_i are

$$\mathrm{E}_0(\theta_i) = \frac{a_i}{a_i + b_i}, \qquad \mathrm{Var}_0(\theta_i) = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)},$$

and we can convert expectations and variances in one set of variables into those of the other via the standard formulae

$$\begin{aligned} \mathrm{E}_0(\eta_i) &= \int_0^1 g(\theta_i) f_0(\theta_i) d\theta_i, \\ \mathrm{Var}_0(\eta_i) &= \int_0^1 g^2(\theta_i) f_0(\theta_i) d\theta_i - \mathrm{E}_0(\theta_i)^2, \end{aligned}$$

where $f_0(\theta_i)$ is the prior density of θ_i. Defining $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$, Bayes linear updates can be made as soon as a prior covariance matrix for $\boldsymbol{\eta}$ has been specified. When $X_i = x_i$ responses are observed out of n_i organisms for dose i, θ_i is updated to

$$\theta_i \mid x_i \sim \mathrm{beta}(A_i, B_i),$$

where $A_i = a_i + x_i$ and $B_i = b_i + n_i - x_i$ as the beta and binomial distributions are conjugate. This will lead to $\mathrm{E}(\eta_i \mid x_i)$ and $\mathrm{Var}(\eta_i \mid x_i)$ which will be of the form of their prior counterparts but using A_i and B_i.

These updates are then propagated through to $\boldsymbol{\eta}$ via the Bayes linear kinematic updating equations, numbered (3.6) and (3.7).

$$\mathrm{E}_1(\boldsymbol{\eta}; x_i) = \mathrm{E}_0(\boldsymbol{\eta}) + \frac{\mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)}{\mathrm{Var}_0(\eta_i)}[\mathrm{E}(\eta_i \mid x_i) - \mathrm{E}_0(\eta_i)], \tag{6.7}$$

$$\mathrm{Var}_1(\boldsymbol{\eta}; x_i) = \mathrm{Var}_0(\boldsymbol{\eta}) - \left(\frac{1}{\mathrm{Var}_0(\eta_i)} - \frac{\mathrm{Var}(\eta_i \mid x_i)}{\mathrm{Var}_0^2(\eta_i)}\right) \mathrm{Cov}_0(\boldsymbol{\eta}, \eta_i)\mathrm{Cov}_0(\eta_i, \boldsymbol{\eta}) \tag{6.8}$$

One such Bayes linear kinematic update is made for each i. From Equation 4.4 a sufficient condition for a unique commutative update to exist is

$$\frac{\mathrm{Var}(\eta_i \mid x_i)}{\mathrm{Var}_0(\eta_i)} < 1$$

for all i. When this solution exists it is given, as we have seen previously, by

$$\mathrm{E}_p(\boldsymbol{\eta}; \boldsymbol{x}) = \mathrm{Var}_p(\boldsymbol{\eta}; \boldsymbol{x}) \left[\sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta}; x_i)\mathrm{E}_1(\boldsymbol{\eta}; x_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\mathrm{E}_0(\boldsymbol{\eta})\right] \tag{6.9}$$

$$\mathrm{Var}_p(\boldsymbol{\eta}; \boldsymbol{x}) = \left[\sum_{i=1}^{p} \mathrm{Var}_1^{-1}(\boldsymbol{\eta}; x_i) - (p-1)\mathrm{Var}_0^{-1}(\boldsymbol{\eta})\right]^{-1}. \tag{6.10}$$

If we consider the design problem its solution is found by maximising the expected utility to give optimal sample size and allocation, using Equation 6.1. This solution is denoted $\boldsymbol{n}^* = (n_1^*, \ldots, n_p^*)'$ where

$$\mathrm{E}[U(s_{\boldsymbol{n}^*})] = \max(\mathrm{E}[U(s_{\boldsymbol{n}})]),$$

for $\boldsymbol{n} \in \mathbb{N}^p$ where $s_{\boldsymbol{n}}$ is the decision to perform the experiment with $\boldsymbol{n} = (n_1, \ldots, n_p)'$ organisms at each of the doses. The expected utility for sample allocation $\boldsymbol{n}$ is given by

$$\mathrm{E}[U(s_{\boldsymbol{n}})] = \sum_{\boldsymbol{x} \in \boldsymbol{X}} f(\boldsymbol{x}) \times \max_{t \in T} \int_{\Theta} f_p(\boldsymbol{\theta}; \boldsymbol{x}) \times U(s_{\boldsymbol{n}}, t, \boldsymbol{\theta}, \boldsymbol{x}) \; d\boldsymbol{\theta}, \tag{6.11}$$

where t is the (possibly notional) terminal decision, $f(\boldsymbol{x})$ is the probability of observing $\boldsymbol{x}$ and $U(s_{\boldsymbol{n}}, t, \boldsymbol{\theta}, \boldsymbol{x})$ is the utility function.

6.4.2 Utility function

We now consider the utility function. Following Farrow & Goldstein (2006) we can represent the utility function in terms of the hierarchy given in Figure 6.9.

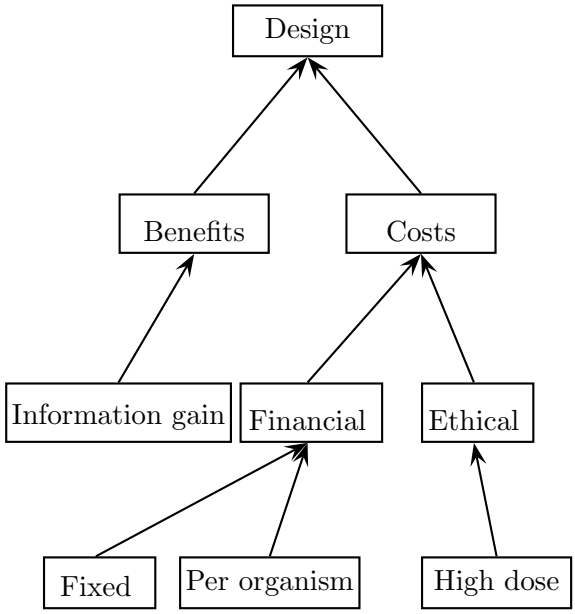

Figure 6.9: Decomposition of the utility function for a bioassay experiment

Thus the overall utility function, generally, can be broken down into utilities for benefit and cost. Benefit shall be measured in terms of the the gain in knowledge that results from performing the experiment. There are two general types of cost which may be relevent to a bioassay experiment; financial costs and ethical costs. Financial costs can

be broken down into the fixed cost of performing an experiment and the additional cost for each organism (which could depend on dose). Ethical costs can be thought of in terms of the adverse effect of giving an organism too high a dose.

If costs and benefits are assumed to be mutually utility independent then the overall utility $U(s_{\boldsymbol{n}}, d, \boldsymbol{\theta}, \boldsymbol{x}) = U$ can be represented by a binary node in terms of the utility of costs U_C and that of benefits U_B.

$$U = r_1 U_B + r_2 U_C + r_3 U_B U_C,$$

where $r_1, r_2 > 0$, $-r_i \leq r_3 < 1 - r_i$ for $i = 1, 2$ and $r_3 = 1 - r_1 - r_2$. If $r_3 = 0$ we have an additive node. Further, if financial and ethical costs can be assumed to be mutually utility independent then the utility for costs can also be represented by a binary node in terms of the utility for ethical costs U_E and the utility for financial costs U_F.

$$U_C = q_1 U_E + q_2 U_F + q_3 U_E U_F,$$

where $q_1, q_2 > 0$, $-q_i \leq q_3 < 1 - q_i$ for $i = 1, 2$ and $q_3 = 1 - q_1 + q_2$. Thus there are three utilities to specify in the analysis; U_B, U_E and U_F. All shall be defined on the standard $[0, 1]$ scale with a utility of 1 being assigned to the best possible outcome and 0 to the worst.

Benefit utility

One way to consider the benefit of performing an experiment is in terms of the gain in knowledge or information which results from the experiment. Much work has been done in this context on both the sample size problem and the design point problem (see Chaloner & Verdinelli (1995), Lindley (1997) and Farrow & Goldstein (2006)).

Farrow & Goldstein (2006) define the Bayes linear utility for information gain which is based upon the reduction in uncertainty between the prior and Bayes linear adjusted variance. They show that this can be calculated as

$$U(\boldsymbol{\beta}) = 1 - \frac{1}{k} \text{trace} \left\{ \text{Var}_0^{-1}(\boldsymbol{\beta}) \text{Var}_{\boldsymbol{\alpha}}(\boldsymbol{\beta}) \right\},$$

where $\boldsymbol{\alpha}$ is a collection of quantities which are observed, $\boldsymbol{\beta}$ is a collection of quantities about which we wish to make inferences and k is the number of elements in the vector $\boldsymbol{\beta}$. They then define the mixed Bayes linear utility which allows for gains in knowledge about certain linear combinations of the elements of $\boldsymbol{\beta}$ to be each given a Bayes linear utility.

We define the Bayes linear kinematic utility. This is

$$U(\boldsymbol{\eta}) = 1 - \frac{1}{p}\text{trace}\left\{\text{Var}_0^{-1}(\boldsymbol{\eta})\text{Var}_p(\boldsymbol{\eta};\boldsymbol{x})\right\}. \tag{6.12}$$

Justification for this choice of benefit utility can be found by considering the prior precision matrix $\text{Var}_0^{-1}(\boldsymbol{\eta})$. This is a symmetric, positive definite matrix and so there exists a lower triangular matrix W such that

$$\text{Var}_0^{-1}(\boldsymbol{\eta}) = WW^T.$$

This is the Choleski decomposition. It also follows that

$$\text{Var}_0(\boldsymbol{\eta}) = (W^T)^{-1}W^{-1}.$$

We can define $d(\boldsymbol{\eta}) = \boldsymbol{\eta} - \text{E}_p(\boldsymbol{\eta};\boldsymbol{x})$, a measure of discrepency between $\boldsymbol{\eta}$ and its Bayes linear kinematic adjusted expectation. Now consider the quantity $W^T d(\boldsymbol{\eta})$. Its variance is

$$\begin{aligned}\text{Var}_0\left(W^T d(\boldsymbol{\eta})\right) &= W^T\text{Var}_0\left(d(\boldsymbol{\eta})\right)W \\ &= W^T\text{Var}_0(\boldsymbol{\eta})W \\ &= W^T(W^T)^{-1}W^{-1}W \\ &= I.\end{aligned}$$

So, a priori, the elements of $W^T d(\boldsymbol{\eta})$ are uncorrelated and each has variance one. This means that they are proportional to the principal components of $d(\boldsymbol{\eta})$. The fact that they are uncorrelated and have the same variance also suggests that adding the expected reductions in variances might be appropriate.

We now need to find a way of measuring the reduction in variance of a multivariate vector. The principal components of $d(\boldsymbol{\eta})$ are $M^T d(\boldsymbol{\eta})$ where M is a matrix, the rows of which are the eigenvectors of $\text{Var}_0(\boldsymbol{\eta})$, such that $M^T M = I$. So

$$\begin{aligned}\text{Var}_0(M^T d(\boldsymbol{\eta})) &= M^T\text{Var}_0(\boldsymbol{\eta})M \\ &= \Lambda,\end{aligned}$$

a diagonal matrix. Now,

$$\begin{aligned}\Lambda^{-1} &= (M)^{-1}\text{Var}_0^{-1}(\boldsymbol{\eta})(M^T)^{-1} \\ &= M^T\text{Var}_0^{-1}(\boldsymbol{\eta})M,\end{aligned}$$

as M is an orthogonal matrix. So, if the principal components of $d(\boldsymbol{\eta})$ are $c_1(\boldsymbol{\eta}), \ldots, c_p(\boldsymbol{\eta})$, then

$$\begin{aligned}\sum_{i=1}^{p} \frac{c_i^2(\boldsymbol{\eta})}{\mathrm{Var}_0(c_i(\boldsymbol{\eta}))} &= (M^T d(\boldsymbol{\eta}))^T \Lambda^{-1} M^T d(\boldsymbol{\eta}) \\ &= d^T(\boldsymbol{\eta}) M M^T Var_0^{-1}(\boldsymbol{\eta}) M M^T d(\boldsymbol{\eta}) \\ &= d^T(\boldsymbol{\eta}) \mathrm{Var}_0^{-1}(\boldsymbol{\eta}) d(\boldsymbol{\eta}).\end{aligned}$$

Thus, equivalent to Equation 6 in Farrow & Goldstein (2006) we have

$$U(\boldsymbol{\eta}) = 1 - \mathrm{E}\left[\frac{1}{p} d^T(\boldsymbol{\eta}) \mathrm{Var}_0^{-1}(\boldsymbol{\eta}) d(\boldsymbol{\eta})\right].$$

To get from this to Equation 6.12 we shall need the two following properties of matrices;

(i) If $\boldsymbol{y}$ is a vector then $\boldsymbol{y}^T\boldsymbol{y} = \mathrm{trace}(\boldsymbol{y}\boldsymbol{y}^T)$.

(ii) If A and B are square matrices then trace(AB) =trace(BA).

Now, if we apply the Choleski decomposition to $\mathrm{Var}_0^{-1}(\boldsymbol{\eta})$ we see that, for some lower triangular matrix W,

$$\begin{aligned} d^T(\boldsymbol{\eta}) \mathrm{Var}_0^{-1}(\boldsymbol{\eta}) d(\boldsymbol{\eta}) &= d^T(\boldsymbol{\eta}) W W^T d(\boldsymbol{\eta}) & \\ &= \mathrm{trace}(W^T d(\boldsymbol{\eta}) d^T(\boldsymbol{\eta}) W) & \text{by (i)} \\ &= \mathrm{trace}(W W^T d(\boldsymbol{\eta}) d^T(\boldsymbol{\eta})) & \text{by (ii)} \\ &= \mathrm{trace}(\mathrm{Var}_0^{-1}(\boldsymbol{\eta}) d(\boldsymbol{\eta}) d^T(\boldsymbol{\eta})). & \end{aligned}$$

If we take the expectation of this

$$\begin{aligned}\mathrm{E}\left[d^T(\boldsymbol{\eta}) \mathrm{Var}_0^{-1}(\boldsymbol{\eta}) d(\boldsymbol{\eta})\right] &= \mathrm{trace}\left(\mathrm{Var}_0^{-1}(\boldsymbol{\eta}) \mathrm{E}[d(\boldsymbol{\eta}) d^T(\boldsymbol{\eta})]\right) \\ &= \mathrm{trace}\left(\mathrm{Var}_0^{-1}(\boldsymbol{\eta}) \mathrm{Var}_p(\boldsymbol{\eta}; \boldsymbol{x})\right),\end{aligned}$$

and so

$$U(\boldsymbol{\eta}) = 1 - \frac{1}{p} \mathrm{trace}\{\mathrm{Var}_0^{-1}(\boldsymbol{\eta}) \mathrm{Var}_p(\boldsymbol{\eta}; \boldsymbol{x})\}.$$

It is also possible to construct a mixed Bayes linear kinematic utility in the same way that Farrow & Goldstein (2006) construct a mixed Bayes linear utility. This would allow us to weight information gain about different factors differently.

Utilities for Cost

Generally people are risk averse when it comes to financial cost. Therefore a linear utility for cost is not appropriate. Instead we shall use a quadratic utility function for financial cost as this reflects a risk averse individual. Thus our utility for financial cost is

$$U_F(C) = 1 - \frac{C^2}{C_{max}^2},$$

where C is the cost of performing the experiment and C_{max} is the maximum amount the decision maker is willing to pay. The financial cost of performing the experiment is given by

$$C = C_0 + \sum_{i=1}^{p} C_i n_i,$$

where C_0 is the fixed cost associated with performing an experiment and C_i is the additional cost associated with an organism being given dose i.

We also need a utility for ethical costs. Assuming that our decision maker's ethical cost utility is linear with respect to dose, the ethical utility function takes the form

$$U_E(d) = 1 - \frac{d}{d_{max}},$$

where $d = \sum_{i=1}^{p} d_i n_i$ and d_{max} is the maximum value d can take. We could perhaps think about this as $\max\{N\}d_p$ where $N = \sum_{i=1}^{p} n_i$.

6.4.3 Possible Link Functions

There are, as we have previously seen, different possible functional forms for $g(\theta_i)$. In this section we consider the logit link and a link function which leads to a pseudo-expectation and variance.

The logit link

This takes the form

$$\eta_i = \log\left(\frac{\theta_i}{1-\theta_i}\right).$$

The prior mean and variance of η_i are

$$\mathrm{E}_0(\eta_i) = \psi(a_i) - \psi(b_i), \qquad \mathrm{Var}_0(\eta_i) = \psi_1(a_i) + \psi_1(b_i).$$

The updated mean and variance, having observed x_i, are of the same form but using

A_i and B_i. The trigamma function is monotonically decreasing on $\mathbb{R}^+$ and so

$$\frac{\mathrm{Var}(\eta_i \mid x_i)}{\mathrm{Var}_0(\eta_i)} < 1$$

for $i = 1, \ldots, p$. Thus a unique commutative solution shall always exist. However, as the posterior variance $\mathrm{Var}_p(\boldsymbol{\eta}; \boldsymbol{x})$ depends upon $\boldsymbol{x}$ in a non-trivial way it will be necessary to evaluate $f(\boldsymbol{x})$ and then sum over $\boldsymbol{x} \in \boldsymbol{X}$ in Equation 6.11 to solve the decision problem. Thus, as we have no explicit form of $f(\boldsymbol{x})$ in our analysis, we shall need an alternative variance if we are to provide a solution.

Use of the mode and curvature of the log-density as in Section 6.3.2 is also unsuitable for this reason.

A second link function

We return to the idea of a pseudo-expectation and variance first explored in Section 3.11. We do not specify the link function explicitly but simply say that it is defined as $\eta_i = g(\theta_i)$ such that the prior 'pseudo-mean' and 'pseudo-variance' are given by

$$\hat{\mathrm{E}}_0(\eta_i) = g_1\left(\frac{a_i}{a_i + b_i}\right), \quad \hat{\mathrm{Var}}_0(\eta_i) = g_2\left(\frac{1}{a_i + b_i}\right),$$

where $g_1()$ and $g_2()$ are suitable monotonic functions. Specifically we take the logit function for g_1 and the identity function for g_2 so that

$$\hat{\mathrm{E}}_0(\eta_i) = \log\left(\frac{a_i}{b_i}\right), \quad \hat{\mathrm{Var}}_0(\eta_i) = \frac{1}{a_i + b_i}.$$

If we observe x_i responses at dose i then the expectation and variance of η_i are updated to

$$\hat{\mathrm{E}}(\eta_i \mid x_i) = \log\left(\frac{a_i + x_i}{b_i + n_i - x_i}\right), \quad \hat{\mathrm{Var}}(\eta_i \mid x_i) = \frac{1}{a_i + b_i + n_i},$$

and so we see that our uncertainty is reduced by the number of observations we make and not by what those observations are.

A Bayes linear kinematic update for $\boldsymbol{\eta}$ can be made each time data are observed using Equations 6.7 and 6.8. Clearly

$$\hat{\mathrm{Var}}(\eta_i \mid x_i) < \hat{\mathrm{Var}}_0(\eta_i)$$

for all i as observing data will lead to positive n_i. This decreases the variance. Thus a unique commutative solution always exists for this choice of link function and it is given by Equations 6.9 and 6.10.

Also, as $\hat{\text{Var}}(\eta_i \mid x_i)$ depends on the data only through n_i, $\hat{\text{Var}}_p(\boldsymbol{\eta}; \boldsymbol{x})$ will not depend explicitly on $\boldsymbol{x}$. Thus the design problem can be solved without full knowledge of $f(\boldsymbol{x})$.

But is the pseudo-approach updating our parameters in a sensible way? To answer this we can return to the plot of $\text{E}(\theta_r \mid X = x)$ for $n = 10$ in the usability experiment, Figure 6.8. We can also plot this adjusted expectation for the pseudo-moment method and compare it to those for the other four methods. A plot of this is given in Figure 6.10.

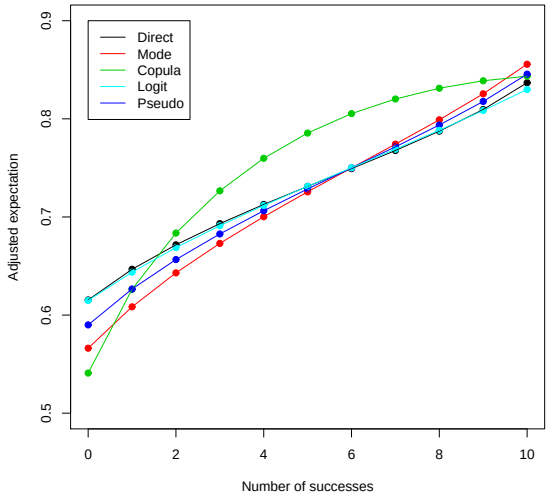

Figure 6.10: A plot of the adjusted expectations of θ_r including the pseudo approach for $n = 10$

We see that the pseudo-moment approach is updating the expectation of θ_r in a very similar way to the direct approach (method 1), the mode and curvature approach (method 2) and the fully Bayesian solution using a logit-Normal prior. Thus, it appears that the pseudo-moment approach is adjusting the parameters in a sensible manner.

6.4.4 Example: Testing the effects of fertilisers on fish

Fertilisers are used worldwide to increase agricultural productivity in both developed and developing nations. They can, however, have an adverse effect on water quality and ultimately lead to the death of aquatic organisms such as fish. This is known as eutrophication.

Chukwu et al (2009) investigated the effect of two fertilisers, NPK 20:10:10 and NPK 12:12:17, and mixtures of these, upon the young (fingerlings) of Oreochromis nitolicus, a fish farmed throughout Africa, South America and South-Eastern Asia. They selected

appropriate doses of each fertiliser and then performed quantal bioassay experiments with the response being the number of fish to die within 4 days of the start of the experiment.

In each experiment at each dose 10 fish were used. For each fertiliser 5 doses were selected:

$$\text{1ml/L, 2ml/L, 4ml/L, 6ml/L and 8ml/L.}$$

Thus 50 fish were used in each experiment. We shall now illustrate our Bayes linear kinematic experimental design procedure.

Suppose that we wish to perform a single experiment with a single fertiliser NPK 20:10:10 and a large experiment contains 50 fish as in Chukwu et al (2009). How many fish should we use and how many should be exposed to each dose?

Prior Elicitation

The elicitation process consists of specifying the prior means and variances of $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_5)'$ and a prior variance matrix $\text{Var}_0(\boldsymbol{\eta})$. The marginal specifications can be achieved by finding values for each a_i and b_i from elicited quantiles.

Initially elicit 3 quantiles for each θ_i, for example the median m_i and upper and lower tertiles t_{2i} and t_{1i}. To do this, questions can be put to the expert in terms of the average proportion of deaths that would be observed over a large number of experiments. As 3 quantiles are being elicited to calculate 2 values (a_i and b_i) there is no exact solution in general. However, we can apply the method of Section 4.5.1 to find suitable parameter values. That is, initially we find exact values of the beta parameters for each combination of 2 of the 3 quantiles;

$$\begin{aligned} (t_{1i}, m_i) &\Rightarrow (a_{1i}, b_{1i}) \\ (m_i, t_{2i}) &\Rightarrow (a_{2i}, b_{2i}) \\ (t_{1i}, t_{2i}) &\Rightarrow (a_{3i}, b_{3i}). \end{aligned}$$

We then find prior means and variances from these on the unrestricted scale via

$$m_{ki} = \log\left(\frac{a_{ki}}{b_{ki}}\right), \quad v_{ki} = \frac{1}{a_{ki} + b_{ki}},$$

for $k = 1, 2, 3$. The prior mean and variance of η_i can then be given as

$$\begin{aligned} m_i &= w_{1i}m_{1i} + w_{2i}m_{2i} + w_{3i}m_{3i} \\ v_i &= \frac{w_{1i}^2}{w_i^2}v_{1i} + \frac{w_{2i}^2}{w_i^2}v_{2i} + \frac{w_{3i}^2}{w_i^2}v_{3i}, \end{aligned}$$

for some weights w_{1i}, w_{2i}, w_{3i} where $w_i^2 = w_{1i}^2 + w_{2i}^2 + w_{3i}^2$. The prior parameter values are found as

$$a_i = \frac{e^{m_i}}{v_i(1+e^{m_i})}, \quad b_i = \frac{1}{v_i(1+e^{m_i})}.$$

In order to elicit covariances between the η_i's we ask the expert to suppose that the value of θ_l is now known for some $l \neq i$ and this has left the value of θ_i unchanged at m_i. New tertiles are then elicited for θ_i in the light of this new information and these are used to find new values $a_i^{'}$ and $b_i^{'}$, where

$$\theta_i \mid \theta_l \sim \text{beta}(a_i^{'}, b_i^{'}).$$

In terms of η_i and η_l this implies that

$$\text{Var}(\eta_l \mid \eta_i) = \frac{1}{a_i^{'} + b_i^{'}}$$

and this gives the prior covariance between η_i and η_l via

$$\frac{\text{Var}(\eta_i \mid \eta_l)}{\text{Var}_0(\eta_i)} = 1 - \text{Corr}_0^2(\eta_i, \eta_l).$$

Thus

$$\text{Cov}_0(\eta_i, \eta_l) = \pm\sqrt{\text{Var}_0(\eta_l)[\text{Var}_0(\eta_i) - \text{Var}(\eta_i \mid \eta_l)]},$$

with the sign given by whether the expert believes their expectation of η_i would increase or decrease upon observation of a higher than expected η_l. This procedure is similar to those developed in Chapter 4 and used in Chapter 5.

However, if we elicit all covariances directly it may be difficult to avoid accidental incoherence as well as prove extremely time consuming. Therefore it may be beneficial to use the following more structured approach. Define

$$\eta_i - \text{E}_0(\eta_i) = F_i,$$

so that F_i is a zero expectation quantity which depends upon dose. We could link these in a first order autoregression

$$F_i = \phi F_{i-1} + \epsilon_i,$$

with $\mathrm{E}(\epsilon_i) = 0$, $\mathrm{Var}(\epsilon_i) = v_\epsilon$ and $\mathrm{Cov}(\epsilon_i, \epsilon_l) = 0$ for $i \neq l$. We could then set the initial variance to

$$\mathrm{Var}_0(F_i) = \frac{v_\epsilon}{1-\phi^2} = v_F,$$

so that the process is stationary. This gives prior covariances

$$\mathrm{Cov}_0(F_i, F_j) = \mathrm{Cov}_0(\eta_i, \eta_j) = \phi^{|j-i|} v_F.$$

Results

Suppose that initially the expert settles on values for the median, upper tertile and lower tertile for each dose as in Table 6.2. This gives values of the beta parameters also given in the table. We use the weights $w_1 = w_2 = w_3 = \frac{1}{3}$ to do this. In order to

	Dose in ml/L				
	1	2	4	6	8
m	0.2	0.5	0.65	0.75	0.9
t_1	0.09	0.37	0.53	0.62	0.73
t_2	0.34	0.64	0.79	0.88	0.97
a	1.970	5.819	6.583	6.122	4.879
b	6.388	6.317	4.184	2.880	1.418

Table 6.2: Initial elicited quantiles and resulting parameter values

find the covariances we set $\phi = 0.93$ and use a stationary variance of $v_F = 0.12$. This necessitates an iterative adjustment of the beta distribution parameter values. New values of a_i and b_i are given in Table 6.3.

	Dose in ml/L				
	1	2	4	6	8
a	1.964	3.996	5.095	5.667	6.457
b	6.369	4.338	3.238	2.666	1.877

Table 6.3: Parameter values adjusted for stationarity of variance

Other values used are the overall cost of an experiment of £20,000, additional costs for each fish at doses $1, \ldots, 5$ of 500, 600, 700, 800, 900 and 1000 (£) respectively, C_{max} =£70,000 and $d_{max} = 400$. In terms of the trade off parameters for cost they were given values $q_1 = q_2 = q_3 = \frac{1}{3}$.

Initially we considered an additive node for the overall utility function with $r_1 = 0.8$, $r_2 = 0.2$ and $r_3 = 0$. This gave an optimal sample allocation of $\boldsymbol{n}^* = (21, 8, 4, 3, 5)$ and hence a sample size of $N = 41$.

Taking a binary node with $r_1 = 0.6$, $r_2 = 0.2$ and $r_3 = 0.2$ produced an optimal sample allocation of $\boldsymbol{n}^* = (19, 6, 3, 2, 3)$ and so an optimal sample size of $N = 33$. It is interesting that in both of these cases, when a maximum sample size of 50 was used, the optimal sample allocation contained fewer then 50 fish. It can be seen that the extra cost of higher doses, both ethical and financial, is resulting in an optimal sample allocation with far fewer fish being given higher doses.

If we use an arbitrary terminal decision of $t =$'report the adjusted expectations of $\boldsymbol{\theta}$' we can see what happens if we observe a certain number of deaths at each dose. For example if, using the additive node and its optimal allocation from above, we observed $\boldsymbol{x} = (3, 5, 3, 3, 5)$ then $\mathrm{E}_5(\boldsymbol{\theta}; \boldsymbol{x}) = (0.23, 0.52, 0.68, 0.76, 0.85)$. If we consider the binary node then observing $\boldsymbol{x} = (3, 3, 2, 1, 3)$ deaths in the optimal allocation would result in $\mathrm{E}_5(\boldsymbol{\theta}; \boldsymbol{x}) = (0.20, 0.45, 0.60, 0.68, 0.78)$. We can compare each of these adjusted expectations with the prior expectation of $\boldsymbol{\theta}$, $\mathrm{E}_0(\boldsymbol{\theta}) = (0.24, 0.48, 0.61, 0.68, 0.77)$.

6.5 Conclusions

In this chapter we investigated the application of the Bayes linear kinematic approach to the design of experiments. Bayesian experimental design has been severely hampered by the necessity, in most realistic models, of approximation, numerical or simulation methods within the maximisations and integrations used to maximise the expected utility and hence solve the design problem. Since, in practice, a large number of such integrations are necessary this has led to severe restrictions in the design problems Bayesian analyes can currently tackle.

The approach developed in this thesis, applied to the design of two types of experiment in this chapter, does not require such numerical integrations or simulation based methods. By giving conjugate prior distributions to the parameters of interest and linking these parameters in a Bayes linear structure all updates have been performed exactly either by fully Bayesian conjugate updating or Bayes linear kinematics.

The two specific applications considered in this chapter were both concerned with binomial counts in which the success probabilities in different groups were considered to be correlated. In the initial usability application there were two probabilities of interest and based on these a decision would be made as to whether to launch or rewrite the product. With just a single observation the Bayes linear kinematic update was straightforward as it was not necessary to check for a commutative update.

A fully Bayesian approach to this problem was adopted by Valks (2005). She used numerical methods to compute the necessary integrals in the calculation of the expected utilities. Our method used no such approximations. The results obtained in our anal-

ysis, the optimal sample size and critical number of successes, were similar to those of Valks (2005). If our method was considered to be an 'approximation' to the 'exact' fully Bayesian solution then, in this analysis at least, the two methods showed good agreement.

In the second application, bioassay, updates over many groups were considered. As our method does not suffer from the computational burdens associated with fully Bayesian approaches we were able to solve both the sample size and design point problems simultaneously. Whereas the benefit utility in the usability application was developed specifically for that problem, in the bioassay application we defined a benefit utility based on information gain. This utility is far more general and could be used in a wide range of applications.

Chapter 7

Conclusions

7.1 Project summary

In this thesis we have concerned ourselves with approaches to Bayesian inference for such things as collections of related binomial and Poisson distributions. Initially, in Chapter 2, we investigated fully Bayesian solutions to the problem in two dimensions, mainly in the form of density multipliers. Two specific types of density multiplier were considered, copula functions and mixtures. We found that the copula family which allowed tractable calculations of posterior distributions to be made, the Farlie-Gumbel-Morgenstern family, had severe restrictions on the prior correlations it was possible to specify. The copula models did, however, preserve marginal distributions allowing prior specifications to be made easily. Using mixtures it was possible to specify any prior correlation between parameters. However, due to the more complex structure of marginal distributions, prior specification was no longer simple even in the two parameter case.

We considered Bayes linear approaches to the two parameter problem in Chapter 3. Initially we considered a model which took advantage of properties of second order exchangeability between Bernoulli trials within a group. We found that this did not fully overcome the problems associated with the known mean-variance relationship in binomial and Poisson distributions. We then applied a modelling approach based on the idea of Bayes linear kinematics, a form of Bayes linear analysis in which changes in belief about some quantities are propagated through to others within a Bayes linear structure. As this is a linear fitting procedure we proposed transforming the binomial and Poisson parameters onto an unrestricted scale before performing the updates. Several transformations were used and the important idea of guide relationships (West *et al.*, 1985) was utilised. It was felt that the transformations led to more effective

updates.

In Chapter 4 we extended the problem to more than two parameters and considered the binomial case. The methodology adopted was the Bayes linear kinematic approach utilising the transformations of the binomial parameters. Conditions for a unique commutative solution to exist were explored and it was found, using the logistic transformation and exact calculation of moments, that a unique commutative solution shall always exist. This is not the case if transformations are not used. We applied the approach to an example in 36 dimensions concerned with the effects of smoking on health. With so many covariances to specify it was found to be useful to utilise ideas from Farrow (2003) to impose a prior covariance structure.

We considered two applications of the transformed Bayes linear kinematic approach to related binomial and Poisson distributions in Chapters 5 and 6. Reliability and survival analysis were the focus of Chapter 5. The two applications to reliability analysis considered were the analyses of failure rates and failure time distributions. The failure rates example was essentially a correlated Poisson distributions problem in more than two parameters. We found that, using a log transformation and exact means and variances for the transformed parameters, a unique commutative Bayes linear kinematic solution exists as long as at least one failure is observed in one of the groups. The other reliability application considered grouped failure times in the form of binomial counts. It was found that taking the complementary log-log transformation is useful in this case for calculation of the reliability. The survival model considered was a piecewise constant hazards model which utilised ideas of system evolution from West *et al.* (1985) for prior specifications. A unique commutative Bayes linear kinematic solution was found in contrast to the non-commutative solution of Gamerman (1991).

Finally, in Chapter 6, we applied our transformed Bayes linear kinematic approach to two problems in the design of experiments; usability testing and bioassay. We solved both problems by maximising expected utility. In the usability testing application we provided a Bayes linear kinematic solution using the same prior specifications and utility functions of Valks (2005) who gave a fully Bayesian solution. We found that our optimal sample size and critical number of users were in close agreement with hers. In the bioassay application we provided a solution to the design problem which answered both of the questions generally considered in Bayesian experimental design; the sample size and design point problems. To do this we introduced the Bayes linear kinematic utility. This requires an adjusted variance which does not explicitly depend on the number of successes observed in the binomial trials.

7.2 Review of objectives

At the start of the thesis the stated aims were to find a methodology for related binomial and Poisson distributions in which

(i) intensive numerical or simulation based methods are not required in the calculation of posterior quantities,

(ii) a careful assessment of genuine prior beliefs can be made for the unknowns in the analysis, and

(iii) realistically complex problems can be solved within a reasonable time frame in the area of Bayesian experimental design.

Both of the fully Bayesian approaches considered, FGM copulas and extensions of these and mixtures, violated condition (ii). In the case of copulas it was not possible to specify strong correlations between parameters which, in general, is a severe restriction. In the case of mixtures, because the marginal distributions for parameters were not in a simple form, finding parameter values which give the prior specifications required would not be a trivial task in non-trivial problems. For mixtures to be widely applicable in this context a general method for prior specification would have to be found.

All of the Bayes linear approaches considered satisfy the first two criteria set out above. In a Bayes linear analysis updating is done by way of a linear fitting procedure. The expectation vector and variance matrix of the quantities of interest are adjusted using standard rules each time anything is observed. Therefore intensive numerical or simulation methods are never necessary and all calculations are tractable.

In a Bayes linear analysis a full second order prior specification is made. That is, for all unknown quantites, expectations, variances and covariances are specified. Thus the Bayes linear approach makes careful assessment of genuine beliefs about relationships between quantities a practical proposition without the imposition of artificial distributional assumptions. Using the Bayes linear kinematic approach unknowns are given conjugate marginal distributions. This allows for quantiles to be elicited and used to set the values of the parameters of the prior distributions. This satisfies condition (ii).

The approach we chose to develop for the remainder of the thesis was Bayes linear kinematics performed on the transformed parameters of the binomial and Poisson distributions. As well as satisfying the first two criteria this approach allowed a Bayes linear analysis to be performed without violating the mean-variance relationship which exists in both the binomial and Poisson distributions. The more standard Bayes linear analysis which took advantage of properties of second order exchangeability did violate these relationships.

Of course we could have chosen to extend the Bayes linear kinematic approach without transformations. We felt, however, that there were two main advantages of the transformed approach. The first was that Bayes linear methods offer a linear fitting procedure. Therefore, as with the case of linear regression, they are most effective when performed on an unrestricted scale. Also, under transformations, we were able to find a unique commutative solution no matter what the observations (or at least almost all of the time) by considering the sufficient condition of Goldstein & Shaw (2004). This was not true in the untransformed case.

Having extended this chosen approach to more dimensions we applied it in several contexts. One such application was the design of experiments and within this bioassay. In this analysis we allocated up to 50 subjects (fish) to 5 different doses and found the expected utility associated with each combination. The Bayes linear kinematic model was able to find the utilities for each of these combinations and find the optimal sample size and allocation. This satisfies condition (iii).

7.3 Future work

This thesis has attempted to provide answers to certain questions. In doing so more have come to light. One involves density multipliers, and in particular mixtures. The theory of mixtures is fairly straightforward and the conjugacy of the joint densities is a very useful property. The problem, as we have already discussed, is the difficulty of prior specification. An area for further work would therefore be to investigate ways to specify prior information in these models. If a general methodology could be put in place to do this then mixtures would provide a tractable fully Bayesian approach to modelling related binomial and Poisson distributions.

An important remaining issue with the Bayes linear kinematic approach is the question of predictive distributions. We have supposed that, after adjustment, we can still use the beta-binomial or gamma-Poisson distribution to find the predictive distribution. Some justification of this is given in Section 3.7.2 but a more formal investigation of predictive distributions for this kind of structure would be informative.

We have examined a Bayes linear kinematic approach to survival analysis for a piecewise constant hazards model. It would be useful to be able to apply our method to a Cox-type proportional hazards model. To do this we could think of each patient at each death time as a Poisson observation Y_i with $y_i = 0$ for all of the patients who do not die at that death time and $y_i = 1$ for the patient who dies. If we give gamma prior distributions to each of the Poisson parameters then we could perform a Bayes linear kinematic analysis.

There are several possible areas of further work in the design of experiments chapter. Within the context of the usability experiment we considered only a single task. This could be extended to multiple tasks with the probability of success in each of the tasks correlated with one another. Having performed the task, the terminal decision was to rewrite or launch the software. Following the rewrite a second usability test could be performed to assess the success of the rewrite. This could be incorporated into the model.

The Taylor expansion approach to specification of the prior covariance for the two different methods is perhaps not the most suitable way to specify the covariance. An alternative method, possibly based on finding the covariance of the transformed parameters which gives a specified rank correlation of the untransformed parameters may be more suitable.

Within the bioassay application we defined the Bayes linear kinematic benefit utility for information gain. This took a similar form to the Bayes linear utility for information gain given by Farrow & Goldstein (2006) who also proposed a mixed Bayes linear benefit utility. We could extend the Bayes linear kinematic benefit utility to create a mixed utility in a similar way.

Bibliography

Aitchison, J. 1986 *The statistical analysis of compositional data*. Chapman and Hall.

Balakrishnan, N. & Lai, C. 2009 *Continuous bivariate distributions: second edition*. Springer.

Bar-Hillel, M. & Neter, E. 1993 How alike is it versus how likely it is: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology* **65**, 1119–1131.

Best, E. W. R. & Walker, C. B. 1964 A Canadian study of smoking and health. *Canadian Journal of Public Health* **55**, 1–14.

Box, G. & Jenkins, G. 1970 *Time series analysis; forecasting and control*. Holden-Day.

Cameron, A. C. & Trivedi, P. K. 1998 *Regression analysis of count data*. Cambridge University Press.

Chaloner, K. M. & Duncan, G. T. 1983 Assessment of a beta prior distribution: pm elicitation. *The Statistician* **32**, 174–180.

Chaloner, K. M. & Verdinelli, I. 1995 Bayesian experimental design: a review. *Statistical Science* **10**, 273–304.

Clayton, D. G. 1978 A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.

Clayton, D. G. 1996 Generalised linear mixed models. In *Markov Chain Monte Carlo in Practice* (ed. W. R. Gilks, S. Richardson & D. J. Spiegelhalter), pp. 275–301. Chapman and Hall.

Congdon, P. 2001 *Bayesian Statistical Modelling*. John Wiley and Sons, Chichester, UK.

Cook, R. & Johnson, M. 1981 A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society. Series B* **43**, 210–218.

Coolen, F. P. A., Goldstein, M. & Munro, M. 2001 Generalised partition testing via bayes linear methods. *Information and Software Technology* **43**, 783–793.

Cox, D. 1972 Regression models and life tables. *Journal of the Royal Statistical Society. Series B* **34**, 187–220.

Davies, O. L. & Goldsmith, P. L. 1972 *Statistical methods in research and production*. ICI Ltd.

De Finetti, B. 1974 *Theory of Probability, vol 1*. Wiley.

De Finetti, B. 1975 *Theory of Probability, vol 2*. Wiley.

DeGroot, M. 1970 *Optimal statistical decisions*. McGraw-Hill.

Dellaportas, P. & Smith, A. F. M. 1993 Bayesian inference for generalised linear and proportional hazards models via Gibbs sampling. *Applied Statistics* **42**, 443–459.

Diaconis, P. & Zabell, S. L. 1982 Updating subjective probability. *Journal of the American Statistical Association* **77**, 822–830.

Doring, F. 1999 Why Bayesian psychology is incomplete. *Philosophy of Science* **66**, S379–S389.

Dumas, J. & Redish, J. 1999 *A practical guide to usability testing*. Intellect books.

Fa-Si-Oen, V. & Pievatolo, A. 2000 The analysis of ruptures in metropolitan gas pipelines. *Quality and Reliability Engineering International* **16**, 17–22.

Farlie, D. J. G. 1960 The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**, 307–323.

Farrow, M. 2003 Building of subjective covariance structures for large complicated systems. *The Statistician* **52**, 553–573.

Farrow, M. & Goldstein, M. 2006 Trade-off sensitive experimental design: a multicriterion, decision theoretic, Bayes linear approach. *Journal of Statistical Planning and Inference* **136**, 498–526.

Field, H. 1978 A note on Jeffrey conditionalization. *Philosophy of Science* **45**, 361–367.

Frank, M. J. 1979 On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aquationes Math* **19**, 194–226.

French, S. & Insua, D. R. 2000 *Statistical decision theory*. Arnold.

Gamerman, D. 1991 Dynamic Bayesian models for survival data. *Applied Statistics* **40**, 63–79.

Gargoum, A. & Smith, J. 1994 Approximating dynamic Gaussian junction trees. *Tech. Rep.* 279. Statistics department, University of Warwick.

Gargoum, A. & Smith, J. 1997 Dynamic generalized linear probability propagation. *Tech. Rep.* 315. Statistics department, University of Warwick.

Garthwaite, P. H., Kadane, J. B. & O'Hagan, A. 2005 Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* **100**, 680–700.

Garthwaite, P. H. & O'Hagan, A. 2000 Quantifying expert opinion in the UK water industry: an experimental study. *The Statistician* **49**, 455–477.

Gavaskar, U. 1988 A comparison of two elicitation methods for a prior distribution of a binomial parameter. *Management Science* **34**, 784–790.

Gittens, J. & Pezeshk, H. 2000 How large should a clinical trial be? *The Statistician* **49**, 177–187.

Goldstein, M. 1986 Exchangeable belief structures. *Journal of the American Statistical Association* **81**, 971–976.

Goldstein, M., Farrow, M. & Spiropoulos, T. 1993 Prediction under the influence: Bayes linear influence diagrams for prediction in a large brewery. *The Statistician* **42**, 445–459.

Goldstein, M. & Shaw, S. 2004 Bayes linear kinematics and Bayes linear Bayes graphical models. *Biometrika* **91**, 425–446.

Goldstein, M. & Wilkinson, D. 2000 Bayes linear analysis for graphical models: The geometric approach to local computation and interpretive graphs. *Statistics and Computing* **10**, 311–324.

Goldstein, M. & Wooff, D. 2007 *Bayes linear statistics*. Wiley.

Gumbel, E. J. 1958 Distributions à plusiers variables dont les marges sont données. *CR Acad Sci Paris Sér I Math* **246**, 2717–2719.

Gumbel, E. J. 1960 Distributions des valeurs extrêmes en plusiers dimensions. *Publ Inst Stat Univ Paris* **9**, 171–173.

Haines, L., Perevozskaya, I. & Rosenberger, W. 2003 Bayesian optimal designs for phase I clinical trials. *Biometrics* **59**, 591–600.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J. & Ostrowski, E. 1994 *A handbook of small datasets*. Chapman and Hall.

Harrison, P. & Stevens, C. 1976 Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society. Series B* **38**, 205–247.

Huang, J. S. & Kotz, S. 1984 Correlation structure in iterated Farlie-Gumbel-Morgenstern distributions. *Biometrika* **71**, 633–636.

Huang, J. S. & Kotz, S. 1999 Modifications of the Farlie-Gumbel-Morgenstern distributions: A tough hill to climb. *Metrika* **49**, 135–145.

Ibrahim, J., Chen, M. & Sinha, D. 2001 *Bayesian survival analysis*. Springer.

Jeffrey, R. C. 1965 *The logic of decisions*. McGraw-Hill.

Johnson, N. L. & Kotz, S. 1977 On some generalized Farlie-Gumbel-Morgenstern distributions-II: Regression, correlation and further generalizations. *Communications in Statistics* **6**, 485–496.

Kadane, J. B. 1998 Experiences in elicitation. *The Statistician* **47**, 3–19.

Kahneman, D. & Tversky, A. 1972 Subjective probability: A judgement of representativeness. *Cognitive Psychology* **3**, 430–454.

Kalbfleisch, J. 1978 Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society. Series B* **40**, 214–221.

Keefer, D. & Bodily, S. 1983 Three-point approximations for continuous random variables. *Management Science* **29**, 595–610.

Keeney, R. L. & Raiffa, H. 1976 *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge university press.

Keeney, R. L. & Raiffa, H. 1993 *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge university press.

Kendall, M. 1938 A new measure of rank correlation. *Biometrika* **30**, 81–89.

Klein, J. P. & Moeschberger, M. L. 1997 *Survival techniques for censored and truncated data*. Springer.

Kruskal, W. 1958 Ordinal measures of association. *Journal of the American Statistical Association* **53**, 814–861.

Kuo, L., Soyer, R. & Wang, F. 1999 Optimal design for quantal bioassay via monte carlo methods. In *Bayesian Statistics 6* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith), pp. 797–802. Oxford University Press.

Lauritzen, S. 1996 *Graphical models*. Oxford science publications.

Lavine, M. & West, M. 1992 A Bayesian method for classification and discrimination. *The Canadian journal of statistics* **20**, 451–461.

Lawless, J. F. 1982 *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, Chichester, UK.

Lin, G. D. 1987 Relationships between two extensions of Farlie-Gumbel-Morgenstern distribution. *AISM* .

Lindley, D. V. 1972 *Bayesian statistics - a review*. SIAM.

Lindley, D. V. 1997 The choice of sample size. *The Statistician* **46**, 129–138.

Martz, H. F., Kvam, P. H. & Abramson, L. R. 1996 Empirical Bayes estimation of the reliability of nuclear-power-plant emergency diesel generators. *Technometrics* **38**, 11–24.

McCullagh, P. & Nelder, J. A. 1989 *Generalized Linear Models*. Chapman and Hall, London.

McLachlan, G. & Peel, D. 2000 *Finite mixture models*. Wiley.

M'Lan, C., Joseph, L. & Wolfson, D. 2008 Bayesian sample size determination for binomial proportions. *Bayesian Analysis* **3**, 269–296.

Monagan, M. B., Geddes, K. O., Heal, K. M., Labahn, G., Vorkoetter, S. M., McCarron, J. & DeMarco, P. 2005 *Maple 10 Programming Guide*. Waterloo ON, Canada: Maplesoft.

Morgenstern, D. 1956 Einfache beispiele zweidmensionaler verteilungen. *Mitteilingsblatt für Mathematische Statistik* **8**, 599–616.

Müller, P. 1999 Simulation-based optimal design. In *Bayesian Statistics 6* (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith), pp. 459–474. Oxford University Press.

Müller, P. 2004 Optimal sample size for multiple testing: the case of gene expression microarrays. *Journal of the American Statistical Association* **99**, 990–1001.

Nelder, J. & Wedderburn, R. 1972 Generalized linear models. *RSSA* **135**, 370–384.

Nelson, R. B. 1999 *An introduction to copulas*. Springer.

Nelson, R. B. 2006 *An introduction to copulas; second edition*. Springer.

Papathomas, M. & O'Hagan, A. 2005 Updating beliefs for binary variables. *Journal of Statistical Planning and Inference* **135**, 324–338.

Payne, S. 1951 *The art of asking questions*. Princeton university press.

Pearson, E. & Tukey, J. 1965 Approximate means and standard deviations based on distances between percentage points of frequency curves. *Biometrika* **52**, 533–546.

Peterson, C. R., Snapper, K. J. & Murphy, A. H. 1972 Credible interval temperature forecasts **53**, 966–970.

Posner, B., Posner, V., Farrow, M. & Brown, A. 1995 Does risk factor reduction correlate with morbidity after CABG? *British Journal of Cardiology* **2**, 8–11.

Posner, B., Posner, V., Farrow, M. & Brown, A. 1996 Morbidity and risk factors after CABG: a second study. *British Journal of Cardiology* **3**, 84–88.

Pratt, J. W., Raiffa, H. & Schlaifer, R. 1995 *Introduction to Statistical Decision Theory*. The MIT press.

R Development Core Team 2011 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Raiffa, H. & Schlaifer, R. 1961 *Applied statistical decision theory*. MIT press.

Revie, M., Bedford, T. & Walls, L. 2010 Evaluation of elicitation methods to quantify Bayes linear models. *Journal of risk and reliability: Special issue on uncertainty* **224**, 322–332.

Schaefer, R. E. & Borcherding, K. 1973 The assessment of subjective probability distributions: a training experiment. *Acta Psychologica* **37**, 117–129.

Schweizer, B. & Wolff, E. F. 1981 On nonparametric measures of dependence for random variables. *The Annals of Statistics* **9**, 879–885.

Settimi, R. & Smith, J. 2000 A comparison of approximate Bayesian forecasting methods for non-Gaussian time series. *Journal of Forecasting* **19**, 135–148.

Slovic, P. 1972 From Shakespeare to Simon: speculations - and some evidence - about man's ability to process information. *Oregon Research Institute Research Bulletin* **12**.

Smith, J. 1979 A generalization of the Bayesian steady forecasting model. *Journal of the Royal Statistical Society. Series B* **41**, 375–387.

Smith, J. 1981 The multiparameter steady model. *Journal of the Royal Statistical Society. Series B* **43**, 256–260.

Smith, J. Q. 1992 *Decision analysis: a Bayesian approach*. Chapman and Hall.

Spearman, C. 1904 The proof and measurement of association between two things. *American Journal of Psychology* **15**, 72–101.

Spiropoulos, T. 1995 Decision support for management using Bayes linear influence diagrams. PhD thesis, University of Sunderland, Chester Road, Sunderland, SR1 3SD, UK.

Tversky, A. & Kahneman, D. 1971 The judgement of frequency and probability by availability of instances. *Oregon Research Institute Research Bulletin* **11**.

Valks, P. 2005 Bayesian decision theoretic approach to experimental design and application to usability experiments. PhD thesis, University of Sunderland, Chester Road, Sunderland, SR1 3SD, UK.

Walker, S. 2003 How many samples?: a Bayesian nonparametric approach. *The Statistician* **52**, 475–482.

West, M., Harrison, P. J. & Migon, H. S. 1985 Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* **80**, 73–83.

Wilkinson, D. J. 1998 *MAS368: Decision making and Bayes linear methods*. Newcastle University.

Wilson, K. & Farrow, M. 2010 Bayes linear kinematics in the analysis of failure rates and failure time distributions. *Journal of Risk and Reliability* **224**, 309–321.

Winkler, R. L. 1967 The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association* **62**, 776–800.

Wooff, D. A. 2000 *[B/D] manual*. Journal of statistical software, http://www.stat.ucla.edu/journals/jss/v05/io2.

Wooff, D. A. & Goldstein, M. 2000 *The Bayes linear programming language [B/D]*. Journal of statistical software, http://www.stat.ucla.edu/journals/jss/v05/io2.