Prediction in Poisson and other Errors in Variables Models

by

Fernando José Malheiro de Magalhães

University of Sheffield

A Thesis submitted to

The University of Sheffield

for the degree

of

Doctor of Philosophy

July 1997

Probability and Statistics Section

School of Mathematics and Statistics

In memory

of

my parents

ACKNOWLEDGEMENTS

To Professor Ian R. Dunsmore, for his supervision, his support and words of encouragement, and for his patience.

To Mr David J. Robson, for his precious help in programming and computing problems.

To my parents, who throughout their lives educated me in a way that made possible to make this course and who taught me to have great enthusiasm in my work. Surely they would love to see this thesis.

To my sister and brother in law, for all the support they gave me, and for all work they did for me solving all kinds of problems during my absence in Portugal.

To all my family and friends in Portugal, who always kept in touch with me, supporting me when I most needed it.

To Instituto Politécnico do Porto and Universidade Portucalense for their financial support during my studies in Sheffield.

To Projecto Modest from PRAXIS XXI / JNICT / FEDER (Portugal).

To Professor Antónia Amaral Turkman, who introduced me to Bayesian Statistics, and who to some extent is responsible for my decision to come to Sheffield, for her friendship and support.

To Professor M. J. Maher (Napier University), who kindly provided the accident data set used to illustrate part of this work.

To all staff and colleagues, for their help during my stay in Sheffield and for the good time we spent together.

To my colleagues and friends Natércia Durão and Cristina Oliveira, who solved all kind of professional problems during my absence from Portugal, and who always were there to encourage me.

To Roland Burkhard, my flatmate, for the good time we spent together, for his patience, support, help, great libertarian thoughts and especially for his friendship.

To all friends I met during the three years I stayed in Sheffield.

PREDICTION IN POISSON AND OTHER ERRORS IN VARIABLES MODELS

by

Fernando José Malheiro de Magalhães

SUMMARY

We want to be able to use information about the traffic flows at road junctions and covariates describing those junctions to predict the number of accidents occurring there. We develop here a Bayesian predictive approach.

Initially we considered three simpler but related problems to assess the efficiency of some approximation techniques, namely:

(I) Given a treatment with an effect that can be described mathematically as of a multiplicative form, we record Poisson countings before and after the treatment is applied. Then, given a new individual with a known counting before the treatment is used, we want to predict the outcome on that individual after the treatment is applied.

(II) After observing the value on an individual before any treatment is applied, we decide, based on that value, which of two treatments to apply, and then register the post-treatment outcome. Given a new individual, with an observed value before he receives any treatment, we aim to derive the predictive distribution for the outcome after one of the treatments is used. (This problem is also considered when several possible treatments are available).

(III) We compare the effects of two treatments, through a two-period crossover design. We assume that both the treatment effect and the period effect are of multiplicative forms.

Estimative and approximation methods are developed for each of these problems. We use the Gibbs sampling approach, normal asymptotic approximations for the posterior distributions and the Laplace approximations. Examples are presented to compare the efficiency and performance of the different methods. We find that the Laplace method performs well, and has computational advantages over the other methods.

Using the knowledge obtained solving these simpler problems we develop solutions for the traffic accidents problem and analyse a real data set. Stepwise procedures for the incorporation of the covariates through the use of Kullback-Leibler measure of divergence are developed.

We also consider the three simpler problems assuming that the observations are exponentially and binomially distributed.

CONTENTS

Page

Page

CHAPTER 8

PREDICTION IN SOME BINOMIAL ERRORS IN VARIABLES MODELS

CHAPTER 1

INTRODUCTION

1.1. Introduction

The final goal of the project is to study a model which, in a Bayesian framework, can be used to predict the number of accidents occurring at a road junction in a given period of time, based on a measurement of the traffic flows at that junction as well as on other covariates describing important characteristics of the junction. Barnett & Wright (1990) and Dunsmore & Robson (1992) considered this problem in a classical framework. In order to achieve this goal initially we study other mathematically related but much simpler problems, with the purpose of analysing how well some estimative or approximative methods work.

The first of those problems, considered in Chapter 2, is what we have called the "multiplicative effect of a treatment problem". In this case, we use observations collected before and after a treatment, which we suppose to have a multiplicative effect, is given to an individual. Then, observing a new individual before the treatment is applied, we derive the predictive distribution for the outcome on that individual after he has received the treatment. In Chapter 2, we suppose that the collected data are Poisson counts. To model this problem, we consider random variables X_i and Y_i, $i=1, 2, ..., n$, representing the observation on the i-th individual before and after the treatment is used, respectively, and suppose that

- given θ_i, $X_i \sim Po(\exp(\theta_i))$,
- given θ_i and α, $Y_i \sim Po(\exp(\alpha + \theta_i))$.

Here α is an unknown parameter which models the multiplicative effect of the treatment, θ_i $(i=1, 2, ..., n)$ are parameters which are needed to model the individuals' characteristics and $Po(\mu)$ represents a Poisson model with mean μ.

As a variation of this problem, in Chapter 3 we consider a situation where two treatments, T_1 and T_2, are available, both with multiplicative, but usually different, effects. The choice of treatments is based on the observation before any treatment is applied. Based on the pairs of observations, before and after treatment, and on a measurement on a new individual who did not receive any treatment, we derive the predictive distributions for the outcome on that new individual, after he has been given either the first or the second treatment. For this "treatment effect under biased allocation problem" with Poisson counts, we consider the model

- given θ_i, $X_i \sim Po(\exp(\theta_i))$,
- given α, β, θ_i and x_i,

$$Y_i \sim Po(\exp(\alpha+\theta_i)), \text{if treatment } T_1 \text{ is used,}$$

$$Y_i \sim Po(\exp(\beta+\theta_i)), \text{if treatment } T_2 \text{ is used,}$$

- given a, T_1 is used if $x_i < a$ and T_2 is used if $x_i \geq a$.

We also generalise this problem to the case when there are more than two possible treatments to use.

The third problem studied is the so called "crossover design to compare two treatments". We suppose we want to compare the effects produced by two possible treatments T_1 and T_2. In order to do that, we give to each individual one treatment in the first period of time, and later, in the second period, we give to that individual, another treatment (either the same or different). We suppose that the effects of the treatments as well as the effects of the periods have multiplicative forms. Based on the observed values on both periods, we derive the predictive distributions which involve the outcome on a new individual who receives either one or other treatment. Supposing that the outcomes are Poisson counts, we solve this problem in Chapter 4. With random variables W_{i1} and W_{i2} representing the outcomes for the i-th individual in the two periods, we use the following statistical model:

Treatment	Period 1	Period 2
T_1	$Po(\exp(\theta_i))$	$Po(\exp(\beta+\theta_i))$
T_2	$Po(\exp(\alpha+\theta_i))$	$Po(\exp(\alpha+\beta+\theta_i))$

Here, the multiplicative treatment effect is modelled by the parameter α and the multiplicative period effect is modelled by the parameter β. The parameters $\theta_1, \theta_2, ..., \theta_n$ are used to model the characteristics of each individual and, therefore, the existent dependence between W_{i1} and W_{i2}.

In Chapters 5 and 6 we consider the accident problem that was the final aim of the project. Let X_{ij} $(i=1, 2, ..., n$ and $j=1, 2, ..., f)$ be random variables representing the j-th traffic flow of the i-th junction and let Y_i be a random variable representing the number of accidents at junction i. We consider $z_{i\ell}$ $(i=1, 2, ..., n$ and $\ell=1, 2, ..., c)$ as being covariates describing the ℓ-th characteristic of the i-th junction. The model considered to predict the number of accidents at a new road junction is

$$X_{ij} \sim Po\left(\exp\left(a_{ij}+k_{ij}\right)\right) \quad , \quad i=1,2,...,n \quad ; \quad j=1,2,...,f,$$

$$Y_i \sim Po\left(\exp\left(\sum_{j=1}^{f}\lambda_j a_{ij}+\sum_{\ell=1}^{c}\beta_\ell z_{i\ell}\right)\right) \quad , \quad i=1,2,...,n$$

where k_{ij} are known constants which relate to the length of the observational period and the time of the day and year that the observations were made. Here, a_{ij} are parameters which are used to model the characteristics of the traffic flows at the junctions; λ and β are vector parameters used to model, respectively, the effect of the traffic flows and of the covariates on the number of accidents. To allow the presentation to be as simple as possible, we begin by considering a situation in which the prediction is based just on two traffic flows (Chapter 5) and then, in Chapter 6, we extend the problem for a general situation with f flows and c covariates in the model.

Finally, in Chapters 7 and 8, we consider the three problems described in Chapters 2, 3 and 4 but supposing that the outcomes are respectively exponentially and binomially distributed.

1.1.1. Notation

Vectors and matrices will be always written in bold type and their dimensions will only be mentioned if they are not obvious. A matrix with n rows and p columns with all elements equal to zero, the null matrix, will be denoted by $\boldsymbol{0}_{np}$ or just $\boldsymbol{0}$. The identity matrix will be denoted by $\boldsymbol{I}_n$ or simply $\boldsymbol{I}$.

1.2. Some Statistical Methods

The basis of all Bayesian methods is to combine the information provided by the data, through the likelihood function $p(\boldsymbol{x} \mid \theta)$, and our beliefs about the model parameters, expressed through the prior distribution $p(\theta)$, in order to derive the posterior distribution $p(\theta \mid \boldsymbol{x})$, using the Bayes' theorem. The posterior distribution is given by

$$p(\theta \mid \boldsymbol{x}) = \frac{p(\theta)\, p(\boldsymbol{x} \mid \theta)}{\int p(\theta)\, p(\boldsymbol{x} \mid \theta) d\theta}$$

or, up to a normalising constant,

$$p(\theta \mid \boldsymbol{x}) \propto p(\theta)\, p(\boldsymbol{x} \mid \theta).$$

Then, the posterior distribution is used to evaluate the predictive distribution of a random variable Y; such a predictive distribution is given by

$$p(y \mid \boldsymbol{x}) = \int p(y \mid \theta) p(\theta \mid \boldsymbol{x}) d\theta.$$

Note that the predictive distribution is then the posterior expectation of the probability distribution $p(y \mid \theta)$. Details about posterior and predictive distributions can be found, for example, in Aitchison & Dunsmore (1975) or Geisser (1993).

In the models we have developed we often find that the posterior distribution is not analytically tractable. Therefore, the practical implementation of our Bayesian methods is quite difficult and requires a great amount of computation. We use some methods which summarise, at least approximately, the important characteristics of the posterior distribution. Bayesian inference summaries often require the evaluation of posterior expected values of functions. Due to the difficult integrals that are usually involved in those expectations, we are in almost all practical situations unable to solve them analytically and we are therefore forced to use numerical methods. But, if the problem has a high dimensionality, those numerical methods are usually not efficient enough. One possible way to overcome that problem is to provide approximations for the integrals.

The next sections outline some of those methods.

1.2.1. The Gibbs Sampling Algorithm

The Gibbs Sampling Algorithm is a Markov Chain Monte Carlo method that can be used to simulate the posterior distribution $p(\theta \mid x)$. The summary presented here is based on the work by Bernardo & Smith (1994), O'Hagan (1994), Dellaportas & Smith (1993), Gelfand & Smith (1990) and Gelfand, Hills, Racine-Poon & Smith (1990).

Let $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ be a parameter vector and suppose that we want to summarise its posterior distribution $p(\theta \mid x)$.

If $p(\theta \mid x)$ is too complex for analytical treatment, it is quite likely that it will be very difficult to generate a random sample from that posterior distribution directly. However, if the full conditional distributions $p(\theta_i \mid \theta_{j \neq i}, x)$ for each individual component of θ, given the data x and specified values of all other components of θ, are easily obtained from $p(\theta \mid x)$, it is sometimes easier to generate samples from them instead. Let us suppose that we are able to generate θ_i from the full conditional distribution $p(\theta_i \mid \theta_{j \neq i}, x)$. In that case, the Gibbs sampling algorithm consists of considering an arbitrary set of starting values $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)})$ and of generating a series of random k-dimensional points $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \ldots$, where $\theta^{(p+1)}$ is derived from $\theta^{(p)}$ in the following way:

- generate $\theta_1^{(p+1)}$ from $p(\theta_1 \mid \theta_2^{(p)}, \theta_3^{(p)}, \ldots, \theta_k^{(p)}, x)$;
- generate $\theta_2^{(p+1)}$ from $p(\theta_2 \mid \theta_1^{(p+1)}, \theta_3^{(p)}, \ldots, \theta_k^{(p)}, x)$;
- generate $\theta_3^{(p+1)}$ from $p(\theta_3 \mid \theta_1^{(p+1)}, \theta_2^{(p+1)}, \theta_4^{(p)} \ldots, \theta_k^{(p)}, x)$; (1.1)

$$\vdots$$

- generate $\theta_k^{(p+1)}$ from $p\left(\theta_k \mid \theta_1^{(p+1)}, \theta_2^{(p+1)}, \ldots, \theta_{k-1}^{(p+1)}, x\right)$.

It has been shown that under mild conditions, if enough iterations are done, this process will converge (see, for example, O'Hagan (1994)). Thus, after t iterations we will obtain a sample $\left(\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_k^{(t)}\right)$ from $p(\theta \mid x)$. Repeating the whole procedure M times, we will obtain M samples from $p(\theta \mid x)$, namely

$$\theta_{(j)}^{(t)} = \left(\theta_{1(j)}^{(t)}, \theta_{2(j)}^{(t)}, \ldots, \theta_{k(j)}^{(t)}\right) \quad , \quad j = 1, 2, \ldots, M$$

which can be used to estimate all summaries about $p(\theta \mid x)$ in which we are interested.

For instance, we can estimate the mean vector of θ by

$$\bar{\theta} = \frac{1}{M} \sum_{j=1}^{M} \theta_{(j)}^{(t)}.$$

More generally, if $g(\theta)$ is any function of θ, its expectation can be estimated by

$$\hat{E}[g(\theta)] = \frac{1}{M} \sum_{j=1}^{M} g\left(\theta_{(j)}^{(t)}\right).$$

In particular, noting that the predictive distribution of a random variable Y can be regarded as the posterior expected value of $p(y \mid \theta)$, we can estimate such a predictive distribution by

$$\hat{p}(y \mid x) = \frac{1}{M} \sum_{j=1}^{M} p\left(y \mid \theta_{(j)}^{(t)}\right) \tag{1.2}$$

The Gibbs sampling algorithm supposes that we are able to sample from the full conditional distributions. In some situations the random generation from $p\left(\theta_i \mid \theta_{j \neq i}, x\right)$ is not straightforward, and we will then have to use techniques such as those summarised later in section 1.2.3.

1.2.2. Asymptotic Results

Supposing that n, the dimension of the available data set, is large enough, the posterior distribution $p(\theta \mid x)$ can be successfully approximated by a k-multivariate normal distribution, and therefore the summaries we are interested in will be readily obtained.

We will consider two different asymptotic normal approximations for the posterior distribution $p(\theta \mid x)$. The first of them is based only on the likelihood function (Bernardo & Smith, 1994) and the other is based on characteristics of the posterior distribution itself (O'Hagan, 1994).

1.2.2.1. Posterior Normality Based on the Likelihood Function

Let us consider the continuous parameter vector $\theta \in \Theta \subseteq \Re^k$ and let $p(\theta \mid x)$ be its posterior distribution. By the definition of $p(\theta \mid x)$,

$$p(\theta \mid x) \propto p(\theta) p(x \mid \theta);$$

or, equivalently, we can write

$$p(\theta \mid x) \propto \exp\{\ln p(\theta) + \ln p(x \mid \theta)\}.$$

Considering both $p(\theta)$ and $p(x \mid \theta)$ as functions of θ, and solving the systems of equations

$$\nabla \ln p(\theta) = 0 \quad \text{and} \quad \nabla \ln p(x \mid \theta) = 0,$$

separately, we will obtain the prior mode m_0 and the maximum likelihood estimate $\hat{\theta}$. Then, let us expand $\ln(p(\theta))$ and $\ln(p(x \mid \theta))$ in Taylor series around m_0 and $\hat{\theta}$, respectively,

$$\ln p(\theta) = \ln p(m_0) - \frac{1}{2}(\theta - m_0)^T H_0 (\theta - m_0) + R_1$$

$$\ln p(x \mid \theta) = \ln p\left(x \mid \hat{\theta}\right) - \frac{1}{2}\left(\theta - \hat{\theta}\right)^T H\left(\hat{\theta}\right)\left(\theta - \hat{\theta}\right) + R_2$$

where R_1 and R_2 are the remainder terms and the matrices H_0 and $H\left(\hat{\theta}\right)$ are defined by

$$H_0 = \left(-\frac{\partial^2 \ln p(\theta)}{\partial\theta_i \partial\theta_j}\right)_{\mid \theta = m_0} \quad \text{and} \quad H\left(\hat{\theta}\right) = \left(-\frac{\partial^2 \ln p(x \mid \theta)}{\partial\theta_i \partial\theta_j}\right)_{\mid \theta = \hat{\theta}}$$

Considering an exact normal prior distribution, R_1=0; however, there is no need of such a requirement in order to obtain a good approximation because under some regularity conditions (Bernardo & Smith, 1994, pp. 287-292), as n becomes large, the remainder terms R_1 and R_2 become negligible, no matter the prior assumptions considered. Hence, we can write that, approximately,

$$p(\theta \mid x) \propto \exp\left\{-\frac{1}{2}\left[(\theta - m_0)^T H_0 (\theta - m_0) + \left(\theta - \hat{\theta}\right)^T H\left(\hat{\theta}\right)\left(\theta - \hat{\theta}\right)\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}(\theta - m_1)^T H_1 (\theta - m_1)\right\}$$

with

$$H_1 = H_0 + H\left(\hat{\theta}\right)$$

$$m_1 = H_1^{-1}\left(H_0 m_0 + H\left(\hat{\theta}\right)\hat{\theta}\right).$$

This shows that for large n, the posterior distribution $p(\theta \mid x)$ can be approximated by a k-multivariate normal distribution with mean vector m_1 and precision matrix H_1.

Intuitively, we expect that as n becomes larger, the prior precision H_0 will become small when compared with the precision provided by the data, $H\left(\hat{\theta}\right)$, and that approximately $H_1 = H\left(\hat{\theta}\right)$ and $m_1 = \hat{\theta}$. Therefore, we can assume that

$$p(\theta \mid x) \cong N_k\left(\hat{\theta}, H^{-1}\left(\hat{\theta}\right)\right) \tag{1.3}$$

where $\hat{\theta}$ is the vector of the maximum likelihood estimates of the parameters and $H(\hat{\theta})$ is the Hessian matrix evaluated at $\hat{\theta}$.

1.2.2.2. Posterior Normality Based on Characteristics of the Posterior Distribution

This asymptotic result, presented by O'Hagan (1994), is similar to the one described in the former section. Given the posterior distribution $p(\theta \mid x)$, we evaluate its mode $\boldsymbol{m}$, solving the system of equations defined by

$$\nabla \ln p(\theta \mid x) = \boldsymbol{0} \tag{1.4}$$

and we expand $\ln p(\theta \mid x)$ in a Taylor series around the posterior mode $\boldsymbol{m}$, ignoring terms involving third and higher order derivatives. Therefore, defining the modal dispersion matrix $\boldsymbol{V}$ such that

$$\boldsymbol{V}^{-1} = \left(-\frac{\partial^2 \ln p(\theta \mid x)}{\partial \theta_i \partial \theta_j} \right)_{|\theta = \boldsymbol{m}}, \tag{1.5}$$

we easily conclude that

$$p(\theta \mid x) \approx N_k(\boldsymbol{m}, \boldsymbol{V}) \tag{1.6}$$

1.2.3. The Rejection Sampling Algorithm

If the conditional distributions are not standard ones from which samples can be easily generated, techniques such as rejection sampling algorithms (Gilks & Wild, 1992) have been developed.

1.2.3.1. Non-Adaptive Rejection Sampling

Let us suppose that we need to generate n random points, independently, from a distribution $p(x)$, known up to a constant of integration. To obtain those n random points, we define an envelope function $p_e(x)$, which is easy to generate from, such that $p_e(x) \geq p(x)$, for all x in the domain of $p(x)$. Because the envelope function has also to be defined just up to a constant of integration, we can choose $p_e(x)$ to be of a similar shape to $p(x)$; then, to ensure that the envelope function really covers $p(x)$, we can multiply it by a constant $M \geq 1$.

We begin by generating a point x_0 from $p_e(x)$ or, more specifically, from $Mp_e(x)$; then, we generate a value y_0 from the uniform $(0,1)$ distribution and finally, if

$$y_0 M p_e(x_0) \leq p(x_0),$$

we accept x_0 as being a random point from $p(x)$; otherwise, we reject the point x_0.

In order to increase the efficiency or acceptance rate, the value of M has to be appropriately selected. It has been shown that the proportion of accepted values is $1/M$. Therefore, it is obvious that we wish M to be as small as possible. As the envelope function $M p_e(x)$ was built in such a way that $p(x) \leq Mp_e(x)$, this implies that the best choice of M is

$$M = \sup_x \frac{p(x)}{p_e(x)} \tag{1.7}$$

The efficiency of this algorithm depends very much on the art of choosing the envelope function, which can sometimes be extremely difficult.

1.2.3.2. Adaptive Rejection Sampling

Adaptive rejection sampling is a successful attempt to improve the efficiency of the non-adaptive rejection algorithm. Full details of this algorithm as well as its proof are presented by Gilks & Wild (1992). Here we will just present a brief summary of this sampling method.

Adaptive rejection sampling can be applied if and only if some conditions are satisfied: the domain D of the random variable X must be connected; $p(x)$ must be a continuous function differentiable everywhere in D; and $h(x)=\ln p(x)$ must be concave everywhere in D, that is, $h'(x)$ must decrease monotonically as x increases in D. An example of a concave $h(x)$ in D is shown in Figure 1.1.

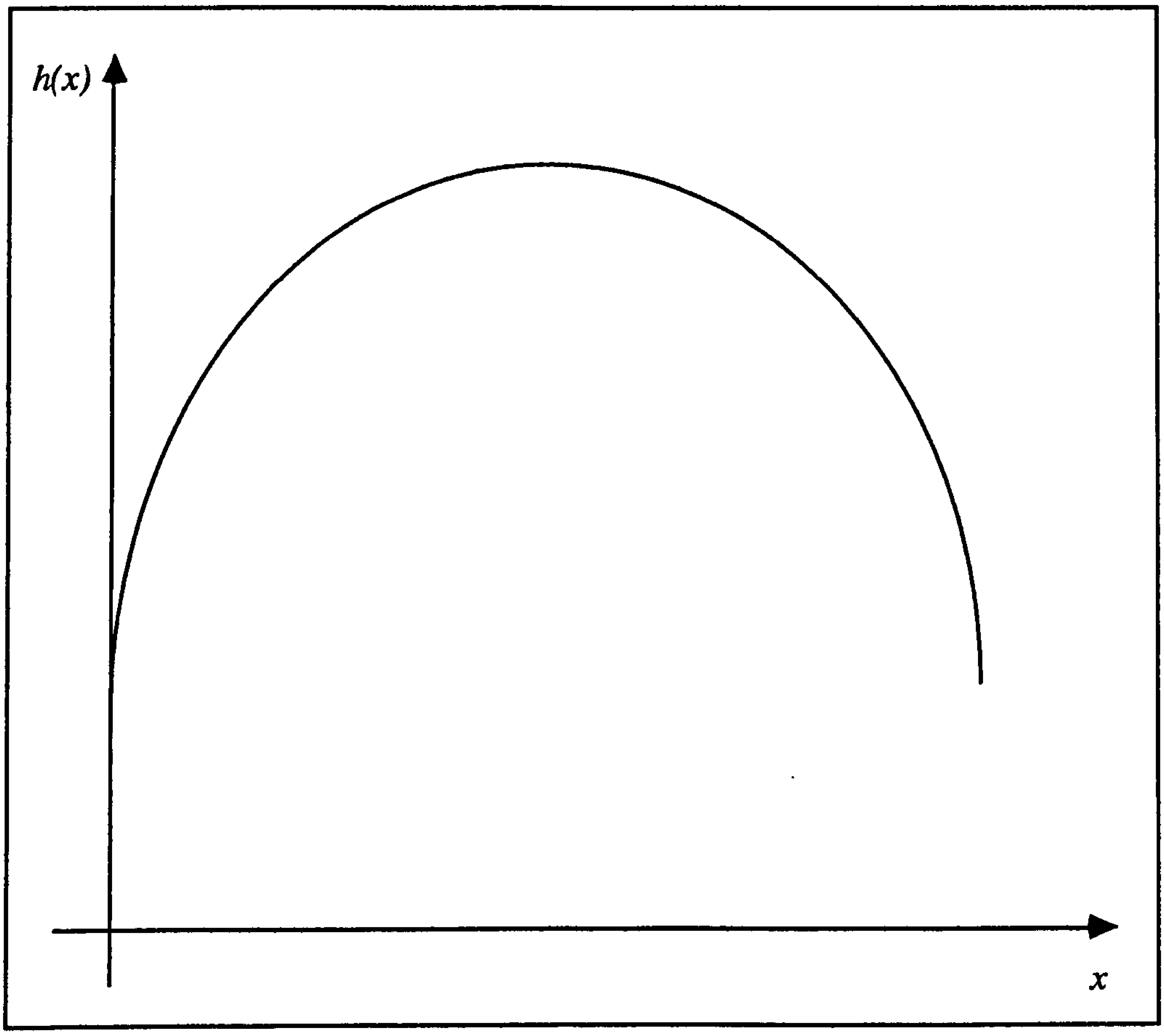

Figure 1.1: An example of a concave function $h(x)$.

The basic idea of the adaptive rejection sampling algorithm is to build an efficient envelope function for $p(x)$, based on an envelope function for $h(x)=\ln p(x)$.

Let us consider a set of k abscissae $T_k = \{x_i, i = 1, 2, \ldots, k\}$ such that $x_1 \leq x_2 \leq \ldots \leq x_k$, and let us suppose that $h(x)$ and $h'(x)$ have been evaluated at the k points of T_k. If, at each of these points, we define the tangents to $h(x)$, we can define a piecewise linear upper hull, $u_k(x)$, as shown by Figure 1.2. Each of those tangents intersect with the next one, defining abscissae z_j, $j=1, 2, ..., k-1$, given by

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})}. \tag{1.8}$$

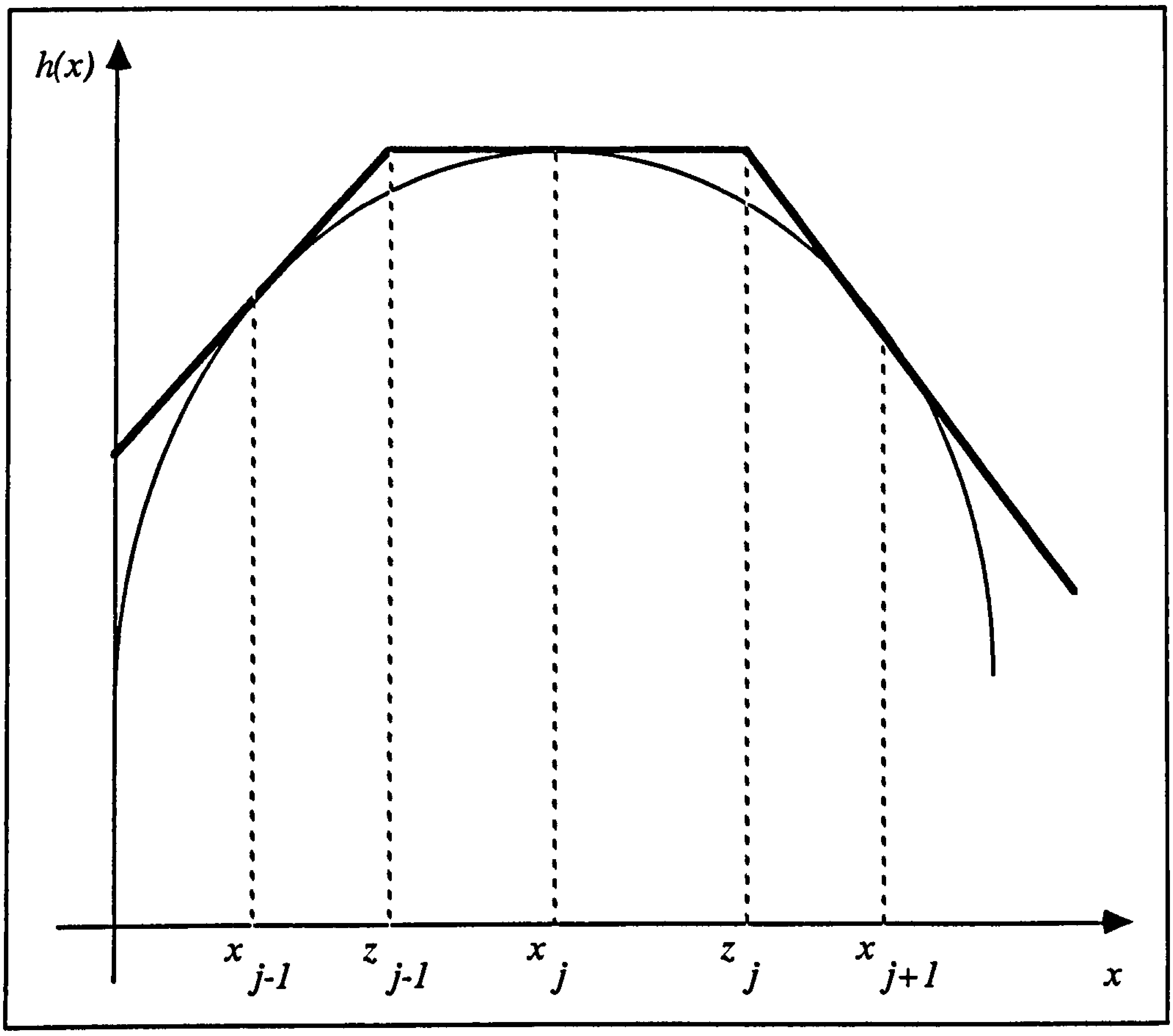

Figure 1.2: A piecewise linear upper hull for *h(x)*.

For $x_j \in [z_{j-1}, z_j]$, $j=1, 2, ..., k$, the piecewise linear functions $u_k(x)$ which form the upper hull are obviously defined by

$$u_k(x) = h(x_j) + (x - x_j)h'(x_j), \tag{1.9}$$

where z_0 is the lower bound of D (or $-\infty$ if D is not bounded below) and z_k is the upper bound of D (or $+\infty$ if D is not bounded above).

Noting that $h(x)=\ln p(x)$ and $u_k(x)$ is an upper hull for $h(x)$, it is natural to define

$$s_k(x) = \frac{\exp\{u_k(x)\}}{\int_D \exp\{u_k(x)\}\, dx} \tag{1.10}$$

to form the envelope function for $p(x)$.

With the aim of improving the algorithm, we can also define a piecewise linear lower hull, $l_k(x)$, formed by the chords between adjacent abscissae in T_k, as shown in Figure 1.3, and given by

$$l_k(x) = \frac{(x_{j+1} - x)h(x_j) + (x - x_j)h(x_{j+1})}{x_{j+1} - x_j}, \tag{1.11}$$

for $j=1, 2, ..., k-1$. For $x < x_1$ or $x > x_k$, we define $l_k(x) = -\infty$. Using this lower hull, and in a similar way as in (1.10), we define a squeezing function on T_k for $p(x)$ as $\exp\{l_k(x)\}$.

To sample n points independently from $p(x)$ by adaptive rejection, we have to perform the following algorithm:

1. Initialise the abscissae in T_k, choosing arbitrarily k points. If the domain D is not bounded below, x_1 must be chosen such that $h'(x_1) > 0$; if D is not bounded above, x_k is chosen such that $h'(x_k) < 0$.

2. Calculate the functions $u_k(x)$, $s_k(x)$ and $l_k(x)$ from (1.9), (1.10) and (1.11), respectively.

3. Generate a value x_0 from $s_k(x)$ and sample a value y_0 from the uniform $(0,1)$ distribution.

4. If

$$y_0 \le \exp\{l_k(x_0) - u_k(x_0)\}$$

then accept x_0; otherwise, evaluate $h(x_0)$ and $h'(x_0)$ and if

$$y_0 \le \exp\{h(x_0) - u_k(x_0)\}$$

then accept x_0; otherwise, reject x_0.

5. If $h(x_0)$ and $h'(x_0)$ were evaluated at the former step, x_0 is included in T_k, defining the set T_{k+1}; the elements of T_{k+1} are re-ordered in ascending order and we define the functions $u_{k+1}(x)$, $s_{k+1}(x)$ and $l_{k+1}(x)$; finally, we return to step 3 until n points have been accepted.

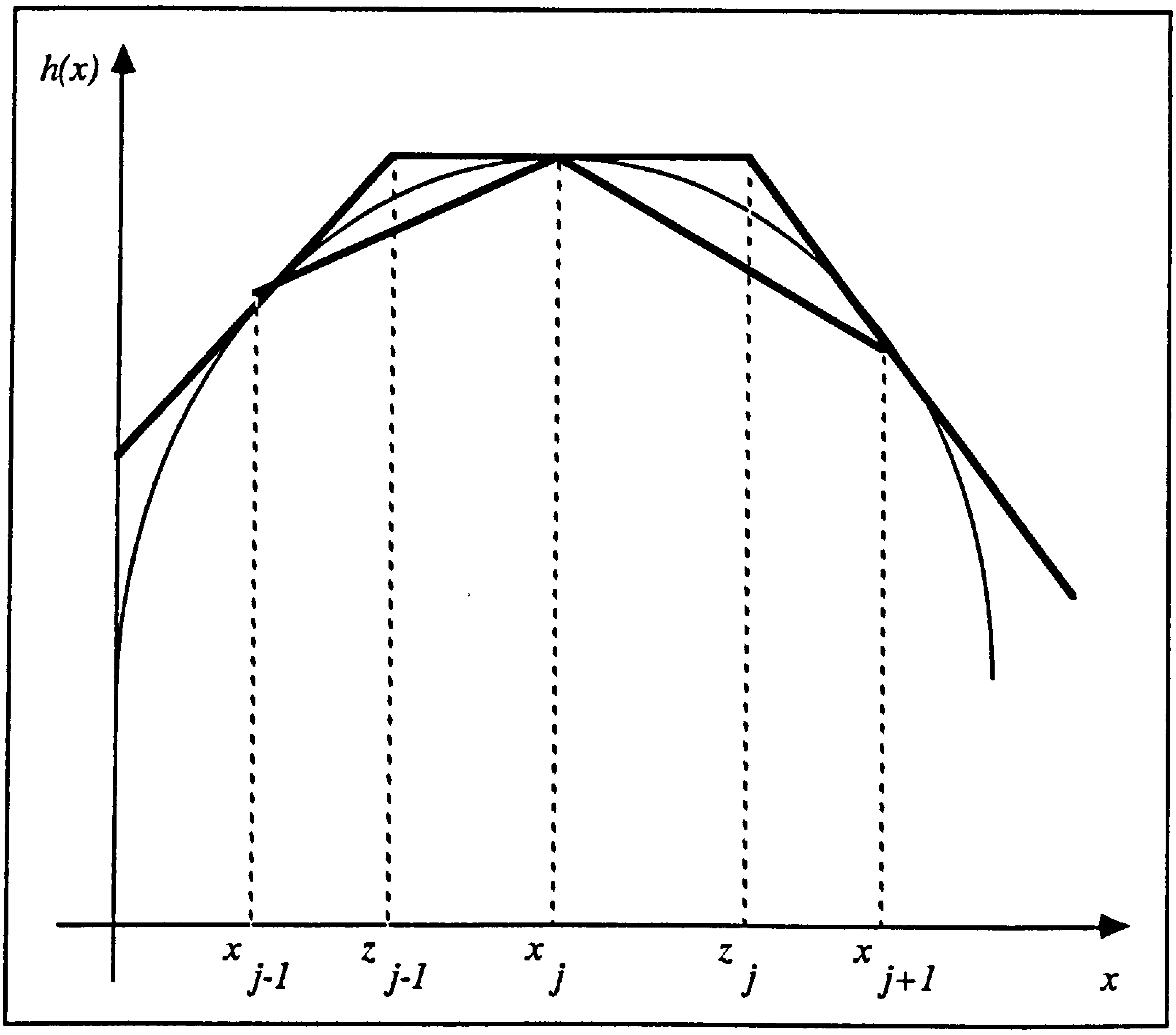

Figure 1.3: A piecewise linear upper hull and a piecewise linear lower hull for $h(x)$.

It is interesting to note that as this algorithm is performed, the number of abscissae considered increases and become closer. This will make the envelope and the squeezing functions to be each time closer to $p(x)$, increasing the efficiency of the algorithm, that is, increasing the proportion of accepted values.

1.2.4. The Laplace Approximation

An alternative approach to approximating the predictive distributions will be through the use of the Laplace approximation (see, for example, Bernardo & Smith (1994)). A summary of this method is presented here.

Given a parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ and a data set x, let $g(\theta)$ be a real function of the parameter vector θ and suppose that we are interested in its posterior expectation

$$E[g(\theta) \mid x] = \int g(\theta) p(\theta \mid x) d\theta$$

$$= \frac{\int g(\theta) p(\theta) p(x \mid \theta) d\theta}{\int p(\theta) p(x \mid \theta) d\theta} \qquad (1.12)$$

If we define two functions $h(\theta)$ and $h^*(\theta)$ such that

$$-n h(\theta) = \ln p(\theta) + \ln p(x \mid \theta) \qquad (1.13)$$

and

$$-n h^*(\theta) = \ln g(\theta) + \ln p(\theta) + \ln p(x \mid \theta), \qquad (1.14)$$

the expectation (1.12) can be written in the alternative form

$$E[g(\theta) \mid x] = \frac{\int \exp\{-n h^*(\theta)\} d\theta}{\int \exp\{-n h(\theta)\} d\theta} \; . \qquad (1.15)$$

Let us now define $\tilde{\theta}$, θ^*, $\tilde{\sigma}$ and σ^* such that

$$-h(\tilde{\theta}) = \sup_\theta\{-h(\theta)\},$$

$$-h^*(\theta^*) = \sup_\theta\{-h^*(\theta)\}, \qquad (1.16)$$

$$\tilde{\sigma} = \left|n\nabla^2 h(\tilde{\theta})\right|^{-1/2},$$

$$\sigma^* = \left|n\nabla^2 h^*(\theta^*)\right|^{-1/2},$$

where $\nabla^2 h$ and $\nabla^2 h^*$ are, respectively, the Hessian matrices of h and h^*. Expanding $h(\theta)$ in a Taylor series around $\tilde{\theta}$ and ignoring all terms with order superior to two, we obtain the following approximation for $h(\theta)$:

$$h(\theta) \approx h(\tilde{\theta}) + (\theta - \tilde{\theta})^T \nabla h(\tilde{\theta}) + \frac{1}{2}(\theta - \tilde{\theta})^T \nabla^2 h(\tilde{\theta})(\theta - \tilde{\theta}).$$

But, by the definition in (1.16) for $\tilde{\theta}$, $\nabla h(\tilde{\theta}) = \mathbf{0}$, and thus,

$$h(\theta) \approx h(\tilde{\theta}) + \frac{1}{2}(\theta - \tilde{\theta})^T \nabla^2 h(\tilde{\theta})(\theta - \tilde{\theta}).$$

The denominator of (1.15) can then be approximated by

$$\int \exp\{-n\,h(\theta)\}d\theta \approx \exp\{-nh(\tilde{\theta})\}\int \exp\left\{-\frac{1}{2}(\theta - \tilde{\theta})^T \boldsymbol{H}^{-1}(\theta - \tilde{\theta})\right\}d\theta$$

with $\boldsymbol{H}^{-1} = n\nabla^2 h(\tilde{\theta})$. The argument of the integral appearing in the right-hand side of this approximation has the form of a multivariate normal distribution $N_k(\tilde{\theta}, \boldsymbol{H})$ and therefore, using the definition in (1.16) for $\tilde{\sigma}$, we have

$$\int \exp\left\{-\frac{1}{2}(\theta - \tilde{\theta})^T \boldsymbol{H}^{-1}(\theta - \tilde{\theta})\right\}d\theta = (2\pi)^{k/2}\tilde{\sigma},$$

and so the denominator in (1.15) becomes

$$\int \exp\{-n\,h(\theta)\}d\theta \approx (2\pi)^{k/2}\,\tilde{\sigma}\,\exp\{-nh(\tilde{\theta})\}.$$

Following an analogous development, the numerator in (1.15) is

$$\int \exp\left\{-nh^*(\theta)\right\} d\theta = (2\pi)^{k/2}\, \sigma^*\, \exp\left\{-nh^*\left(\theta^*\right)\right\}$$

and thus, the Laplace approximation for $E\left[g(\theta) \mid x\right]$ is given by

$$E\left[g(\theta) \mid x\right] \cong \left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\left\{-nh^*\left(\theta^*\right) + nh\left(\tilde{\theta}\right)\right\}. \qquad (1.17)$$

CHAPTER 2

THE MULTIPLICATIVE EFFECT OF A TREATMENT: PREDICTION IN A POISSON ERRORS IN VARIABLES MODEL

Suppose we have independent random variables (X_1,Y_1), (X_2,Y_2), ..., (X_n,Y_n), where X_i and Y_i are Poisson counts observed on individual i before and after a treatment is used, respectively. We also suppose that the treatment has an effect upon the individuals which can be expressed in a multiplicative form. To describe such a situation, we consider the following statistical model

- given θ_i, $X_i \sim Po(\exp(\theta_i))$,
- given θ_i and α, $Y_i \sim Po(\exp(\alpha + \theta_i))$.

Here α is an unknown parameter which models the multiplicative effect of the treatment and θ_i $(i = 1, 2, \ldots, n)$ are parameters which are used to model the individuals' characteristics. We use this form of parameterisation, with log-link functions for the mean values, because the parameter values are then unrestricted real numbers, which is a requirement necessary for some of the estimative and approximative methods we will use.

Our aim is to make predictions for a future Y_{n+1} based on an observed x_{n+1} and on a data set $D^n = \{(x_i, y_i),\ i = 1, 2, \ldots, n\}$; that is, we want to derive the predictive distribution of the random variable Y_{n+1}.

Dunsmore & Robson (1997) consider this model in a different parametric framework. Here we develop and compare various approximations.

2.1. A Classical Approach

The maximum likelihood estimates of the parameters can be used, in a classical framework, to derive a simple estimative approximation for the predictive distribution of Y_{n+1}. Let us suppose that, given θ_{n+1}, $X_{n+1} \sim Po(\exp(\theta_{n+1}))$, and to simplify the notation we define a parameter vector $\theta^n = (\theta_1, \theta_2, \ldots, \theta_n)$. The likelihood function will be

$$L(\theta^n, \theta_{n+1}, \alpha;\, D^n, x_{n+1}) = \prod_{i=1}^{n+1}\{p(x_i \mid \theta_i)\}\prod_{i=1}^{n}\{p(y_i \mid \theta_i, \alpha)\}$$

$$\propto \exp\left\{-\sum_{i=1}^{n} e^{\theta_i}(1+e^{\alpha}) - e^{\theta_{n+1}}\right\}\times$$

$$\times \exp\left\{\sum_{i=1}^{n}\theta_i(x_i+y_i) + \alpha\sum_{i=1}^{n} y_i + \theta_{n+1}x_{n+1}\right\},$$

and thus

$$\ln L(\theta^n, \theta_{n+1}, \alpha;\, D^n, x_{n+1}) = \ln(const) - \sum_{i=1}^{n} e^{\theta_i}(1+e^{\alpha}) - e^{\theta_{n+1}} + \sum_{i=1}^{n}\theta_i(x_i+y_i) +$$

$$+\alpha\sum_{i=1}^{n} y_i + \theta_{n+1}x_{n+1},$$

leading to the maximum likelihood estimates

$$\hat{\theta}_{n+1} = \ln(x_{n+1}), \quad \hat{\alpha} = \ln\left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}\right), \quad \hat{\theta}_i = \ln\left(\frac{x_i + y_i}{1+e^{\hat{\alpha}}}\right),\ i = 1, 2, \ldots, n. \quad (2.1)$$

Thus, the predictive distribution for Y_{n+1} can be approximated, in a classical framework, by

$$p(y_{n+1} \mid \hat{\theta}_{n+1}, \hat{\alpha}) = Po(\exp(\hat{\alpha} + \hat{\theta}_{n+1})).$$

Note that this classical approach will be impossible to implement if $x_{n+1} = 0$.

2.2. A Bayesian Approach

2.2.1. The Exact Predictive Distribution

The basis of the Bayesian approach is to consider prior distributions for the parameters, derive the posterior distribution using the information in the observed data set and obtain the predictive distribution for Y_{n+1}, given by

$$p(y_{n+1} \mid D^n, x_{n+1}) = \int_{\Re^2} p(y_{n+1} \mid \theta_{n+1}, \alpha) p(\theta_{n+1}, \alpha \mid D^n, x_{n+1}) \, d\theta_{n+1} \, d\alpha . \qquad (2.2)$$

We will consider a hierarchical prior structure. At the first stage, we take

$$p(\theta^n, \theta_{n+1}, \alpha \mid \xi, \eta) = \prod_{i=1}^{n+1} \{p(\theta_i \mid \xi)\} p(\alpha \mid \eta),$$

whilst at the second stage we assume

$$p(\xi, \eta) = p(\xi) \, p(\eta).$$

An appropriate structure here would be of the form

$$e^{\theta_i} \sim Ga(k, e^{\xi}) \qquad e^{\alpha} \sim Ga(g, e^{\eta})$$

$$e^{\xi} \sim Ga(u, v) \qquad e^{\eta} \sim Ga(r, s)$$

where $Ga(a, b)$ represents a gamma distribution with density function proportional to $\psi^{a-1}\exp\{-b\psi\}, \psi > 0$, and where k, g, u, v, r and s are known constants.

Considering the equivalent forms in Table A1.1, the joint prior distribution for $(\theta^n, \theta_{n+1}, \alpha, \xi, \eta)$ will be given by

$$p(\theta^n, \theta_{n+1}, \alpha, \xi, \eta) = \prod_{i=1}^{n+1} \{p(\theta_i \mid \xi)\} p(\alpha \mid \eta) p(\xi) p(\eta)$$

$$\propto \exp\left\{-\sum_{i=1}^{n+1} e^{\xi+\theta_i} - e^{\eta+\alpha} - ve^{\xi} - se^{\eta}\right\}\times$$

$$\times \exp\left\{\sum_{i=1}^{n+1} k(\xi+\theta_i) + g(\eta+\alpha) + u\xi + r\eta\right\}$$

and the joint posterior distribution is then given by

$$p\left(\theta^n, \theta_{n+1}, \alpha, \xi, \eta \mid D^n, x_{n+1}\right) \propto \exp\left\{-\sum_{i=1}^{n} e^{\theta_i}\left(1+e^{\alpha}+e^{\xi}\right) - e^{\theta_{n+1}}\left(1+e^{\xi}\right) - ve^{\xi}\right\}\times$$

$$\times\exp\left\{-e^{\eta}\left(e^{\alpha}+s\right) + \sum_{i=1}^{n}\theta_i\left(x_i+y_i+k\right) + \theta_{n+1}\left(x_{n+1}+k\right)\right\}\times$$

$$\times\exp\left\{\alpha\left(g+\sum_{i=1}^{n} y_i\right) + \eta(g+r) + ((n+1)k+u)\xi\right\}. \qquad (2.3)$$

From (2.2) all we need in order to derive the predictive distribution is the marginal posterior distribution of $\left(\theta_{n+1}, \alpha\right)$. Whilst we can derive the marginal posterior distribution for $\left(\theta_{n+1}, \alpha, \xi\right)$, by integrating (2.3) in order to remove θ^n and η, we are unable to integrate out ξ. Defining

$$S_x = \sum_{i=1}^{n} x_i \qquad S_y = \sum_{i=1}^{n} y_i \qquad S_t = S_x + S_y + nk = \sum_{i=1}^{n}\left(x_i+y_i+k\right)$$

we obtain the posterior distribution for $\left(\theta_{n+1}, \alpha, \xi\right)$ as

$$p\left(\theta_{n+1}, \alpha, \xi \mid D^n, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\theta_{n+1}}\left(1+e^{\xi}\right) - ve^{\xi}\right\}}{\left(e^{\alpha}+s\right)^{g+r}}\times$$

$$\times\frac{\exp\left\{\theta_{n+1}\left(x_{n+1}+k\right) + \alpha\left(g+S_y\right) + ((n+1)k+u)\xi\right\}}{\left(1+e^{\alpha}+e^{\xi}\right)^{S_t}}. \qquad (2.4)$$

To derive the predictive distribution, we have to adapt (2.2), and the predictive distribution of Y_{n+1} will be given by

$$p\left(y_{n+1} \mid D^n, x_{n+1}\right) = \int_{\Re^3} p\left(y_{n+1} \mid \theta_{n+1}, \alpha\right) p\left(\theta_{n+1}, \alpha, \xi \mid D^n, x_{n+1}\right) d\theta_{n+1}\, d\alpha\, d\xi$$

$$\propto \frac{1}{y_{n+1}!} \int_{\Re^3} \frac{\exp\left\{-e^{\alpha+\theta_{n+1}} - e^{\theta_{n+1}}\left(1+e^{\xi}\right) - v e^{\xi}\right\}}{\left(e^{\alpha}+s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{\left(\alpha+\theta_{n+1}\right) y_{n+1} + \theta_{n+1}\left(x_{n+1}+k\right) + \alpha\left(g+S_y\right)\right\}}{\left(1+e^{\alpha}+e^{\xi}\right)^{S_t}} \times$$

$$\times \exp\left\{\left((n+1)k+u\right)\xi\right\} d\theta_{n+1}\, d\alpha\, d\xi$$

$$\propto \frac{\Gamma\left(x_{n+1}+y_{n+1}+k\right)}{y_{n+1}!} \int_{\Re^2} \frac{\exp\left\{-v e^{\xi}\right\} \exp\left\{\alpha\left(g+S_y+y_{n+1}\right)\right\}}{\left(e^{\alpha}+s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{\left((n+1)k+u\right)\xi\right\} d\alpha\, d\xi}{\left(1+e^{\alpha}+e^{\xi}\right)^{S_t+x_{n+1}+y_{n+1}+k}} . \qquad (2.5)$$

The integrations with respect to α and ξ cannot be carried out analytically. Therefore, in a practical situation, we will have to use a two-dimensional numerical integration technique to obtain the exact distribution (2.5).

For the special case where we have second stage vague priors $(u, v, r, s \to 0)$, the predictive distribution of Y_{n+1} simplifies to the explicit form

$$p\left(y_{n+1} \mid D^n, x_{n+1}\right) \propto \frac{B\left(S_t\, , x_{n+1}+y_{n+1}+k\right)}{B\left(S_y-1\, , y_{n+1}+1\right)} . \qquad (2.6)$$

In the general situation the predictive distribution of Y_{n+1} has to be computed numerically. In this case, the integral to solve has not a high dimensionality and therefore we are not expecting major numerical problems. However, to set the scene for later models we also derive estimates or approximations to the predictive distribution.

2.2.2. Estimation Via Gibbs Sampling

The Gibbs sampling algorithm, presented in section 1.2.1, can be used to estimate the predictive distribution of Y_{n+1}, given by (2.5) or (2.6). To implement this algorithm, we need to know the full conditional distributions from which we will draw random points. Those full conditional distributions are easily derived from the posterior distribution for $(\theta_{n+1}, \alpha, \xi)$. In the general case, the posterior distribution for $(\theta_{n+1}, \alpha, \xi)$ is given by (2.4). Thus, the full conditional distributions will be

$$p(\theta_{n+1} \mid \alpha, \xi, D^n, x_{n+1}) \propto \exp\left\{-e^{\theta_{n+1}}\left(1+e^{\xi}\right)\right\}\exp\left\{\theta_{n+1}(x_{n+1}+k)\right\},$$

$$p(\alpha \mid \theta_{n+1}, \xi, D^n, x_{n+1}) \propto \frac{\exp\left\{\alpha\left(g+S_y\right)\right\}}{\left(e^{\alpha}+s\right)^{g+r}\left(1+e^{\alpha}+e^{\xi}\right)^{S_t}}, \tag{2.7}$$

$$p(\xi \mid \theta_{n+1}, \alpha, D^n, x_{n+1}) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}}+v\right)\right\}\exp\left\{\left((n+1)k+u\right)\xi\right\}}{\left(1+e^{\alpha}+e^{\xi}\right)^{S_t}}.$$

In the particular case with second stage vague priors, these become, respectively,

$$p(\theta_{n+1} \mid \alpha, \xi, D^n, x_{n+1}) \propto \exp\left\{-e^{\theta_{n+1}}\left(1+e^{\xi}\right)\right\}\exp\left\{\theta_{n+1}(x_{n+1}+k)\right\},$$

$$p(\alpha \mid \theta_{n+1}, \xi, D^n, x_{n+1}) \propto \frac{\exp\left\{\alpha S_y\right\}}{\left(1+e^{\alpha}+e^{\xi}\right)^{S_t}}, \tag{2.8}$$

$$p(\xi \mid \theta_{n+1}, \alpha, D^n, x_{n+1}) \propto \frac{\exp\left\{-e^{\xi+\theta_{n+1}}\right\}\exp\left\{(n+1)k\xi\right\}}{\left(1+e^{\alpha}+e^{\xi}\right)^{S_t}}.$$

In both cases, the random sampling of values for θ_{n+1} is straightforward noting that the full conditional distribution is a transformed gamma distribution (see Table A1.1). Hence, those values will simply be the logarithm of values which are drawn from a gamma distribution. Typically the generation of values of α from (2.7) will require the use of rejection sampling (section 1.2.3). For the special case of second stage vague priors we note that the full conditional distribution in (2.8) is a transformed beta distribution (see Table A1.2). Therefore, in this special case, the generation of α can be done generating β from a beta distribution and using the transformation

$$\alpha = \ln\left\{\left(1+e^{\xi}\right)\left(\frac{1-\beta}{\beta}\right)\right\}.$$

The generation of ξ either from (2.7) or (2.8) requires the use of rejection sampling.

As pointed out in section 1.2.3, one of the difficulties for implementing the rejection sampling algorithm is to be able to find a good envelope function in order to have a satisfactory rate of acceptable values. If we find difficulty, in defining an efficient envelope function for the distributions in (2.7) and (2.8), we can overcome the problem by drawing random values from the full conditional distributions derived from the joint posterior distribution for $\left(\theta^n, \theta_{n+1}, \alpha, \xi, \eta\right)$, given by (2.3). Those full conditional distributions are

$$p\left(\theta_i \mid \theta_{j\neq i}, \theta_{n+1}, \alpha, \xi, \eta, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\theta_i}\left(1+e^{\alpha}+e^{\xi}\right)\right\}\exp\left\{\theta_i\left(x_i+y_i+k\right)\right\},$$

$$i=1, 2, \ldots, n,$$

$$p\left(\theta_{n+1} \mid \theta_1, \ldots, \theta_n, \alpha, \xi, \eta, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\theta_{n+1}}\left(1+e^{\xi}\right)\right\}\exp\left\{\theta_{n+1}\left(x_{n+1}+k\right)\right\},$$

$$p\left(\alpha \mid \theta_1, \ldots, \theta_n, \theta_{n+1}, \xi, \eta, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\alpha}\left(\sum_{i=1}^{n} e^{\theta_i}+e^{\eta}\right)\right\}\exp\left\{\alpha\left(g+S_y\right)\right\},$$

$$p\left(\xi \mid \theta_1, \ldots, \theta_n, \theta_{n+1}, \alpha, \eta, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\xi}\left(v+\sum_{i=1}^{n+1} e^{\theta_i}\right)\right\}\exp\left\{\left((n+1)k+u\right)\xi\right\},$$

$$p\left(\eta \mid \theta_1, \ldots, \theta_n, \theta_{n+1}, \alpha, \xi, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\eta}\left(e^{\alpha}+s\right)\right\}\exp\left\{\eta\left(g+r\right)\right\}.$$

The random sampling of values from any of these full conditional distributions is very simple, noting that all of them are transformed gamma distributions (see Table A1.1).

Using this set of full conditional distributions, we avoid the use of the rejection sampling algorithm and all possible related problems. However, we will then have a cost to pay: the number of values to sample in each cycle of the Gibbs sampling algorithm increases from three to $(n+3)$, which can make the algorithm much more time consuming. Hence, this option should be used just when suitable and efficient envelope functions for (2.7) and (2.8) cannot be found.

Taking one of those sets of full conditional distributions, we perform the Gibbs sampling routine with t iterations to obtain each one of the M samples. Although we generate values for more parameters, we just have to register the values for θ_{n+1} and α, because they are the only ones we need in order to estimate the predictive distribution. So, let us consider that

$$\left(\theta_{n+1(j)}^{(t)}, \alpha_{(j)}^{(t)}\right), \quad j = 1, 2, \ldots, M$$

are the M samples generated by Gibbs sampling. We can use (1.2) to estimate the predictive distribution of Y_{n+1}. Such estimate will be

$$\hat{p}\left(y_{n+1} \mid D^n, x_{n+1}\right) = \frac{1}{M} \sum_{j=1}^{M} \frac{\exp\left\{-\mu_j\right\} \mu_j^{y_{n+1}}}{y_{n+1}!},$$

where $\mu_j = \exp\left\{\alpha_{(j)}^{(t)} + \theta_{n+1(j)}^{(t)}\right\}, \quad j = 1, 2, \ldots, M.$

2.2.3. Estimation Via Asymptotic Results

Suppose n is large enough that the posterior distribution can be successfully approximated by a multivariate normal distribution, as presented in section 1.2.2. As the full conditional distributions derived from a multivariate normal distribution are univariate normals, the predictive distribution of Y_{n+1} is very easy to estimate through the Gibbs sampling algorithm. The next two sections will present two possible normal approximations for the posterior distribution, following the methods summarised in sections 1.2.2.1 and 1.2.2.2.

2.2.3.1. Posterior Normality Based on the Likelihood Function

The asymptotic result presented in section 1.2.2.1 takes into consideration just the likelihood function. Therefore, no matter which prior structure we consider, by (1.3), we can write that the posterior distribution for a parameter vector $\left(\theta^n, \theta_{n+1}, \alpha\right)$ is approximated by

$$p\left(\theta^{n}, \theta_{n+1}, \alpha \mid D^{n}, x_{n+1}\right) \cong N_{n+2}\left(\hat{\mu}, H^{-1}(\hat{\mu})\right),$$

where $\hat{\mu}^{T} = \left(\hat{\theta}^{n}, \hat{\theta}_{n+1}, \hat{\alpha}\right)$ is a vector which components are the maximum likelihood estimates of the parameters and $H(\hat{\mu})$ is the Hessian matrix evaluated at $\hat{\mu}$.

The maximum likelihood estimates are shown in (2.1). We find that

$$H(\hat{\mu}) = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}^{T} & V_{22} \end{pmatrix},$$

where

$$V_{11} = diag\left(a_i, i = 1, 2, \ldots, n\right), \quad V_{12}^{T} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ b_1 & b_2 & \cdots & b_n \end{pmatrix}, \quad V_{22} = diag(c, d),$$

and

$$a_i = e^{\hat{\theta}_i}\left(1 + e^{\hat{\alpha}}\right), \qquad b_i = e^{\hat{\theta}_i + \hat{\alpha}}, \; i = 1, 2, \ldots, n,$$

$$c = e^{\hat{\theta}_{n+1}}, \qquad d = e^{\hat{\alpha}} \sum_{i=1}^{n} e^{\hat{\theta}_i}.$$

Note that to predict Y_{n+1} we just need the marginal posterior distribution for (θ_{n+1}, α), which is

$$p\left(\theta_{n+1}, \alpha \mid D^{n}, x_{n+1}\right) \cong N_2\left(\left(\hat{\theta}_{n+1}, \hat{\alpha}\right)^{T}, W\right),$$

where

$$W = \left(V_{22} - V_{12}^{T} V_{11}^{-1} V_{12}\right)^{-1} = diag(1/c, 1/f),$$

with

$$f = d - \sum_{i=1}^{n} \frac{b_i^2}{a_i}.$$

Thus θ_{n+1} and α are a posteriori independent with distributions

$$p\left(\theta_{n+1} \mid D^n, x_{n+1}\right) = N\left(\hat{\theta}_{n+1}, 1/c\right) = N\left(\ln\left(x_{n+1}\right), \frac{1}{x_{n+1}}\right),$$

$$p\left(\alpha \mid D^n, x_{n+1}\right) = N\left(\hat{\alpha}, 1/f\right) = N\left(\ln\left(\frac{S_y}{S_x}\right), \frac{S_x + S_y}{S_x S_y}\right).$$

Since this approximation requires the maximum likelihood estimates of the parameters, the implementation of this method is impossible if $x_{n+1} = 0$.

2.2.3.2. Posterior Normality Based on Characteristics of the Posterior Distribution

Following the method presented in section 1.2.2.2, the posterior distribution (2.4) for $\left(\theta_{n+1}, \alpha, \xi\right)$ can be approximated by a multivariate normal distribution,

$$p\left(\theta_{n+1}, \alpha, \xi \mid D^n, x_{n+1}\right) \approx N_3(m, V),$$

where m is the posterior mode, defined by (1.4), and V is the modal dispersion matrix, defined by (1.5).

The posterior mode $m = \left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\xi}\right)$ is obtained from the numerical solution of the equations

$$e^{\theta_{n+1}}\left(1 + e^{\xi}\right) - x_{n+1} - k = 0,$$

$$g + S_y - \frac{(g + r)e^{\alpha}}{e^{\alpha} + s} - \frac{S_t e^{\alpha}}{1 + e^{\alpha} + e^{\xi}} = 0, \tag{2.9}$$

$$e^{\xi}\left(e^{\theta_{n+1}} + v\right) + \frac{S_t e^{\xi}}{1 + e^{\alpha} + e^{\xi}} - (n + 1)k - u = 0,$$

whilst the inverse of the modal dispersion matrix is given by

$$V^{-1} = \begin{pmatrix} \tilde{a} & 0 & \tilde{b} \\ 0 & \tilde{c} & \tilde{d} \\ \tilde{b} & \tilde{d} & \tilde{f} \end{pmatrix}, \tag{2.10}$$

where

$$\tilde{a} = e^{\tilde{\theta}_{n+1}}\left(1+e^{\tilde{\xi}}\right), \qquad \tilde{b} = e^{\tilde{\theta}_{n+1}+\tilde{\xi}}, \qquad \tilde{c} = \frac{(g+r)se^{\tilde{\alpha}}}{\left(e^{\tilde{\alpha}}+s\right)^2} + \frac{e^{\tilde{\alpha}}\left(1+e^{\tilde{\xi}}\right)S_t}{\left(1+e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)^2},$$

$$\tilde{d} = -\frac{e^{\tilde{\alpha}+\tilde{\xi}}S_t}{\left(1+e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)^2}, \qquad \tilde{f} = e^{\tilde{\xi}}\left(e^{\tilde{\theta}_{n+1}}+v\right) + \frac{e^{\tilde{\xi}}\left(1+e^{\tilde{\alpha}}\right)S_t}{\left(1+e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)^2}.$$

The appropriate conditional distributions to be used in the Gibbs sampling routine are given by

$$p\left(\theta_{n+1} \mid \alpha, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_{n+1} - \frac{\tilde{b}}{\tilde{a}}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{a}}\right),$$

$$p\left(\alpha \mid \theta_{n+1}, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\alpha} - \frac{\tilde{d}}{\tilde{c}}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{c}}\right),$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, D^n, x_{n+1}\right) = N\left(\tilde{\xi} - \frac{\left(\theta_{n+1} - \tilde{\theta}_{n+1}\right)\tilde{b} + \left(\alpha - \tilde{\alpha}\right)\tilde{d}}{\tilde{f}}, \frac{1}{\tilde{f}}\right).$$

For the special case when we are taking second stage vague priors $(u, v, r, s \rightarrow 0)$ explicit modal values can be obtained from (2.9) as

$$\tilde{\theta}_{n+1} = \ln\left\{\frac{\left(S_x + x_{n+1}\right)\left(x_{n+1} + k\right)}{S_x + x_{n+1} + (n+1)k}\right\}, \qquad \tilde{\alpha} = \ln\left\{\frac{\left(S_x + x_{n+1} + (n+1)k\right)S_y}{\left(S_x + x_{n+1}\right)\left(S_x + nk\right)}\right\},$$

$$\tilde{\xi} = \ln\left\{\frac{(n+1)k}{S_x + x_{n+1}}\right\}, \tag{2.11}$$

and the constants $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ and $\tilde{f}$ become

$$\tilde{a} = x_{n+1} + k, \qquad \tilde{b} = \frac{(x_{n+1} + k)(n+1)k}{S_x + x_{n+1} + (n+1)k},$$

$$\tilde{c} = \frac{(S_x + nk)S_y}{S_t}, \qquad \tilde{d} = -\frac{(S_x + nk)(n+1)k\,S_y}{(S_x + x_{n+1} + (n+1)k)\,S_t}, \tag{2.12}$$

$$\tilde{f} = \frac{(n+1)k\left[(S_x + x_{n+1})(S_x + nk) + (S_x + x_{n+1} + (n+1)k)\,S_y\right](S_x + nk)}{(S_x + x_{n+1} + (n+1)k)^2\,S_t} +$$

$$+ \frac{(x_{n+1} + k)(n+1)k}{S_x + x_{n+1} + (n+1)k}.$$

2.2.4. Laplace Approximation

Given the posterior distribution (2.4) for $(\theta_{n+1}, \alpha, \xi)$, the predictive distribution of Y_{n+1} is defined by

$$p(y_{n+1} \mid D^n, x_{n+1}) \propto \frac{1}{y_{n+1}!} \int_{\Re^3} \exp\left\{-e^{\alpha + \theta_{n+1}}\right\} \exp\left\{(\alpha + \theta_{n+1}) y_{n+1}\right\} \times$$

$$\times\, p(\theta_{n+1}, \alpha, \xi \mid D^n, x_{n+1})\, d\theta_{n+1}\, d\alpha\, d\xi.$$

Following the Laplace method in section 1.2.4, we consider this as

$$E\left[g(\theta_{n+1}, \alpha, \xi) \mid D^n, x_{n+1}\right]$$

where

$$g(\theta_{n+1}, \alpha, \xi) = \exp\left\{-e^{\alpha + \theta_{n+1}}\right\} \exp\left\{(\alpha + \theta_{n+1}) y_{n+1}\right\},$$

and, as in (1.13) and (1.14), define functions $h(\theta_{n+1}, \alpha, \xi)$ and $h^*(\theta_{n+1}, \alpha, \xi)$ by

$$-n h(\theta_{n+1}, \alpha, \xi) = \ln p(\theta_{n+1}, \alpha, \xi) + \ln p(D^n, x_{n+1} \mid \theta_{n+1}, \alpha, \xi)$$

$$= -e^{\theta_{n+1}}(1 + e^{\xi}) - v e^{\xi} + \theta_{n+1}(x_{n+1} + k) + \alpha(g + S_y) +$$

$$+((n+1)k+u)\xi-(g+r)\ln\left(e^{\alpha}+s\right)-S_t\ln\left(1+e^{\alpha}+e^{\xi}\right)$$

and

$$-nh^*(\theta_{n+1},\alpha,\xi)=\ln g(\theta_{n+1},\alpha,\xi)+\ln p(\theta_{n+1},\alpha,\xi)+\ln p\left(D^n,x_{n+1}\mid\theta_{n+1},\alpha,\xi\right)$$

$$=-e^{\alpha+\theta_{n+1}}+(\alpha+\theta_{n+1})y_{n+1}-e^{\theta_{n+1}}\left(1+e^{\xi}\right)-ve^{\xi}+\theta_{n+1}(x_{n+1}+k)+$$

$$+\alpha(g+S_y)+((n+1)k+u)\xi-(g+r)\ln\left(e^{\alpha}+s\right)-$$

$$-S_t\ln\left(1+e^{\alpha}+e^{\xi}\right).$$

Defining $\left(\tilde{\theta}_{n+1},\tilde{\alpha},\tilde{\xi}\right)$ and $\tilde{\sigma}$ as in (1.16) we see that $\left(\tilde{\theta}_{n+1},\tilde{\alpha},\tilde{\xi}\right)=m$, solution of (2.9) and

$$\tilde{\sigma}=\left\{\tilde{a}\tilde{c}\tilde{f}-\tilde{b}^2\tilde{c}-\tilde{d}^2\tilde{a}\right\}^{-1/2},$$

from (2.10).

In a similar way, we define $\left(\theta^*_{n+1},\alpha^*,\xi^*\right)$ and σ^* as in (1.16) and find that the optimal $\left(\theta^*_{n+1},\alpha^*,\xi^*\right)$ is obtained using a numerical method to solve the equations

$$e^{\theta_{n+1}}\left(1+e^{\alpha}+e^{\xi}\right)-x_{n+1}-y_{n+1}-k=0,$$

$$e^{\alpha+\theta_{n+1}}+\frac{(g+r)e^{\alpha}}{e^{\alpha}+s}+\frac{e^{\alpha}S_t}{1+e^{\alpha}+e^{\xi}}-g-S_y-y_{n+1}=0, \tag{2.13}$$

$$e^{\xi}\left(e^{\theta_{n+1}}+v\right)+\frac{e^{\xi}S_t}{1+e^{\alpha}+e^{\xi}}-(n+1)k-u=0.$$

To derive the optimal σ^* we define

$$a^*=e^{\theta^*_{n+1}}\left(1+e^{\alpha^*}+e^{\xi^*}\right),\quad m^*=e^{\theta^*_{n+1}+\alpha^*},\quad b^*=e^{\theta^*_{n+1}+\xi^*},\quad d^*=-\frac{e^{\alpha^*+\xi^*}S_t}{\left(1+e^{\alpha^*}+e^{\xi^*}\right)^2},$$

$$c^* = e^{\alpha^* + \theta^*_{n+1}} + \frac{(g+r)se^{\alpha^*}}{\left(e^{\alpha^*}+s\right)^2} + \frac{e^{\alpha^*}\left(1+e^{\xi^*}\right)S_t}{\left(1+e^{\alpha^*}+e^{\xi^*}\right)^2}, \qquad f^* = e^{\xi^*}\left(e^{\theta^*_{n+1}}+v\right) + \frac{e^{\xi^*}\left(1+e^{\alpha^*}\right)S_t}{\left(1+e^{\alpha^*}+e^{\xi^*}\right)^2},$$

so that

$$n\nabla^2 h^*\left(\theta^*_{n+1}, \alpha^*, \xi^*\right) = \begin{pmatrix} a^* & m^* & b^* \\ m^* & c^* & d^* \\ b^* & d^* & f^* \end{pmatrix},$$

and thus

$$\sigma^* = \left\{a^* c^* f^* + 2m^* b^* d^* - b^{*2} c^* - d^{*2} a^* - m^{*2} f^*\right\}^{-1/2}.$$

Finally, by (1.17), the Laplace approximation for the predictive distribution of Y_{n+1} is given by

$$p\left(y_{n+1} \mid D^n, x_{n+1}\right) \propto \frac{1}{y_{n+1}!}\left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\left\{-nh^*\left(\theta^*_{n+1}, \alpha^*, \xi^*\right) + nh\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\xi}\right)\right\}.$$

Simplifications occur when we consider second stage vague priors because the systems of equations involved can then be solved explicitly. The values of $\tilde{\theta}_{n+1}$, $\tilde{\alpha}$ and $\tilde{\xi}$ follow as in (2.11) and the constants $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, $\tilde{d}$ and $\tilde{f}$ are defined as in (2.12). Similarly we obtain

$$\theta^*_{n+1} = \ln\left\{\frac{\left(S_x + x_{n+1}\right)\left(x_{n+1} + y_{n+1} + k\right)}{S_t + x_{n+1} + y_{n+1} + k}\right\}, \quad \alpha^* = \ln\left\{\frac{S_y + y_{n+1}}{S_x + x_{n+1}}\right\}, \quad \xi^* = \ln\left\{\frac{(n+1)k}{S_x + x_{n+1}}\right\},$$

as being the solution of (2.13) and the constants a^*, m^*, b^*, c^*, d^* and f^* become

$$a^* = x_{n+1} + y_{n+1} + k, \qquad m^* = \frac{\left(S_y + y_{n+1}\right)\left(x_{n+1} + y_{n+1} + k\right)}{S_t + x_{n+1} + y_{n+1} + k},$$

$$b^* = \frac{\left(x_{n+1} + y_{n+1} + k\right)(n+1)k}{S_t + x_{n+1} + y_{n+1} + k}, \qquad d^* = -\frac{\left(S_y + y_{n+1}\right)(n+1)kS_t}{\left(S_t + x_{n+1} + y_{n+1} + k\right)^2},$$

$$c^{*}=\frac{\left(S_y+y_{n+1}\right)\left(x_{n+1}+y_{n+1}+k\right)}{S_t+x_{n+1}+y_{n+1}+k}+\frac{\left(S_y+y_{n+1}\right)\left(S_x+x_{n+1}+(n+1)k\right)S_t}{\left(S_t+x_{n+1}+y_{n+1}+k\right)^2},$$

$$f^{*}=\frac{(n+1)k\left(x_{n+1}+y_{n+1}+k\right)}{S_t+x_{n+1}+y_{n+1}+k}+\frac{(n+1)k\left(S_x+S_y+x_{n+1}+y_{n+1}\right)S_t}{\left(S_t+x_{n+1}+y_{n+1}+k\right)^2}.$$

2.2.5. Examples and Conclusions

In order to illustrate the performance of the methods used in this chapter and also to evaluate their efficiency, we will present two numerical applications, considering two data sets used by Dunsmore & Robson (1994a).

Example 1:

The data shown in Table 2.1 are simulated values of traffic counts x_i and corresponding number of accidents y_i $(i = 1, 2, \ldots, 20)$ at 20 road junctions.

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
12	1	40	4	16	5	17	2
24	2	19	2	23	4	10	2
28	3	16	3	29	4	22	5
27	2	16	1	27	6	22	2
18	1	15	3	18	2	34	6

Table 2.1. Traffic counts x_i and number of accidents y_i $(i = 1, 2, \ldots, 20)$ for 20 road junction (simulated values)

Predictions are given for y_{21} corresponding to $x_{21} = 14$. In the analyses we assume a vague second stage prior structure $(u, v, r, s \rightarrow 0)$. With this choice of the second stage parameters we are able to derive explicitly the exact predictive distribution, and, therefore, it is possible to compare it with the predictive probabilities obtained using estimative or approximative techniques.

We notice throughout the development that for such a case specification of g is not necessary; we just need to specify k. We do this by matching the first two marginal moments of the X_is (Dunsmore & Robson (1997), Gelfand & Smith (1990) and Gaver & O'Muircheartaigh (1987)). The theoretical moments are

$$E(X_i) = E(E(X_i \mid \theta_i)) = E(e^{\theta_i}) = \frac{k}{e^{\xi}}$$

$$V(X_i) = V(E(X_i \mid \theta_i)) + E(V(X_i \mid \theta_i)) = V(e^{\theta_i}) + E(e^{\theta_i}) = \frac{k(1+e^{\xi})}{e^{2\xi}}$$

and the empirical moments are $\bar{x}$ and s_x^2, the sample mean and the sample variance, respectively. By equating the moments we obtain

$$k = \frac{\bar{x}^2}{s_x^2 - \bar{x}}. \tag{2.14}$$

From (2.6), we see that the predictive distribution of Y_{21}, considering vague second stage priors, reduces to the ratio between two beta functions. Because it is much easier to evaluate them when their arguments are integer, we rounded the value of k above, taking $k=13$.

Figure 2.1 shows the exact predictive distribution of Y_{21} and the different estimates and approximations we derived in this chapter (see Table A3.1, Appendix 3). Normal approximation 1 refers to the Bernardo & Smith (1994) approach and normal approximation 2 refers to the O'Hagan's (1994) approach. The Gibbs sampling estimate was done generating values from (2.8) with $M=500$ and $t=100$.

Clearly the Laplace and the Gibbs methods provide excellent approximations to the exact predictive distribution. Equally it is clear that the anticipated problems with the multivariate normal approximations with a sample of only *20* manifested themselves, although O'Hagan's (1994) suggestion seems much superior to the more usual posterior normal approximation suggested by Bernardo & Smith (1994). This latter normal approximation seems to tend to diverge from the exact distribution in the direction of the plug-in estimate. We can point out two possible reasons for that. Firstly, while the normal approximation suggested by Bernardo & Smith (1994) is based here on *22* parameters, the one suggested by O'Hagan (1994) is based here on only *3* parameters. Secondly, O'Hagan (1994) takes into consideration the prior structure chosen, in opposition to the approximation suggested by Bernardo & Smith (1994) which just takes into consideration the likelihood function and the maximum likelihood estimates.

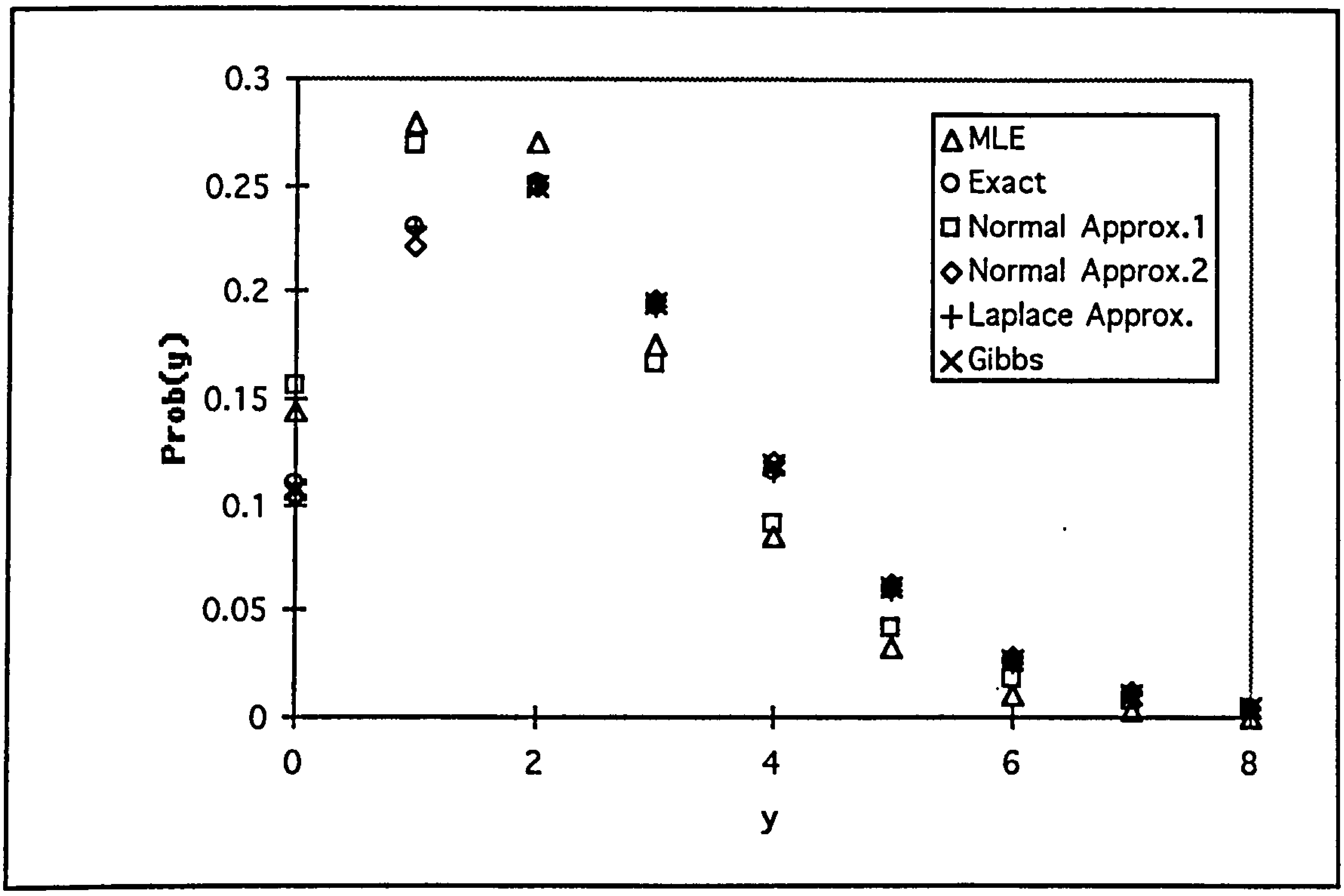

Figure 2.1: The predictive distribution of Y_{21} (the exact one and its approximations and estimates), using the data in Table 2.1.

Example 2:

In Table 2.2 we have a subset of the data reported by Svensson (1981) on an experiment carried out in Sweden during the summers of 1961 and 1962 into the effect of speed limits on road traffic accidents. The multiplicative Poisson model discussed here was shown by Svensson (1981) to provide a good fit to the data. On *43* days in 1961 when no speed limits were imposed, the number of traffic accidents were recorded. On the corresponding days in 1962 a general speed limit was imposed and the numbers of accidents were again recorded. The pairs of values (x_i, y_i) are shown in Table 2.2.

x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i	x_i	y_i
29	17	40	23	28	16	17	20	15	13	21	13	24	9	15	10
32	17	22	12	24	7	11	11	27	15	37	29	32	17	25	17
20	15	40	25	21	9	18	16	35	25	21	25	25	16	34	22
42	21	27	17	34	26	47	41	36	25	15	12	26	17	21	24
39	26	39	16	21	15	15	12	17	22	20	24	24	16	30	25
25	14	8	15	21	9										

Table 2.2: Number of traffic accidents: x_i with no speed restrictions in 1961; y_i with speed restrictions in 1962 $(i = 1, 2, ..., 43)$.

Corresponding to $x_{44} = 20$, we want to make predictions about Y_{44}. We assume once again a vague second stage prior structure. As before, we use (2.14) to choose k; for this example, we took $k=12$.

The exact predictive distribution of Y_{44} and the different estimates and approximations are shown in Figure 2.2 (see Table A3.2, Appendix 3).

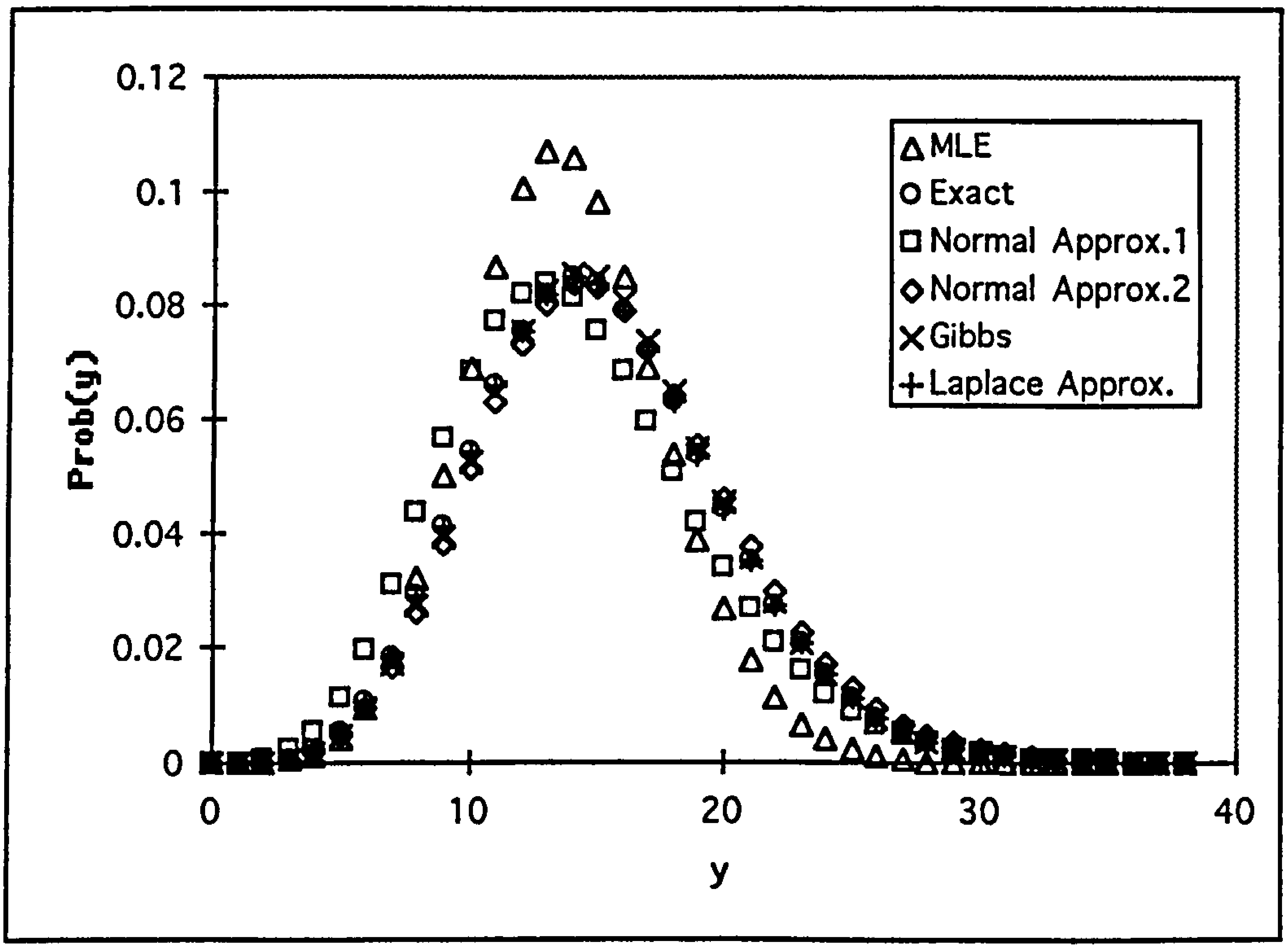

Figure 2.2: The predictive distribution of Y_{44} (the exact one and its approximations and estimates), using the data in Table 2.2.

The conclusions are similar to those in the first example, namely that the Gibbs and the Laplace methods provide very good approximations of the predictive probabilities. However, in this case, the normal approximations, in particular the one suggested by O'Hagan (1994), also provide good results. This may be due to the fact that the sample size is larger. The computational speed of the Laplace method, in comparison to Gibbs sampling, is a strong point in its favour.

Therefore, we conclude that the Laplace approximation and the Gibbs sampling can provide alternative and reliable approaches when the evaluation of the exact predictive probabilities are to be avoided due to the high dimensionality of the integrals involved. The use of the usual posterior normal approximations can be suspect because of the high dimensionality of the parameters, although O'Hagan's (1994) approach improves matters somewhat.

CHAPTER 3

TREATMENTS EFFECTS IN A BIASED ALLOCATION MODEL: PREDICTION IN A POISSON ERRORS IN VARIABLES MODEL

Robbins and Zhang (1991) considered a model in which (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) are independent random variables such that

- given ψ_i, $X_i \sim Po(\psi_i)$,
- given γ, ϕ, ψ_i and x_i,

  $Y_i \sim Po(\gamma\,\psi_i)$ if treatment T_1 is used,

  $Y_i \sim Po(\phi\,\psi_i)$ if treatment T_2 is used,
- given a, T_1 is used if $x_i < a$ and T_2 is used if $x_i \geq a$.

Here, γ and ϕ are unknown parameters which are used to model the multiplicative effects of treatments T_1 and T_2, respectively; ψ_i, $i = 1, 2, \ldots, n$, are parameters modelling particular characteristics of the individuals. We are thus considering again multiplicative treatments effects, but where the assignment of treatment depends on the covariate value X.

Robbins and Zhang (1991) considered this problem as one of estimation and took the difference $\gamma - \phi$ or the ratio γ / ϕ as a measure of the differential treatment effect. They treated $\psi_1, \psi_2, \ldots, \psi_n$ as nuisance parameters, and derived the following consistent estimators for γ and ϕ:

$$\hat{\gamma} = \frac{\sum_{i=1}^{n} y_i I(x_i < a)}{\sum_{i=1}^{n} x_i I(x_i < a+1)} \quad , \quad \hat{\phi} = \frac{\sum_{i=1}^{n} y_i I(x_i \geq a)}{\sum_{i=1}^{n} x_i I(x_i \geq a+1)} ,$$

where $I(.)$ represents the indicator function.

We consider the problem from the predictive aspect. Suppose that in the past we observed a group of n individuals, so that the available data are

$$D^n = \{(x_i, y_i), \, i = 1, 2, \ldots, n\}.$$

If we analyse a new individual, with measurement x_{n+1} before any treatment is applied, our aim is to make predictions about the outcomes after that individual receives one of the treatments.

To avoid problems later with the non-negativity constraints on the parameters in the model of Robbins and Zhang (1991), we will again use the alternative parameterisation of a logarithmic link function. Thus, we will assume that X_i and Y_i ($i=1, 2, \ldots, n$) are random variables such that

- given θ_i, $X_i \sim Po(\exp(\theta_i))$,
- given α, β, θ_i and x_i,

$$Y_i \sim Po(\exp(\alpha + \theta_i)), \text{if treatment } T_1 \text{ is used,} \tag{3.1}$$

$$Y_i \sim Po(\exp(\beta + \theta_i)), \text{if treatment } T_2 \text{ is used,}$$

- given a, T_1 is used if $x_i < a$ and T_2 is used if $x_i \geq a$.

In this new model, α and β are unknown parameters used to model the multiplicative effect of the treatments and $\theta_1, \theta_2, \ldots, \theta_n$ are nuisance parameters.

We will assume that for the $(n+1)$-th individual,

$$X_{n+1} \sim Po(\exp(\theta_{n+1})),$$

$$Y_{n+1,1} \sim Po(\exp(\alpha + \theta_{n+1})),$$

$$Y_{n+1,2} \sim Po(\exp(\beta + \theta_{n+1})),$$

where $Y_{n+1,1}$ and $Y_{n+1,2}$ are random variables representing the outcomes on the $(n+1)$-th individual if treatment T_1 or T_2 is used, respectively. The dependence between the outcomes for the individual from the two treatments is modelled through the common parameter θ_{n+1}. Our aim is to make predictive statements about $Y_{n+1,1}$ and $Y_{n+1,2}$ either individually or jointly.

In order to simplify the presentation of results, we define a treatment indicator

$$\delta_i = \begin{cases} 1, & \text{if } x_i < a \\ 0, & \text{if } x_i \geq a \end{cases},$$

for i=1, 2, ..., n and let

$$n_1 = \sum_{i=1}^{n} \delta_i \qquad n_2 = \sum_{i=1}^{n} (1-\delta_i) \qquad n = n_1 + n_2$$

$$S_{x1} = \sum_{i=1}^{n} \delta_i x_i \qquad S_{x2} = \sum_{i=1}^{n} (1-\delta_i) x_i \qquad T_x = \sum_{i=1}^{n} x_i = S_{x1} + S_{x2}$$

$$S_{y1} = \sum_{i=1}^{n} \delta_i y_i \qquad S_{y2} = \sum_{i=1}^{n} (1-\delta_i) y_i \qquad T_y = \sum_{i=1}^{n} y_i = S_{y1} + S_{y2}$$

$$W_1 = \sum_{i=1}^{n} \delta_i (x_i + y_i + k) = S_{x1} + S_{y1} + kn_1$$

$$W_2 = \sum_{i=1}^{n} (1-\delta_i)(x_i + y_i + k) = S_{x2} + S_{y2} + kn_2$$

3.1. A Classical Approach

Considering the model (3.1), the data set D^n and x_{n+1}, we can obtain a simple estimate for the predictive distributions of $Y_{n+1,1}$, $Y_{n+1,2}$ and $(Y_{n+1,1}, Y_{n+1,2})$, using the maximum likelihood estimates of the parameters involved.

With $\theta^n = (\theta_1, \theta_2, \ldots, \theta_n)$, the likelihood function is given by

$$L\left(\theta^n, \theta_{n+1}, \alpha, \beta; D^n, x_{n+1}\right) = \prod_{i=1}^{n}\left\{p(x_i \mid \theta_i)\left[p(y_i \mid \theta_i, \alpha)\right]^{\delta_i}\left[p(y_i \mid \theta_i, \beta)\right]^{1-\delta_i}\right\} \times$$

$$\times\, p\left(x_{n+1} \mid \theta_{n+1}\right)$$

$$\propto \exp\left\{-\sum_{i=1}^{n+1} e^{\theta_i} - \sum_{i=1}^{n} \delta_i e^{\alpha+\theta_i} - \sum_{i=1}^{n}(1-\delta_i)e^{\beta+\theta_i}\right\} \times$$

$$\times \exp\left\{\sum_{i=1}^{n+1} \theta_i x_i + \sum_{i=1}^{n}(\alpha+\theta_i)y_i\delta_i + \sum_{i=1}^{n}(\beta+\theta_i)y_i(1-\delta_i)\right\},$$

and its logarithm is

$$\ln L\left(\theta^n, \theta_{n+1}, \alpha, \beta; D^n, x_{n+1}\right) = \ln(const) - \sum_{i=1}^{n+1} e^{\theta_i} + \sum_{i=1}^{n+1}\theta_i x_i - \sum_{i=1}^{n}\delta_i e^{\alpha+\theta_i} +$$

$$+ \sum_{i=1}^{n}(\alpha+\theta_i)y_i\delta_i - \sum_{i=1}^{n}(1-\delta_i)e^{\beta+\theta_i} + \sum_{i=1}^{n}(\beta+\theta_i)y_i(1-\delta_i)\ ;$$

The maximum likelihood estimate $\left(\hat{\theta}^n, \hat{\theta}_{n+1}, \hat{\alpha}, \hat{\beta}\right)$ is given by

$$\hat{\theta}_{n+1} = \ln\left(x_{n+1}\right), \qquad \hat{\theta}_i = \ln\left(\frac{x_i + y_i}{1 + \delta_i e^{\hat{\alpha}} + (1-\delta_i)e^{\hat{\beta}}}\right), \quad i = 1, 2, \ldots, n,$$

$$\hat{\alpha} = \ln\left(\frac{S_{y1}}{S_{x1}}\right), \qquad \hat{\beta} = \ln\left(\frac{S_{y2}}{S_{x2}}\right). \tag{3.2}$$

Plug-in estimates for the predictive distributions are then given by

$$p\left(y_{n+1,1} \mid \hat{\theta}_{n+1}, \hat{\alpha}\right) = Po\left(\exp\left(\hat{\alpha} + \hat{\theta}_{n+1}\right)\right),$$

$$p\left(y_{n+1,2} \mid \hat{\theta}_{n+1}, \hat{\beta}\right) = Po\left(\exp\left(\hat{\beta} + \hat{\theta}_{n+1}\right)\right),$$

$$p\left(y_{n+1,1}, y_{n+1,2} \mid \hat{\theta}_{n+1}, \hat{\alpha}, \hat{\beta}\right) = p\left(y_{n+1,1} \mid \hat{\theta}_{n+1}, \hat{\alpha}\right)p\left(y_{n+1,2} \mid \hat{\theta}_{n+1}, \hat{\beta}\right).$$

A simple plug-in estimate for the probability function of the random variable $Z = Y_{n+1,2} - Y_{n+1,1}$, for example, would be given by

$$P(Z = z) = \sum_{i=\max(0,-z)}^{\infty} \frac{\exp\left\{-e^{\hat{\theta}_{n+1}}\left(e^{\hat{\alpha}} + e^{\hat{\beta}}\right)\right\}\exp\left\{\hat{\alpha}\, i + \hat{\beta}(i+z) + \hat{\theta}_{n+1}(2i+z)\right\}}{i!\,(i+z)!}.$$

Clearly problems arise with the plug-in method if $x_{n+1} = 0$, or if any of S_{x1}, S_{y1}, S_{x2}, S_{y2} is zero.

3.2. A Bayesian Approach

In the Bayesian approach, the predictive distributions of $Y_{n+1,1}$, $Y_{n+1,2}$ and $\left(Y_{n+1,1}, Y_{n+1,2}\right)$ are given, respectively, by

$$p\left(y_{n+1,1} \mid D^n, x_{n+1}\right) = \int_{\Re^2} p\left(y_{n+1,1} \mid \theta_{n+1}, \alpha\right) p\left(\theta_{n+1}, \alpha \mid D^n, x_{n+1}\right) d\theta_{n+1}\, d\alpha, \qquad (3.3)$$

$$p\left(y_{n+1,2} \mid D^n, x_{n+1}\right) = \int_{\Re^2} p\left(y_{n+1,2} \mid \theta_{n+1}, \beta\right) p\left(\theta_{n+1}, \beta \mid D^n, x_{n+1}\right) d\theta_{n+1}\, d\beta \qquad (3.4)$$

and

$$p\left(y_{n+1,1}, y_{n+1,2} \mid D^n, x_{n+1}\right) = \int_{\Re^3} p\left(y_{n+1,1} \mid \theta_{n+1}, \alpha\right) p\left(y_{n+1,2} \mid \theta_{n+1}, \beta\right) \times$$

$$\times\, p\left(\theta_{n+1}, \alpha, \beta \mid D^n, x_{n+1}\right) d\theta_{n+1}\, d\alpha\, d\beta. \qquad (3.5)$$

Notice that $\theta_1, \theta_2, \ldots, \theta_n$ are nuisance parameters and we only require the posterior distribution of $\left(\theta_{n+1}, \alpha, \beta\right)$.

3.2.1. The Exact Predictive Distributions

We consider a hierarchical prior structure; at the first stage we take

$$p\left(\theta^{n}, \theta_{n+1}, \alpha, \beta \mid \xi, \eta, \zeta\right) = \prod_{i=1}^{n+1}\left\{p\left(\theta_i \mid \xi\right)\right\} p\left(\alpha \mid \eta\right) p\left(\beta \mid \zeta\right),$$

and at the second stage we take

$$p(\xi, \eta, \zeta) = p(\xi) p(\eta) p(\zeta),$$

where

$$e^{\theta_i} \sim Ga\left(k, e^{\xi}\right) \qquad e^{\alpha} \sim Ga\left(h, e^{\eta}\right) \qquad e^{\beta} \sim Ga\left(g, e^{\zeta}\right)$$

$$e^{\xi} \sim Ga(l, m) \qquad e^{\eta} \sim Ga(u, v) \qquad e^{\zeta} \sim Ga(r, s)$$

with k, h, g, l, m, u, v, r and s specified.

Again, using the equivalent forms in Table A1.1, we obtain the posterior distribution

$$p\left(\theta^{n}, \theta_{n+1}, \alpha, \beta, \xi, \eta, \zeta \mid D^{n}, x_{n+1}\right) \propto \exp\left\{-\sum_{i=1}^{n+1} e^{\theta_i} - \sum_{i=1}^{n} \delta_i e^{\alpha+\theta_i} - \sum_{i=1}^{n}\left(1-\delta_i\right) e^{\beta+\theta_i}\right\} \times$$

$$\times \exp\left\{-\sum_{i=1}^{n+1} e^{\xi+\theta_i} - e^{\eta+\alpha} - e^{\zeta+\beta} - m e^{\xi} - v e^{\eta} - s e^{\zeta}\right\} \times$$

$$\times \exp\left\{\sum_{i=1}^{n+1} \theta_i x_i + \sum_{i=1}^{n}\left(\alpha+\theta_i\right) y_i \delta_i + \sum_{i=1}^{n}\left(\beta+\theta_i\right) y_i\left(1-\delta_i\right)\right\} \times$$

$$\times \exp\left\{k \sum_{i=1}^{n+1}\left(\xi+\theta_i\right) + h(\eta+\alpha) + g(\zeta+\beta) + l\xi + u\eta + r\zeta\right\}.$$

The elimination of $\theta_1, \theta_2, \ldots, \theta_n, \eta$ and ζ yields the posterior distribution of $(\theta_{n+1}, \alpha, \beta, \xi)$ given by

$$p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^{n}, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\theta_{n+1}}\left(1+e^{\xi}\right) - m e^{\xi}\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{\theta_{n+1}\left(x_{n+1}+k\right) + \left((n+1)k + l\right)\xi + \alpha\left(h+S_{y1}\right) + \beta\left(g+S_{y2}\right)\right\}}{\left(1+e^{\alpha}+e^{\xi}\right)^{W_1}\left(1+e^{\beta}+e^{\xi}\right)^{W_2}}, \qquad (3.6)$$

but we are unable to eliminate ξ. It follows that the predictive distributions are given by

$$p\left(y_{n+1,1} \mid D^n, x_{n+1}\right) \propto \frac{\Gamma\left(x_{n+1} + y_{n+1,1} + k\right)}{y_{n+1,1}!} \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\}\exp\left\{((n+1)k+l)\xi\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{\alpha\left(h+S_{y1}+y_{n+1,1}\right)+\beta\left(g+S_{y2}\right)\right\}d\alpha\, d\beta\, d\xi}{\left(1+e^{\alpha}+e^{\xi}\right)^{W_1+x_{n+1}+y_{n+1,1}+k}\left(1+e^{\beta}+e^{\xi}\right)^{W_2}}, \tag{3.7}$$

$$p\left(y_{n+1,2} \mid D^n, x_{n+1}\right) \propto \frac{\Gamma\left(x_{n+1} + y_{n+1,2} + k\right)}{y_{n+1,2}!} \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\}\exp\left\{((n+1)k+l)\xi\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{\alpha\left(h+S_{y1}\right)+\beta\left(g+S_{y2}+y_{n+1,2}\right)\right\}d\alpha\, d\beta\, d\xi}{\left(1+e^{\alpha}+e^{\xi}\right)^{W_1}\left(1+e^{\beta}+e^{\xi}\right)^{W_2+x_{n+1}+y_{n+1,2}+k}} \tag{3.8}$$

and

$$p\left(y_{n+1,1}, y_{n+1,2} \mid D^n, x_{n+1}\right) \propto \frac{\Gamma\left(x_{n+1} + y_{n+1,1} + y_{n+1,2} + k\right)}{y_{n+1,1}!\, y_{n+1,2}!} \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{((n+1)k+l)\xi+\alpha\left(h+S_{y1}+y_{n+1,1}\right)+\beta\left(g+S_{y2}+y_{n+1,2}\right)\right\}d\alpha\, d\beta\, d\xi}{\left(1+e^{\alpha}+e^{\xi}\right)^{W_1}\left(1+e^{\beta}+e^{\xi}\right)^{W_2}\left(1+e^{\alpha}+e^{\beta}+e^{\xi}\right)^{x_{n+1}+y_{n+1,1}+y_{n+1,2}+k}}. \tag{3.9}$$

The evaluation of these predictive distributions therefore requires the use of a numerical integration technique to solve the three-dimensional integrals involved.

Although no simple analytical form is available for (3.7), (3.8) and (3.9), it is possible to obtain the marginal (but dependent) predictive probabilities in the case of vague second stage priors $(l, m, u, v, r, s \to 0)$ explicitly, namely:

$$p\left(y_{n+1,1} \mid D^n, x_{n+1}\right) \propto \frac{B\left(x_{n+1}+y_{n+1,1}+k,\, W_1\right)}{B\left(y_{n+1,1}+1,\, S_{y1}-1\right)} \tag{3.10}$$

and

$$p\left(y_{n+1,2} \mid D^n, x_{n+1}\right) \propto \frac{B\left(x_{n+1} + y_{n+1,2} + k,\, W_2\right)}{B\left(y_{n+1,2} + 1,\, S_{y2} - 1\right)} . \tag{3.11}$$

No explicit form for the exact joint predictive distribution (3.9) results, however.

3.2.2. Estimation Via Gibbs Sampling

We can use the Gibbs sampling algorithm (section 1.2.1) to estimate the predictive distributions without the need of evaluating the problematic integrals in (3.7), (3.8) and (3.9). From (3.6) it is easy to derive the full conditional distributions which are

$$p\left(\theta_{n+1} \mid \alpha, \beta, \xi, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\theta_{n+1}}\left(1+e^{\xi}\right)\right\}\exp\left\{\theta_{n+1}\left(x_{n+1}+k\right)\right\},$$

$$p\left(\alpha \mid \theta_{n+1}, \beta, \xi, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\alpha\left(h+S_{y1}\right)\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(1+e^{\alpha}+e^{\xi}\right)^{W_1}} , \tag{3.12}$$

$$p\left(\beta \mid \theta_{n+1}, \alpha, \xi, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\beta\left(g+S_{y2}\right)\right\}}{\left(e^{\beta}+s\right)^{g+r}\left(1+e^{\beta}+e^{\xi}\right)^{W_2}} ,$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, \beta, D^n, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}}+m\right)\right\}\exp\left\{\left((n+1)k+l\right)\xi\right\}}{\left(1+e^{\alpha}+e^{\xi}\right)^{W_1}\left(1+e^{\beta}+e^{\xi}\right)^{W_2}} .$$

Using the fact (see Table A1.1) that the full conditional distribution of θ_{n+1} is a transformed gamma distribution and using rejection sampling to generate values of α, β and ξ from (3.12), we obtain a sample after t iterations of $\left(\theta_{n+1}^{(t)}, \alpha^{(t)}, \beta^{(t)}, \xi^{(t)}\right)$. Repeating the whole procedure M times we obtain samples

$$\left(\theta_{n+1(j)}^{(t)}, \alpha_{(j)}^{(t)}, \beta_{(j)}^{(t)}, \xi_{(j)}^{(t)}\right) , \; j = 1, 2, \ldots, M.$$

The predictive distributions (3.7), (3.8) and (3.9) can now be estimated by

$$p\left(y_{n+1,1} \mid D^n, x_{n+1}\right) = \frac{1}{M}\sum_{j=1}^{M}\frac{\exp\left\{-\mu_j\right\}\mu_j^{y_{n+1,1}}}{y_{n+1,1}!} ,$$

$$p\left(y_{n+1,2} \mid D^n, x_{n+1}\right) = \frac{1}{M} \sum_{j=1}^{M} \frac{\exp\left\{-v_j\right\} v_j^{y_{n+1,2}}}{y_{n+1,2}!}, \tag{3.13}$$

$$p\left(y_{n+1,1}, y_{n+1,2} \mid D^n, x_{n+1}\right) = \frac{1}{M} \sum_{j=1}^{M} \left\{ \frac{\exp\left\{-\mu_j\right\} \mu_j^{y_{n+1,1}}}{y_{n+1,1}!} \frac{\exp\left\{-v_j\right\} v_j^{y_{n+1,2}}}{y_{n+1,2}!} \right\},$$

where $\mu_j = \exp\left\{\alpha_{(j)}^{(t)} + \theta_{n+1(j)}^{(t)}\right\}$ and $v_j = \exp\left\{\beta_{(j)}^{(t)} + \theta_{n+1(j)}^{(t)}\right\}$, for $j = 1, 2, \ldots, M$. Notice that, although it is necessary to generate values of ξ in this Gibbs routine, these values of this hyperparameter are not required further for our prediction problem.

For the case of vague second stage priors, α and β can be generated from transformed beta distributions (see Table A1.2).

3.2.3. Estimation Via Asymptotic Results

As the sample size increases, the number of parameters in our model increases as well. Problems may therefore arise over assumptions of asymptotic normality of the overall posterior distribution. Since $\theta_1, \theta_2, \ldots, \theta_n$ are nuisance parameters, however, and our concern in evaluating (3.3), (3.4) and (3.5) is only in $\left(\theta_{n+1}, \alpha, \beta\right)$, we have proceeded with the asymptotic approximation for $p\left(\theta_{n+1}, \alpha, \beta \mid D^n, x_{n+1}\right)$, when possible, or for $p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n, x_{n+1}\right)$.

3.2.3.1. Posterior Normality Based on the Likelihood Function

We consider first the asymptotic posterior normality of $\left(\theta^n, \theta_{n+1}, \alpha, \beta\right)$ following Bernardo & Smith (1994) and summarised in section 1.2.2.1. With the maximum likelihood estimates $\left(\hat{\theta}^n, \hat{\theta}_{n+1}, \hat{\alpha}, \hat{\beta}\right)$ as given in (3.2), we find that

$$p\left(\theta^n, \theta_{n+1}, \alpha, \beta \mid D^n, x_{n+1}\right) \approx N\left(\left(\hat{\theta}^n, \hat{\theta}_{n+1}, \hat{\alpha}, \hat{\beta}\right)^T, \begin{pmatrix} V_{11} & V_{12} \\ V_{12}^T & V_{22} \end{pmatrix}^{-1}\right)$$

where

$$V_{11} = diag\left(\hat{a}_i\,,\, i = 1, 2, \ldots, n\right), \quad V_{12}^T = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \hat{b}_1 & \hat{b}_2 & \cdots & \hat{b}_n \\ \hat{c}_1 & \hat{c}_2 & \cdots & \hat{c}_n \end{pmatrix}, \quad V_{22} = diag\left(\hat{d}, \hat{f}, \hat{g}\right)$$

and

$$\hat{b}_i = \delta_i\, e^{\hat{\alpha}+\hat{\theta}_i}, \qquad \hat{c}_i = \left(1-\delta_i\right)e^{\hat{\beta}+\hat{\theta}_i}, \qquad \hat{a}_i = e^{\hat{\theta}_i} + \hat{b}_i + \hat{c}_i, \qquad i = 1, 2, \ldots, n$$

$$\hat{d} = e^{\hat{\theta}_{n+1}}, \qquad \hat{f} = e^{\hat{\alpha}} \sum_{i=1}^{n} \delta_i\, e^{\hat{\theta}_i}, \qquad \hat{g} = e^{\hat{\beta}} \sum_{i=1}^{n} \left(1-\delta_i\right) e^{\hat{\theta}_i}.$$

The marginal posterior distribution for $\left(\theta_{n+1}, \alpha, \beta\right)$ will then be

$$p\left(\theta_{n+1}, \alpha, \beta \mid D^n, x_{n+1}\right) \cong N_3\left(\left(\hat{\theta}_{n+1}, \hat{\alpha}, \hat{\beta}\right)^T, W\right)$$

where

$$W = \left(V_{22} - V_{12}^T V_{11}^{-1} V_{12}\right)^{-1} = diag\left(\frac{1}{\hat{d}}, \frac{1}{\hat{h}}, \frac{1}{\hat{m}}\right),$$

with

$$\hat{h} = \hat{f} - \sum_{i=1}^{n} \frac{\hat{b}_i^2}{\hat{a}_i} \qquad \text{and} \qquad \hat{m} = \hat{g} - \sum_{i=1}^{n} \frac{\hat{c}_i^2}{\hat{a}_i}.$$

Thus θ_{n+1}, α and β are independent a posteriori, with

$$p\left(\theta_{n+1} \mid D^n, x_{n+1}\right) = N\left(\hat{\theta}_{n+1}, \frac{1}{\hat{d}}\right) = N\left(\ln\left(x_{n+1}\right), \frac{1}{x_{n+1}}\right),$$

$$p\left(\alpha \mid D^n, x_{n+1}\right) = N\left(\hat{\alpha}, \frac{1}{\hat{h}}\right) = N\left(\ln\left(\frac{S_{y1}}{S_{x1}}\right), \frac{S_{x1}+S_{y1}}{S_{x1}S_{y1}}\right), \tag{3.14}$$

$$p\left(\beta \mid D^n, x_{n+1}\right) = N\left(\hat{\beta}, \frac{1}{\hat{m}}\right) = N\left(\ln\left(\frac{S_{y2}}{S_{x2}}\right), \frac{S_{x2}+S_{y2}}{S_{x2}S_{y2}}\right).$$

The Gibbs sampling algorithm is then trivial, with $t=1$, and the estimation of the predictive distributions we are interested in is straightforward.

Again problems exist if $x_{n+1} = 0$, if $x_i = y_i = 0$ for one $i = 1, 2, \ldots, n$, or if any of $S_{x1}, S_{y1}, S_{x2}, S_{y2}$ is zero.

3.2.3.2. Posterior Normality Based on Characteristics of the Posterior Distribution

Alternatively we use the asymptotic approximation suggested by O'Hagan (1994) and summarised in section 1.2.2.2, and have that

$$p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n, x_{n+1}\right) \cong N_4(m, V)$$

where the posterior mode $m=\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$ is obtained solving iteratively the system of four equations

$$\begin{aligned}
&e^{\theta_{n+1}}\left(1+e^{\xi}\right)-x_{n+1}-k=0, \\
&h+S_{y1}-\frac{(h+u)e^{\alpha}}{e^{\alpha}+v}-\frac{W_1 e^{\alpha}}{1+e^{\alpha}+e^{\xi}}=0, \\
&g+S_{y2}-\frac{(g+r)e^{\beta}}{e^{\beta}+s}-\frac{W_2 e^{\beta}}{1+e^{\beta}+e^{\xi}}=0, \\
&e^{\xi}\left(e^{\theta_{n+1}}+m\right)-(n+1)k-l+\frac{W_1 e^{\xi}}{1+e^{\alpha}+e^{\xi}}+\frac{W_2 e^{\xi}}{1+e^{\beta}+e^{\xi}}=0,
\end{aligned} \tag{3.15}$$

whilst the precision matrix V^{-1} takes the form

$$V^{-1}=\begin{pmatrix} \tilde{a} & 0 & 0 & \tilde{b} \\ 0 & \tilde{c} & 0 & \tilde{d} \\ 0 & 0 & \tilde{f} & \tilde{q} \\ \tilde{b} & \tilde{d} & \tilde{q} & \tilde{t} \end{pmatrix}$$

where

$$\tilde{a} = e^{\tilde{\theta}_{n+1}}\left(1+e^{\tilde{\xi}}\right), \qquad \tilde{b} = e^{\tilde{\theta}_{n+1}+\tilde{\xi}}, \qquad \tilde{c} = \frac{(h+u)ve^{\tilde{\alpha}}}{\left(e^{\tilde{\alpha}}+v\right)^2} + \frac{W_1 e^{\tilde{\alpha}}\left(1+e^{\tilde{\xi}}\right)}{\left(1+e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)^2},$$

$$\tilde{d} = -\frac{W_1 e^{\tilde{\alpha}+\tilde{\xi}}}{\left(1+e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)^2}, \qquad \tilde{f} = \frac{(g+r)se^{\tilde{\beta}}}{\left(e^{\tilde{\beta}}+s\right)^2} + \frac{W_2 e^{\tilde{\beta}}\left(1+e^{\tilde{\xi}}\right)}{\left(1+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2}, \qquad \tilde{q} = -\frac{W_2 e^{\tilde{\beta}+\tilde{\xi}}}{\left(1+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2},$$

$$\tilde{t} = e^{\tilde{\xi}}\left(e^{\tilde{\theta}_{n+1}}+m\right) + \frac{W_1 e^{\tilde{\xi}}\left(1+e^{\tilde{\alpha}}\right)}{\left(1+e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)^2} + \frac{W_2 e^{\tilde{\xi}}\left(1+e^{\tilde{\beta}}\right)}{\left(1+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2}.$$

The relevant full conditional distributions are then given by

$$p\left(\theta_{n+1} \mid \alpha, \beta, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_{n+1} - \frac{\tilde{b}}{\tilde{a}}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{a}}\right),$$

$$p\left(\alpha \mid \theta_{n+1}, \beta, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\alpha} - \frac{\tilde{d}}{\tilde{c}}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{c}}\right), \tag{3.16}$$

$$p\left(\beta \mid \theta_{n+1}, \alpha, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\beta} - \frac{\tilde{q}}{\tilde{f}}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{f}}\right),$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, \beta, D^n, x_{n+1}\right) = N\left(\tilde{\xi} - \frac{\left(\theta_{n+1} - \tilde{\theta}_{n+1}\right)\tilde{b} + \left(\alpha - \tilde{\alpha}\right)\tilde{d} + \left(\beta - \tilde{\beta}\right)\tilde{q}}{\tilde{t}}, \frac{1}{\tilde{t}}\right).$$

If we consider the special case with vague second stage priors ($l, m, u, v, r, s \to 0$), this approximation can be simplified to some extent, since we can obtain the posterior mode $\boldsymbol{m}$ explicitly as

$$\tilde{\theta}_{n+1} = \ln\left\{\frac{\left(x_{n+1}+k\right)\left(T_x + x_{n+1}\right)}{T_x + x_{n+1} + (n+1)k}\right\}, \qquad \tilde{\alpha} = \ln\left\{\frac{\left(T_x + x_{n+1} + (n+1)k\right)S_{y1}}{\left(T_x + x_{n+1}\right)\left(W_1 - S_{y1}\right)}\right\}, \tag{3.17}$$

$$\tilde{\beta} = \ln\left\{\frac{\left(T_x + x_{n+1} + (n+1)k\right)S_{y2}}{\left(T_x + x_{n+1}\right)\left(W_2 - S_{y2}\right)}\right\}, \qquad \tilde{\xi} = \ln\left\{\frac{(n+1)k}{T_x + x_{n+1}}\right\}.$$

The full conditional distributions defined by (3.16) then become

$$p\left(\theta_{n+1} \mid \alpha, \beta, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_{n+1} - \tilde{w}\left(\xi - \tilde{\xi}\right), \frac{1}{x_{n+1} + k}\right),$$

$$p\left(\alpha \mid \theta_{n+1}, \beta, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\alpha} + \tilde{w}\left(\xi - \tilde{\xi}\right), \frac{W_1}{S_{y1}\left(W_1 - S_{y1}\right)}\right), \tag{3.18}$$

$$p\left(\beta \mid \theta_{n+1}, \alpha, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\beta} + \tilde{w}\left(\xi - \tilde{\xi}\right), \frac{W_2}{S_{y2}\left(W_2 - S_{y2}\right)}\right),$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, \beta, D^n, x_{n+1}\right) = N\left(\tilde{\xi} - \frac{\left(\theta_{n+1} - \tilde{\theta}_{n+1}\right)\tilde{b} + \left(\alpha - \tilde{\alpha}\right)\tilde{d} + \left(\beta - \tilde{\beta}\right)\tilde{q}}{\tilde{t}}, \frac{1}{\tilde{t}}\right),$$

where

$$\tilde{w} = \frac{(n+1)k}{T_x + x_{n+1} + (n+1)k}, \qquad \tilde{b} = \left(x_{n+1} + k\right)\tilde{w},$$

$$\tilde{d} = -\frac{S_{y1}\left(W_1 - S_{y1}\right)}{W_1}\tilde{w}, \qquad \tilde{q} = -\frac{S_{y2}\left(W_2 - S_{y2}\right)}{W_2}\tilde{w},$$

$$\tilde{t} = \tilde{b} + \tilde{w}\left(\frac{\left(T_x + x_{n+1}\right)W_1 + (n+1)kS_{y1}}{T_x + x_{n+1} + (n+1)k}\frac{W_1 - S_{y1}}{W_1} + \frac{\left(T_x + x_{n+1}\right)W_2 + (n+1)kS_{y2}}{T_x + x_{n+1} + (n+1)k}\frac{W_2 - S_{y2}}{W_2}\right)$$

Notice that, in contrast with what happened with the approximation developed in section 3.2.3.1, the full conditional distributions are not independent of each other. Therefore, we now have to perform an iterative process to generate each value in the Gibbs routine.

Finally, the estimation of $p\left(y_{n+1,1}, y_{n+1,2} \mid D^n, x_{n+1}\right)$ and both marginal predictive distributions can be undertaken through (3.13) using (3.16) or (3.18) as the appropriate conditional distributions.

3.2.4. Laplace Approximation

In section 1.2.4 we summarised how the Laplace approximation can be used to obtain the predictive probabilities, given the posterior distribution. The basic idea of this method is to see the integrals involved in the predictive distributions as being posterior expectations of real functions.

3.2.4.1. Marginal Predictive Distribution of $Y_{n+1,1}$

Let us begin by approximating the predictive probabilities of $Y_{n+1,1}$ given by (3.3). The integral here can be regarded as $E\left[g_1(\theta_{n+1}, \alpha, \beta, \xi) \mid D^n, x_{n+1}\right]$ where

$$g_1(\theta_{n+1}, \alpha, \beta, \xi) = \exp\left\{-e^{\alpha+\theta_{n+1}}\right\}\exp\left\{(\alpha+\theta_{n+1})y_{n+1,1}\right\}.$$

As in (1.13) and (1.14), we define functions $h_1(\theta_{n+1}, \alpha, \beta, \xi)$ and $h_1^*(\theta_{n+1}, \alpha, \beta, \xi)$ such that

$$\begin{aligned}-n h_1(\theta_{n+1}, \alpha, \beta, \xi) = {} & -e^{\theta_{n+1}}\left(1+e^{\xi}\right) - me^{\xi} + \theta_{n+1}(x_{n+1}+k) + ((n+1)k+l)\xi + \\ & + \alpha\left(h+S_{y1}\right) + \beta\left(g+S_{y2}\right) - (h+u)\ln\left(e^{\alpha}+v\right) - \\ & -(g+r)\ln\left(e^{\beta}+s\right) - W_1\ln\left(1+e^{\alpha}+e^{\xi}\right) - W_2\ln\left(1+e^{\beta}+e^{\xi}\right)\end{aligned} \tag{3.19}$$

and

$$\begin{aligned}-n h_1^*(\theta_{n+1}, \alpha, \beta, \xi) = {} & -e^{\alpha+\theta_{n+1}} + (\alpha+\theta_{n+1})y_{n+1,1} - e^{\theta_{n+1}}\left(1+e^{\xi}\right) - me^{\xi} + \\ & + \theta_{n+1}(x_{n+1}+k) + ((n+1)k+l)\xi + \alpha\left(h+S_{y1}\right) + \beta\left(g+S_{y2}\right) - \\ & -(h+u)\ln\left(e^{\alpha}+v\right) - (g+r)\ln\left(e^{\beta}+s\right) - W_1\ln\left(1+e^{\alpha}+e^{\xi}\right) - \\ & -W_2\ln\left(1+e^{\beta}+e^{\xi}\right)\end{aligned}$$

and we also define $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$ and $\tilde{\sigma}$ satisfying the first and third expressions in (1.16).

Noting that (3.19) is just the logarithm of the posterior distribution (3.6) and analysing the definition for $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$, we conclude that $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right) = m$, the posterior mode which is the solution of the system of equations (3.15); noting also that the Hessian matrix in the definition of $\tilde{\sigma}$ is the inverse of the modal dispersion matrix defined in section 3.2.3.2, we conclude that

$$\tilde{\sigma} = \left\{\tilde{a}\tilde{c}\tilde{f}\tilde{t} - \tilde{a}\tilde{c}\tilde{q}^2 - \tilde{a}\tilde{f}\tilde{d}^2 - \tilde{c}\tilde{f}\tilde{b}^2\right\}^{-1/2}. \tag{3.20}$$

Then, we define $\left(\theta^*_{n+1(1)}, \alpha^*_1, \beta^*_1, \xi^*_1\right)$ and σ^*_1 such that the second and fourth expressions in (1.16) hold. The optimal $\left(\theta^*_{n+1(1)}, \alpha^*_1, \beta^*_1, \xi^*_1\right)$ is obtained solving, through a numerical technique, a system formed by the equations

$$\begin{aligned}
&e^{\theta_{n+1}}\left(1 + e^{\alpha} + e^{\xi}\right) - x_{n+1} - y_{n+1,1} - k = 0,\\
&e^{\alpha+\theta_{n+1}} + \frac{(h+u)e^{\alpha}}{e^{\alpha}+v} + \frac{W_1 e^{\alpha}}{1+e^{\alpha}+e^{\xi}} - h - S_{y1} - y_{n+1,1} = 0,\\
&\frac{(g+r)e^{\beta}}{e^{\beta}+s} + \frac{W_2 e^{\beta}}{1+e^{\beta}+e^{\xi}} - g - S_{y2} = 0,\\
&e^{\xi}\left(e^{\theta_{n+1}} + m\right) + \frac{W_1 e^{\xi}}{1+e^{\alpha}+e^{\xi}} + \frac{W_2 e^{\xi}}{1+e^{\beta}+e^{\xi}} - (n+1)k - l = 0.
\end{aligned} \tag{3.21}$$

To derive the optimal σ^*_1 we consider the second order derivatives of $h^*_1\left(\theta_{n+1}, \alpha, \beta, \xi\right)$ evaluated at $\left(\theta^*_{n+1(1)}, \alpha^*_1, \beta^*_1, \xi^*_1\right)$ and define

$$m^*_1 = e^{\alpha^*_1 + \theta^*_{n+1(1)}}, \qquad b^*_1 = e^{\theta^*_{n+1(1)} + \xi^*_1}, \qquad a^*_1 = e^{\theta^*_{n+1(1)}} + m^*_1 + b^*_1,$$

$$c^*_1 = e^{\alpha^*_1 + \theta^*_{n+1(1)}} + \frac{(h+u)v e^{\alpha^*_1}}{\left(e^{\alpha^*_1}+v\right)^2} + \frac{W_1 e^{\alpha^*_1}\left(1+e^{\xi^*_1}\right)}{\left(1+e^{\alpha^*_1}+e^{\xi^*_1}\right)^2}, \qquad d^*_1 = -\frac{W_1 e^{\alpha^*_1+\xi^*_1}}{\left(1+e^{\alpha^*_1}+e^{\xi^*_1}\right)^2},$$

$$f_1^* = \frac{(g+r)se^{\beta_1^*}}{\left(e^{\beta_1^*}+s\right)^2} + \frac{W_2 e^{\beta_1^*}\left(1+e^{\xi_1^*}\right)}{\left(1+e^{\beta_1^*}+e^{\xi_1^*}\right)^2}, \qquad q_1^* = -\frac{W_2 e^{\beta_1^*+\xi_1^*}}{\left(1+e^{\beta_1^*}+e^{\xi_1^*}\right)^2},$$

$$t_1^* = e^{\xi_1^*}\left(e^{\theta_{n+1(1)}^*}+m\right) + \frac{W_1 e^{\xi_1^*}\left(1+e^{\alpha_1^*}\right)}{\left(1+e^{\alpha_1^*}+e^{\xi_1^*}\right)^2} + \frac{W_2 e^{\xi_1^*}\left(1+e^{\beta_1^*}\right)}{\left(1+e^{\beta_1^*}+e^{\xi_1^*}\right)^2}.$$

The Hessian matrix can then be written as

$$n\nabla^2 h_1^*\left(\theta_{n+1(1)}^*, \alpha_1^*, \beta_1^*, \xi_1^*\right) = \begin{pmatrix} a_1^* & m_1^* & 0 & b_1^* \\ m_1^* & c_1^* & 0 & d_1^* \\ 0 & 0 & f_1^* & q_1^* \\ b_1^* & d_1^* & q_1^* & t_1^* \end{pmatrix}$$

and thus,

$$\sigma_1^* = \left\{a_1^* c_1^* f_1^* t_1^* - a_1^* c_1^* q_1^{*2} - a_1^* f_1^* d_1^{*2} - f_1^* t_1^* m_1^{*2} + m_1^{*2} q_1^{*2} + \right.$$

$$\left. + 2m_1^* d_1^* b_1^* f_1^* - c_1^* f_1^* b_1^{*2}\right\}^{-1/2}.$$

Finally, the approximation for the predictive distribution of $Y_{n+1,1}$ is given by

$$p\left(y_{n+1,1} \mid D^n, x_{n+1}\right) \approx \frac{1}{y_{n+1,1}!}\left(\frac{\sigma_1^*}{\bar{\sigma}}\right)\exp\left\{-n h_1^*\left(\theta_{n+1(1)}^*, \alpha_1^*, \beta_1^*, \xi_1^*\right) + \right.$$

$$\left. + n h_1\left(\bar{\theta}_{n+1}, \bar{\alpha}, \bar{\beta}, \bar{\xi}\right)\right\}.$$

3.2.4.2. Marginal Predictive Distribution of $Y_{n+1,2}$

The Laplace approximation for the predictive distribution (3.4) of $Y_{n+1,2}$ can be derived in a similar way. We define the function

$$g_2\left(\theta_{n+1}, \alpha, \beta, \xi\right) = \exp\left\{-e^{\beta+\theta_{n+1}}\right\}\exp\left\{\left(\beta+\theta_{n+1}\right)y_{n+1,2}\right\}$$

and also two functions $h_2\left(\theta_{n+1}, \alpha, \beta, \xi\right)$ and $h_2^*\left(\theta_{n+1}, \alpha, \beta, \xi\right)$ such that

$$-nh_2(\theta_{n+1}, \alpha, \beta, \xi) = \ln p(\theta_{n+1}, \alpha, \beta, \xi) + \ln p(D^n, x_{n+1} \mid \theta_{n+1}, \alpha, \beta, \xi)$$

$$= -nh_1(\theta_{n+1}, \alpha, \beta, \xi)$$

and

$$-nh_2^*(\theta_{n+1}, \alpha, \beta, \xi) = -e^{\beta+\theta_{n+1}} + (\beta+\theta_{n+1})y_{n+1,2} - e^{\theta_{n+1}}(1+e^{\xi}) - me^{\xi} +$$

$$+\theta_{n+1}(x_{n+1}+k) + ((n+1)k+l)\xi + \alpha(h+S_{y1}) + \beta(g+S_{y2}) -$$

$$-(h+u)\ln(e^{\alpha}+v) - (g+r)\ln(e^{\beta}+s) - W_1\ln(1+e^{\alpha}+e^{\xi}) -$$

$$-W_2\ln(1+e^{\beta}+e^{\xi}).$$

We also define $(\tilde{\theta}_{n+1(2)}, \tilde{\alpha}_2, \tilde{\beta}_2, \tilde{\xi}_2)$, $\tilde{\sigma}_2$, $(\theta^*_{n+1(2)}, \alpha^*_2, \beta^*_2, \xi^*_2)$ and σ^*_2 satisfying (1.16).

Clearly $(\tilde{\theta}_{n+1(2)}, \tilde{\alpha}_2, \tilde{\beta}_2, \tilde{\xi}_2) = (\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}) = m$, solution of (3.15) and $\tilde{\sigma}_2 = \tilde{\sigma}$, given by (3.20).

The system formed by the equations

$$\begin{aligned}
&e^{\theta_{n+1}}(1+e^{\beta}+e^{\xi}) - x_{n+1} - y_{n+1,2} - k = 0, \\
&\frac{(h+u)e^{\alpha}}{e^{\alpha}+v} + \frac{W_1e^{\alpha}}{1+e^{\alpha}+e^{\xi}} - h - S_{y1} = 0, \\
&e^{\beta+\theta_{n+1}} + \frac{(g+r)e^{\beta}}{e^{\beta}+s} + \frac{W_2e^{\beta}}{1+e^{\beta}+e^{\xi}} - g - S_{y2} - y_{n+1,2} = 0, \\
&e^{\xi}(e^{\theta_{n+1}}+m) + \frac{W_1e^{\xi}}{1+e^{\alpha}+e^{\xi}} + \frac{W_2e^{\xi}}{1+e^{\beta}+e^{\xi}} - (n+1)k - l = 0,
\end{aligned} \tag{3.22}$$

solved numerically, yields $(\theta^*_{n+1(2)}, \alpha^*_2, \beta^*_2, \xi^*_2)$.

The Hessian matrix in the definition of σ^*_2 will then take the form

$$n\nabla^2 h_2^*\left(\theta^*_{n+1(2)}, \alpha_2^*, \beta_2^*, \xi_2^*\right) = \begin{pmatrix} a_2^* & 0 & m_2^* & b_2^* \\ 0 & c_2^* & 0 & d_2^* \\ m_2^* & 0 & f_2^* & q_2^* \\ b_2^* & d_2^* & q_2^* & t_2^* \end{pmatrix}$$

where

$$m_2^* = e^{\beta_2^* + \theta^*_{n+1(2)}}, \qquad b_2^* = e^{\theta^*_{n+1(2)} + \xi_2^*}, \qquad a_2^* = e^{\theta^*_{n+1(2)}} + m_2^* + b_2^*,$$

$$c_2^* = \frac{(h+u)ve^{\alpha_2^*}}{\left(e^{\alpha_2^*} + v\right)^2} + \frac{W_1 e^{\alpha_2^*}\left(1 + e^{\xi_2^*}\right)}{\left(1 + e^{\alpha_2^*} + e^{\xi_2^*}\right)^2}, \qquad d_2^* = -\frac{W_1 e^{\alpha_2^* + \xi_2^*}}{\left(1 + e^{\alpha_2^*} + e^{\xi_2^*}\right)^2},$$

$$f_2^* = e^{\beta_2^* + \theta^*_{n+1(2)}} + \frac{(g+r)se^{\beta_2^*}}{\left(e^{\beta_2^*} + s\right)^2} + \frac{W_2 e^{\beta_2^*}\left(1 + e^{\xi_2^*}\right)}{\left(1 + e^{\beta_2^*} + e^{\xi_2^*}\right)^2}, \qquad q_2^* = -\frac{W_2 e^{\beta_2^* + \xi_2^*}}{\left(1 + e^{\beta_2^*} + e^{\xi_2^*}\right)^2},$$

$$t_2^* = e^{\xi_2^*}\left(e^{\theta^*_{n+1(2)}} + m\right) + \frac{W_1 e^{\xi_2^*}\left(1 + e^{\alpha_2^*}\right)}{\left(1 + e^{\alpha_2^*} + e^{\xi_2^*}\right)^2} + \frac{W_2 e^{\xi_2^*}\left(1 + e^{\beta_2^*}\right)}{\left(1 + e^{\beta_2^*} + e^{\xi_2^*}\right)^2}.$$

Therefore,

$$\sigma_2^* = \left\{a_2^* c_2^* f_2^* t_2^* - a_2^* c_2^* q_2^{*2} - a_2^* f_2^* d_2^{*2} - c_2^* t_2^* m_2^{*2} + d_2^{*2} m_2^{*2} + \right.$$

$$\left. + 2m_2^* c_2^* b_2^* q_2^* - c_2^* f_2^* b_2^{*2}\right\}^{-1/2},$$

and the Laplace approximation for $p\left(y_{n+1,2} \mid D^n, x_{n+1}\right)$ will be given by

$$p\left(y_{n+1,2} \mid D^n, x_{n+1}\right) \cong \frac{1}{y_{n+1,2}!}\left(\frac{\sigma_2^*}{\tilde{\sigma}}\right) \exp\left\{-n h_2^*\left(\theta^*_{n+1(2)}, \alpha_2^*, \beta_2^*, \xi_2^*\right) + \right.$$

$$\left. + n h_1\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\right\}.$$

3.2.4.3. Joint Predictive Distribution of $(Y_{n+1,1}, Y_{n+1,2})$

The derivation of the approximated joint predictive distribution (3.5) is analogous. We define the functions $g_c(\theta_{n+1}, \alpha, \beta, \xi)$, $h_c(\theta_{n+1}, \alpha, \beta, \xi)$ and $h_c^*(\theta_{n+1}, \alpha, \beta, \xi)$ such that

$$g_c(\theta_{n+1}, \alpha, \beta, \xi) = \exp\left\{-e^{\alpha+\theta_{n+1}} - e^{\beta+\theta_{n+1}}\right\}\exp\left\{(\alpha+\theta_{n+1})y_{n+1,1} + (\beta+\theta_{n+1})y_{n+1,2}\right\},$$

$$-n\,h_c(\theta_{n+1}, \alpha, \beta, \xi) = \ln p(\theta_{n+1}, \alpha, \beta, \xi) + \ln p(D^n, x_{n+1} \mid \theta_{n+1}, \alpha, \beta, \xi)$$

$$= -nh_1(\theta_{n+1}, \alpha, \beta, \xi)$$

and

$$-n\,h_c^*(\theta_{n+1}, \alpha, \beta, \xi) = -e^{\alpha+\theta_{n+1}} + (\alpha+\theta_{n+1})y_{n+1,1} - e^{\beta+\theta_{n+1}} + (\beta+\theta_{n+1})y_{n+1,2} -$$

$$-e^{\theta_{n+1}}(1+e^{\xi}) - me^{\xi} + \theta_{n+1}(x_{n+1}+k) + ((n+1)k+l)\xi +$$

$$+\alpha(h+S_{y1}) + \beta(g+S_{y2}) - (h+u)\ln(e^{\alpha}+v) -$$

$$-(g+r)\ln(e^{\beta}+s) - W_1\ln(1+e^{\alpha}+e^{\xi}) - W_2\ln(1+e^{\beta}+e^{\xi}),$$

and we also define $(\tilde{\theta}_{n+1(c)}, \tilde{\alpha}_c, \tilde{\beta}_c, \tilde{\xi}_c)$, $\tilde{\sigma}_c$, $(\theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c, \xi^*_c)$ and σ^*_c such that (1.16) is satisfied.

Once again, as $h_c(\theta_{n+1}, \alpha, \beta, \xi) = h_1(\theta_{n+1}, \alpha, \beta, \xi)$, we conclude that $(\theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c, \xi^*_c) = m$, obtained solving numerically the system (3.15), and $\sigma^*_c = \tilde{\sigma}$, defined by (3.20). Hence, to derive the Laplace approximation for the joint predictive distribution, we just have to work out the quantities based on the function $h_c^*(\theta_{n+1}, \alpha, \beta, \xi)$, namely $(\theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c, \xi^*_c)$ and σ^*_c. Solving numerically the system formed by the equations

$$e^{\theta_{n+1}}(1+e^{\alpha}+e^{\beta}+e^{\xi}) - x_{n+1} - y_{n+1,1} - y_{n+1,2} - k = 0,$$

$$e^{\alpha+\theta_{n+1}} - h - S_{y1} - y_{n+1,1} + \frac{(h+u)e^{\alpha}}{e^{\alpha}+v} + \frac{W_1 e^{\alpha}}{1+e^{\alpha}+e^{\xi}} = 0, \qquad (3.23)$$

$$e^{\beta+\theta_{n+1}} - g - S_{y2} - y_{n+1,2} + \frac{(g+r)e^{\beta}}{e^{\beta}+s} + \frac{W_2 e^{\beta}}{1+e^{\beta}+e^{\xi}} = 0,$$

$$e^{\xi}\left(e^{\theta_{n+1}} + m\right) + \frac{W_1 e^{\xi}}{1+e^{\alpha}+e^{\xi}} + \frac{W_2 e^{\xi}}{1+e^{\beta}+e^{\xi}} - (n+1)k - l = 0,$$

we obtain $\left(\theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c, \xi^*_c\right)$. Defining the constants

$$m^*_c = e^{\alpha^*_c+\theta^*_{n+1(c)}}, \quad p^*_c = e^{\beta^*_c+\theta^*_{n+1(c)}}, \quad b^*_c = e^{\theta^*_{n+1(c)}+\xi^*_c}, \quad a^*_c = e^{\theta^*_{n+1(c)}} + m^*_c + p^*_c + b^*_c,$$

$$c^*_c = e^{\alpha^*_c+\theta^*_{n+1(c)}} + \frac{(h+u)ve^{\alpha^*_c}}{\left(e^{\alpha^*_c}+v\right)^2} + \frac{W_1 e^{\alpha^*_c}\left(1+e^{\xi^*_c}\right)}{\left(1+e^{\alpha^*_c}+e^{\xi^*_c}\right)^2}, \qquad d^*_c = -\frac{W_1 e^{\alpha^*_c+\xi^*_c}}{\left(1+e^{\alpha^*_c}+e^{\xi^*_c}\right)^2},$$

$$f^*_c = e^{\beta^*_c+\theta^*_{n+1(c)}} + \frac{(g+r)se^{\beta^*_c}}{\left(e^{\beta^*_c}+s\right)^2} + \frac{W_2 e^{\beta^*_c}\left(1+e^{\xi^*_c}\right)}{\left(1+e^{\beta^*_c}+e^{\xi^*_c}\right)^2}, \qquad q^*_c = -\frac{W_2 e^{\beta^*_c+\xi^*_c}}{\left(1+e^{\beta^*_c}+e^{\xi^*_c}\right)^2},$$

$$t^*_c = e^{\xi^*_c}\left(e^{\theta^*_{n+1(c)}} + m\right) + \frac{W_1 e^{\xi^*_c}\left(1+e^{\alpha^*_c}\right)}{\left(1+e^{\alpha^*_c}+e^{\xi^*_c}\right)^2} + \frac{W_2 e^{\xi^*_c}\left(1+e^{\beta^*_c}\right)}{\left(1+e^{\beta^*_c}+e^{\xi^*_c}\right)^2},$$

the Hessian matrix in the definition of σ^*_c can be written as

$$n\nabla^2 h^*_c\left(\theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c, \xi^*_c\right) = \begin{pmatrix} a^*_c & m^*_c & p^*_c & b^*_c \\ m^*_c & c^*_c & 0 & d^*_c \\ p^*_c & 0 & f^*_c & q^*_c \\ b^*_c & d^*_c & q^*_c & t^*_c \end{pmatrix},$$

which leads to

$$\sigma^*_c = \left\{a^*_c c^*_c f^*_c t^*_c - a^*_c c^*_c q^{*2}_c - a^*_c f^*_c d^{*2}_c - f^*_c t^*_c m^{*2}_c + m^{*2}_c q^{*2}_c - 2m^*_c d^*_c p^*_c q^*_c + \right.$$
$$\left. + 2m^*_c d^*_c b^*_c f^*_c - c^*_c t^*_c p^{*2}_c + 2p^*_c c^*_c b^*_c q^*_c + d^{*2}_c p^{*2}_c - c^*_c f^*_c b^{*2}_c\right\}^{-1/2},$$

and the Laplace approximation for the joint predictive distribution will be given by

$$p\left(y_{n+1,1}, y_{n+1,2} \mid D^n, x_{n+1}\right) \propto \frac{1}{y_{n+1,1}!\, y_{n+1,2}!}\left(\frac{\sigma_c^*}{\tilde{\sigma}}\right) \exp\left\{-nh_c^*\left(\theta_{n+1(c)}^*, \alpha_c^*, \beta_c^*, \xi_c^*\right) + \right.$$
$$\left. + nh_1\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\right\}.$$

3.2.4.4. Vague Second Stage Priors Case

The implementation of the approximations can be simplified to a great extent when we consider vague second stage priors $(l, m, u, v, r, s \to 0)$, since some of the systems of equations involved in the solution of the problem yield analytical solutions.

The system (3.15), which originates $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$, has (3.17) as solution. $\left(\theta_{n+1(1)}^*, \alpha_1^*, \beta_1^*, \xi_1^*\right)$, which is the solution of the system (3.21), will in this case be

$$\theta_{n+1(1)}^* = \ln\left\{\frac{\left(x_{n+1} + y_{n+1,1} + k\right)\left(T_x + x_{n+1}\right)\left(W_1 - S_{y1} + x_{n+1} + k\right)}{\left(T_x + x_{n+1} + (n+1)k\right)\left(W_1 + x_{n+1} + y_{n+1,1} + k\right)}\right\}, \quad \xi_1^* = \ln\left\{\frac{(n+1)k}{T_x + x_{n+1}}\right\},$$

$$\alpha_1^* = \ln\left\{\frac{\left(S_{y1} + y_{n+1,1}\right)\left(T_x + x_{n+1} + (n+1)k\right)}{\left(T_x + x_{n+1}\right)\left(W_1 - S_{y1} + x_{n+1} + k\right)}\right\}, \qquad \beta_1^* = \ln\left\{\frac{S_{y2}\left(T_x + x_{n+1} + (n+1)k\right)}{\left(T_x + x_{n+1}\right)\left(W_2 - S_{y2}\right)}\right\}.$$

The system of equations (3.22), used to derive $\left(\theta_{n+1(2)}^*, \alpha_2^*, \beta_2^*, \xi_2^*\right)$, will have as solution

$$\theta_{n+1(2)}^* = \ln\left\{\frac{\left(x_{n+1} + y_{n+1,2} + k\right)\left(T_x + x_{n+1}\right)\left(W_2 - S_{y2} + x_{n+1} + k\right)}{\left(T_x + x_{n+1} + (n+1)k\right)\left(W_2 + x_{n+1} + y_{n+1,2} + k\right)}\right\}, \quad \xi_2^* = \ln\left\{\frac{(n+1)k}{T_x + x_{n+1}}\right\},$$

$$\alpha_2^* = \ln\left\{\frac{S_{y1}\left(T_x + x_{n+1} + (n+1)k\right)}{\left(T_x + x_{n+1}\right)\left(W_1 - S_{y1}\right)}\right\}, \qquad \beta_2^* = \ln\left\{\frac{\left(S_{y2} + y_{n+1,2}\right)\left(T_x + x_{n+1} + (n+1)k\right)}{\left(T_x + x_{n+1}\right)\left(W_2 - S_{y2} + x_{n+1} + k\right)}\right\}.$$

The system (3.23), which solution is $\left(\theta_{n+1(c)}^*, \alpha_c^*, \beta_c^*, \xi_c^*\right)$, has to be obtained through a numerical technique, even when we are considering vague second stage priors.

3.2.4.5. Predictive Distribution of $Z=Y_{n+1,2}-Y_{n+1,1}$

The predictive distribution of the random variable $Z = Y_{n+1,2} - Y_{n+1,1}$ can be written as

$$p\left(z \mid D^n, x_{n+1}\right) = \sum_{i=\max(0,-z)}^{\infty} \frac{1}{i!(i+z)!} E\left[g\left(\theta_{n+1}, \alpha, \beta, \xi\right) \mid D^n, x_{n+1}\right] \quad (3.24)$$

where

$$g\left(\theta_{n+1}, \alpha, \beta, \xi\right) = \exp\left\{-e^{\theta_{n+1}}\left(e^{\alpha} + e^{\beta}\right)\right\}\exp\left\{\alpha i + \beta(i+z) + \theta_{n+1}(2i+z)\right\}.$$

We will approximate the posterior expected value in (3.24) through the Laplace approximation. For that, we define a function $h_1\left(\theta_{n+1}, \alpha, \beta, \xi\right)$ as in (3.19), and we also define $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$ as being the solution of the system of equations (3.15) and $\tilde{\sigma}$ as in (3.20). For the vague second stage priors case, the solution of (3.15) can be written explicitly as in (3.17).

Then, we consider a function $h^*\left(\theta_{n+1}, \alpha, \beta, \xi\right)$ such that

$$\begin{aligned}
-n h^*\left(\theta_{n+1}, \alpha, \beta, \xi\right) = &-e^{\theta_{n+1}}\left(1 + e^{\alpha} + e^{\beta} + e^{\xi}\right) + \alpha\left(h + S_{y1} + i\right) + \beta\left(g + S_{y2} + i + z\right) + \\
&+ \theta_{n+1}\left(x_{n+1} + k + 2i + z\right) - m e^{\xi} + ((n+1)k + l)\xi - \\
&- (h+u)\ln\left(e^{\alpha} + v\right) - (g+r)\ln\left(e^{\beta} + s\right) - W_1 \ln\left(1 + e^{\alpha} + e^{\xi}\right) - \\
&- W_2 \ln\left(1 + e^{\beta} + e^{\xi}\right)
\end{aligned}$$

and we derive $\left(\theta^*_{n+1}, \alpha^*, \beta^*, \xi^*\right)$ solving numerically the equations

$$e^{\theta_{n+1}}\left(1 + e^{\alpha} + e^{\beta} + e^{\xi}\right) - x_{n+1} - k - z - 2i = 0,$$

$$e^{\theta_{n+1}+\alpha} + \frac{(h+u)e^{\alpha}}{e^{\alpha} + v} + \frac{W_1 e^{\alpha}}{1 + e^{\alpha} + e^{\xi}} - h - S_{y1} - i = 0, \quad (3.25)$$

$$e^{\theta_{n+1}+\beta} + \frac{(g+r)e^{\beta}}{e^{\beta}+s} + \frac{W_2 e^{\beta}}{1+e^{\beta}+e^{\xi}} - g - S_{y2} - i - z = 0,$$

$$e^{\xi}\left(e^{\theta_{n+1}} + m\right) + \frac{W_1 e^{\xi}}{1+e^{\alpha}+e^{\xi}} + \frac{W_2 e^{\xi}}{1+e^{\beta}+e^{\xi}} - (n+1)k - l = 0.$$

Then, if we define

$$m^* = e^{\theta^*_{n+1}+\alpha^*}, \qquad p^* = e^{\theta^*_{n+1}+\beta^*}, \qquad b^* = e^{\theta^*_{n+1}+\xi^*}, \qquad a^* = e^{\theta^*_{n+1}} + m^* + p^* + b^*,$$

$$c^* = e^{\theta^*_{n+1}+\alpha^*} + \frac{(h+u)ve^{\alpha^*}}{\left(e^{\alpha^*}+v\right)^2} + \frac{W_1 e^{\alpha^*}\left(1+e^{\xi^*}\right)}{\left(1+e^{\alpha^*}+e^{\xi^*}\right)^2}, \qquad d^* = -\frac{W_1 e^{\alpha^*+\xi^*}}{\left(1+e^{\alpha^*}+e^{\xi^*}\right)^2},$$

$$f^* = e^{\theta^*_{n+1}+\beta^*} + \frac{(g+r)se^{\beta^*}}{\left(e^{\beta^*}+s\right)^2} + \frac{W_2 e^{\beta^*}\left(1+e^{\xi^*}\right)}{\left(1+e^{\beta^*}+e^{\xi^*}\right)^2}, \qquad q^* = -\frac{W_2 e^{\beta^*+\xi^*}}{\left(1+e^{\beta^*}+e^{\xi^*}\right)^2},$$

$$t^* = e^{\xi^*}\left(e^{\theta^*_{n+1}} + m\right) + \frac{W_1 e^{\xi^*}\left(1+e^{\alpha^*}\right)}{\left(1+e^{\alpha^*}+e^{\xi^*}\right)^2} + \frac{W_2 e^{\xi^*}\left(1+e^{\beta^*}\right)}{\left(1+e^{\beta^*}+e^{\xi^*}\right)^2},$$

we can write σ^*, defined by (1.16), as

$$\sigma^* = \Big\{a^*c^*f^*t^* - a^*c^*q^{*2} - a^*f^*d^{*2} - f^*t^*m^{*2} + m^{*2}q^{*2} - 2m^*d^*p^*q^* +$$

$$+ 2m^*d^*b^*f^* - c^*t^*p^{*2} + 2p^*c^*b^*q^* + d^{*2}p^{*2} - c^*f^*b^{*2}\Big\}^{-1/2},$$

and the Laplace approximation for (3.24) will be

$$p\left(z \mid D^n, x_{n+1}\right) \tilde{\propto} \sum_{i=max(0,-z)}^{\infty} \frac{1}{i!(i+z)!}\left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\Big\{-nh^*\left(\theta^*_{n+1}, \alpha^*, \beta^*, \xi^*\right) +$$

$$+ nh_1\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\Big\}.$$

No explicit solution for (3.25) can be obtained assuming vague second stage priors.

3.3. Example and Conclusions

The methods used to solve the "treatment effects in a biased allocation model" problem will now be illustrated by an application with a simulated data set.

Let us consider the data shown in Table 3.1. These n=20 values were simulated from models with $e^{\alpha} = 0.3$ and $e^{\beta} = 1.4$, with a=6, and for a random selection of θ_i values.

x_i	δ_i	y_i	x_i	δ_i	y_i	x_i	δ_i	y_i	x_i	δ_i	y_i
7	0	3	2	1	2	11	0	22	17	0	16
8	0	9	13	0	12	6	0	13	3	1	0
9	0	9	4	1	2	9	0	10	2	1	2
13	0	16	6	0	2	6	0	10	2	1	1
5	1	1	7	0	4	10	0	16	8	0	11

Table 3.1: Simulated data set

Predictions are given for $y_{21,1}$ and $y_{21,2}$ corresponding to $x_{21} = 4$. In order to be able to compare the exact predictive probabilities with the ones obtained with the estimative and approximative methods, we assume a vague second stage prior structure $(l, m, u, v, r, s \to 0)$.

Matching the first two marginal moments of the X_is, as in section 2.2.5, we take

$$k = \frac{\bar{x}^2}{s_x^2 - \bar{x}}.$$

In fact, we rounded this value, taking k=6.

Figure 3.1 shows the joint predictive distribution of (Y_{211}, Y_{212}) derived using the Gibbs sampling algorithm (section 3.2.2) with M=500 and t=100. When we considered the asymptotic normal approximation suggested by Bernardo & Smith (1994), developed in section 3.2.3.1, we obtained the joint predictive distribution shown in Figure 3.2. This result was obtained performing the Gibbs sampling routine using the full conditional distributions (3.14) with M=500 and t=1. Performing the Gibbs sampling algorithm with

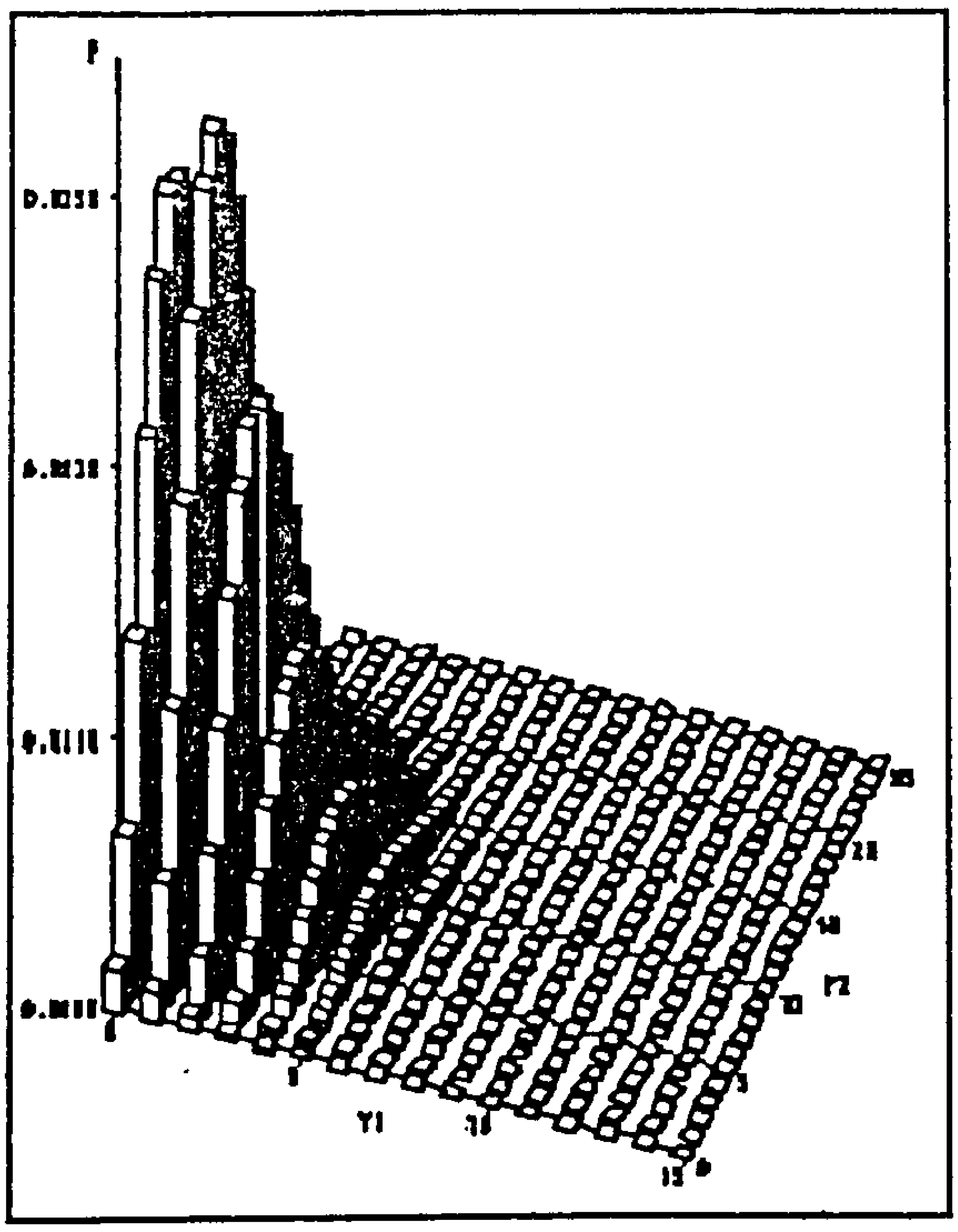

Figure 3.1: The joint predictive distribution (Gibbs Sampling Algorithm)

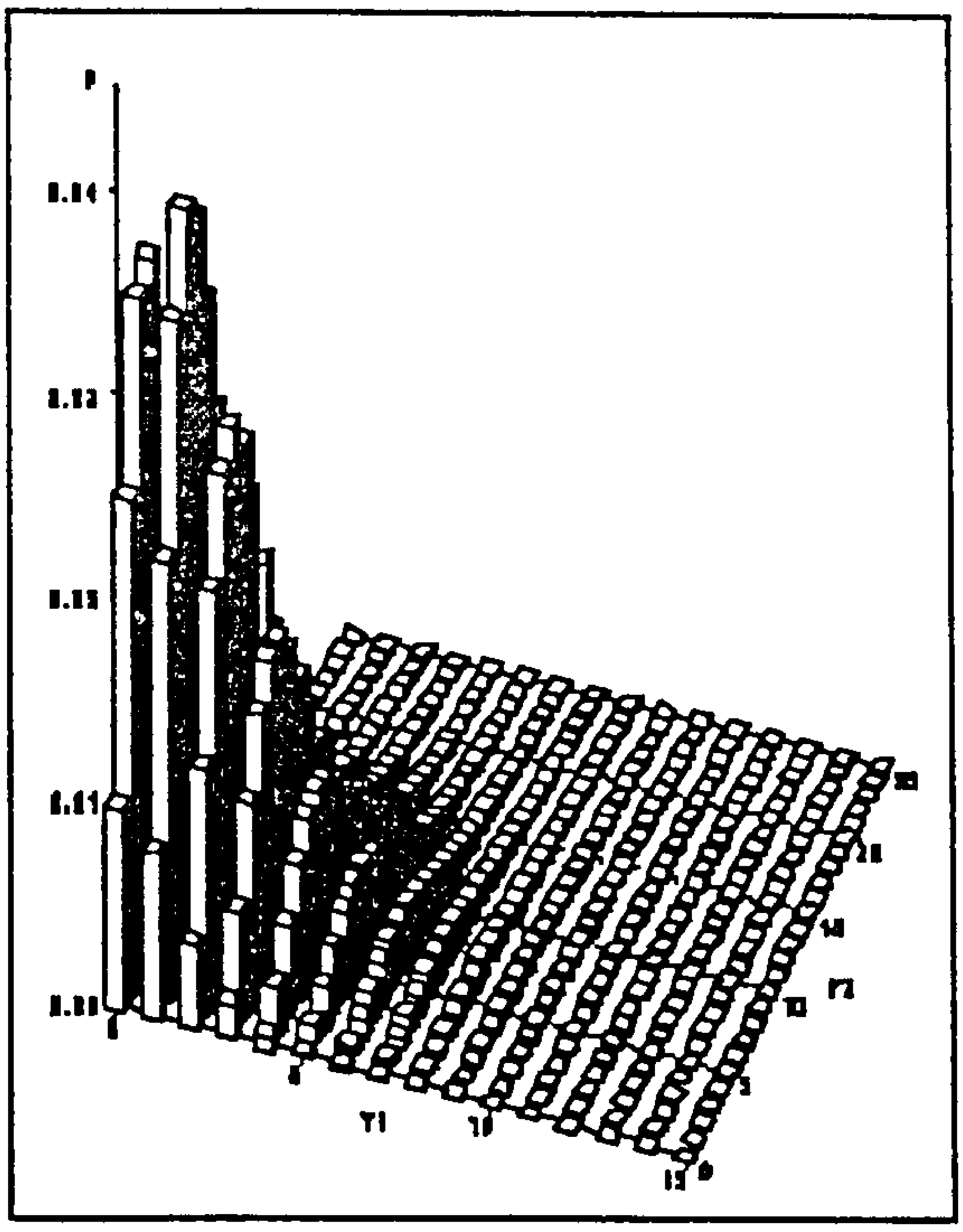

Figure 3.2: The joint predictive distribution (Bernardo & Smith suggestion)

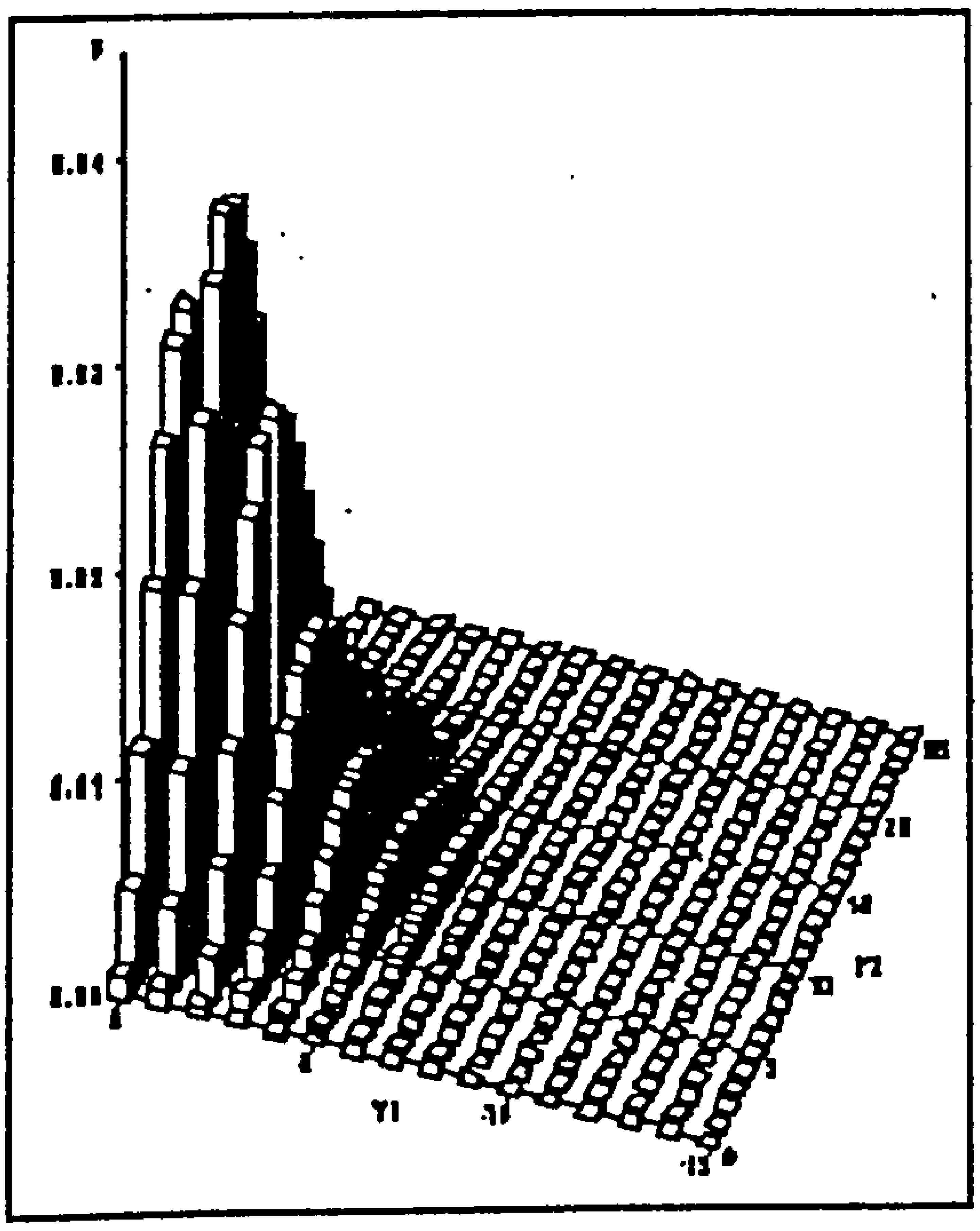

Figure 3.3: The joint predictive distribution (O'Hagan's suggestion)

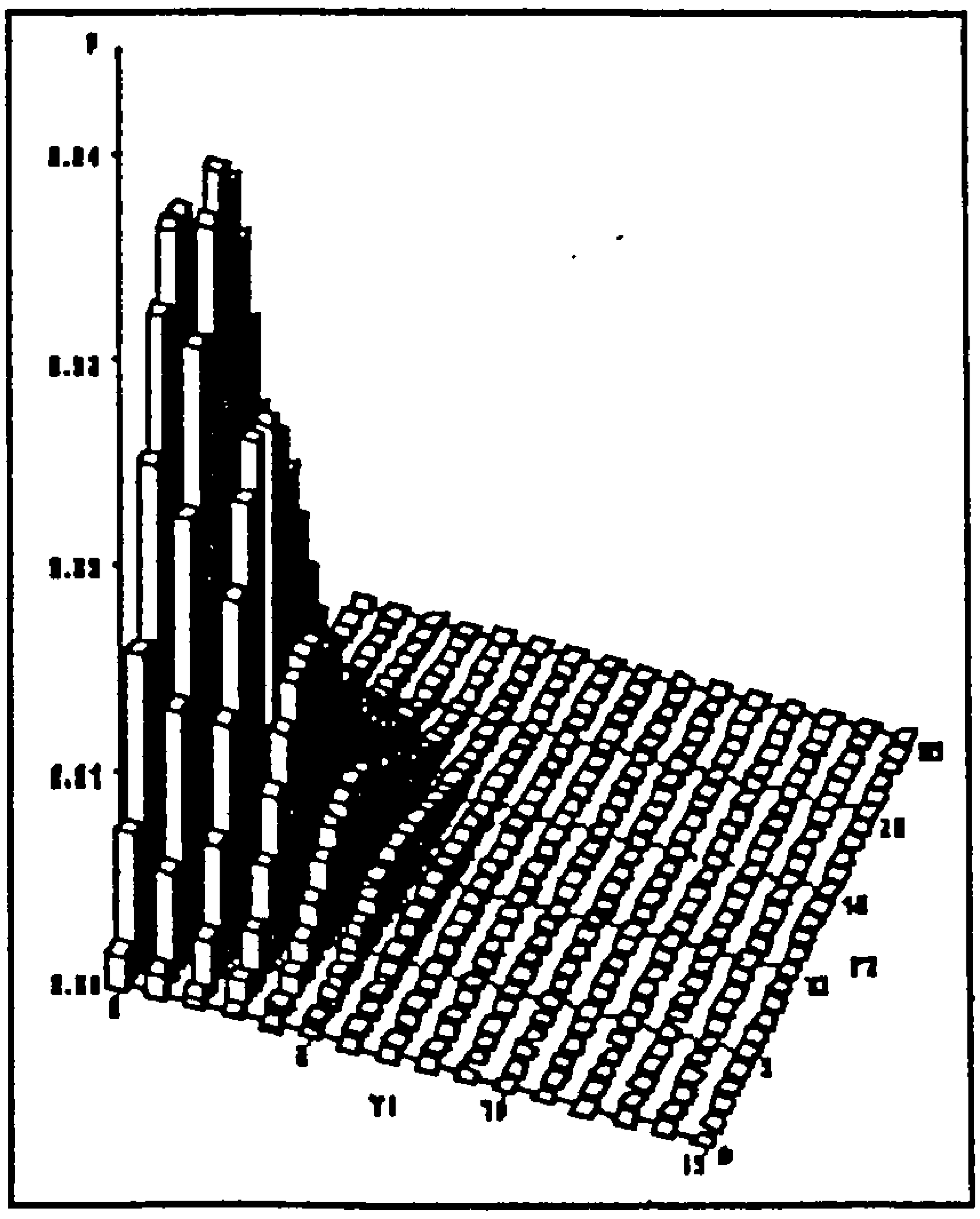

Figure 3.4: The joint predictive distribution (Laplace approximation)

the full conditional distributions (3.18), with $M=500$ and $t=100$ for the O'Hagan's (1994) approach developed in section 3.2.3.2, we obtained the joint predictive distribution presented in Figure 3.3. Figure 3.4 shows the joint predictive distribution obtained when we used the Laplace approximation, developed in section 3.2.4.

Observation of these four figures allows us to notice how remarkably similar the results are when using the different methods to obtain the joint predictive probabilities.

A clearer picture emerges if we consider the marginal predictive functions for Y_{211} and Y_{212} separately. Figures 3.5 and 3.6 show the different approximations and estimates together with the exact forms from (3.10) and (3.11). We also included in these figures the plug-in estimates for the predictive distributions, derived in section 3.1. Normal approximation 1 refers to the Bernardo & Smith (1994) approach and normal approximation 2 refers to O'Hagan's (1994) approach (see Tables A3.3 and A3.4, Appendix 3).

Observation of these figures allows us to confirm the conclusions drawn in section 2.2.5, namely, that the Laplace and the Gibbs methods are excellent when used to approximate the exact predictive distributions. The posterior normal approximations are not so reliable although O'Hagan's (1994) suggestion seems to be much superior.

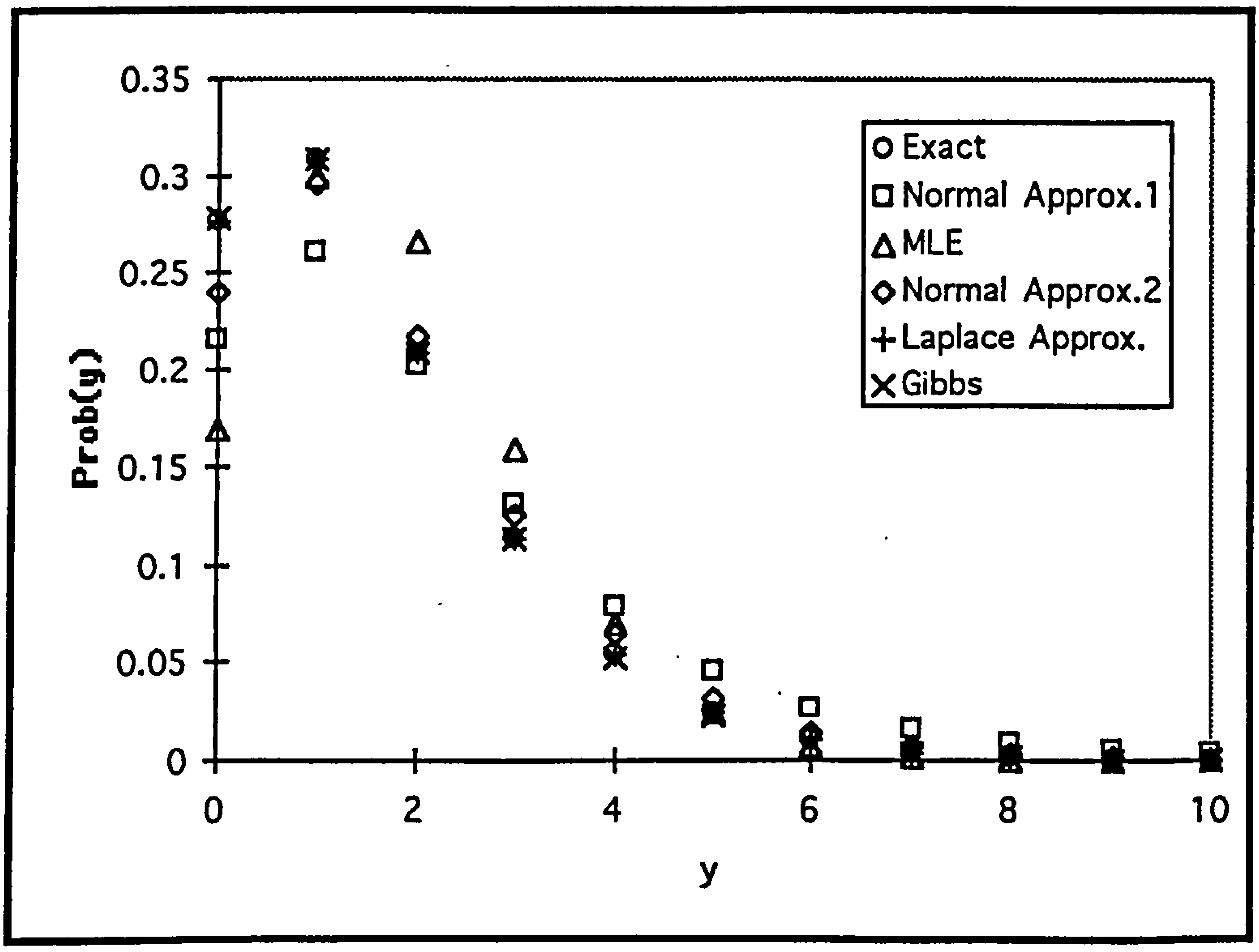

Figure 3.5: The predictive distribution of Y_{211} (the exact one and its approximations and estimates)

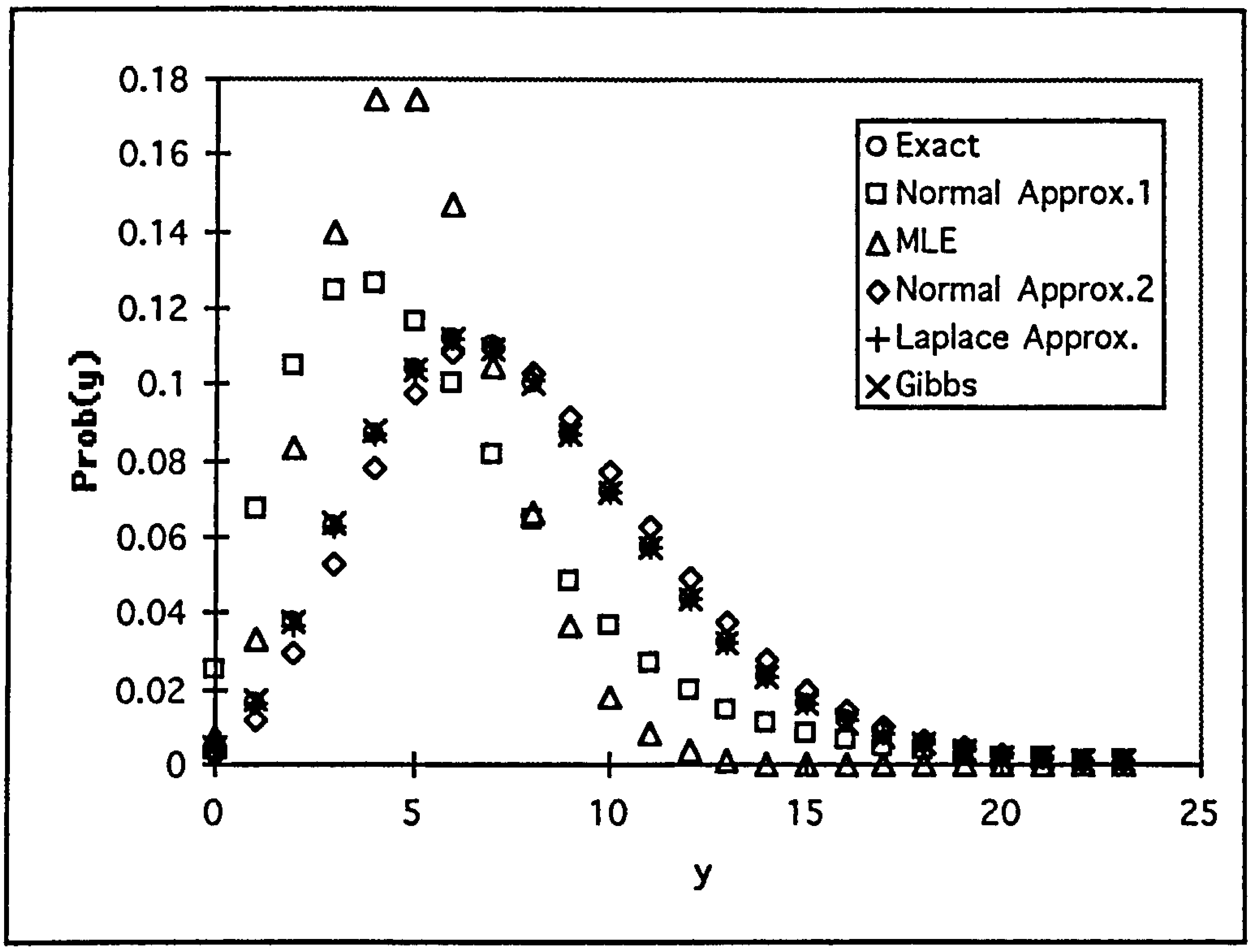

Figure 3.6: The predictive distribution of Y_{212} (the exact one and its approximations and estimates)

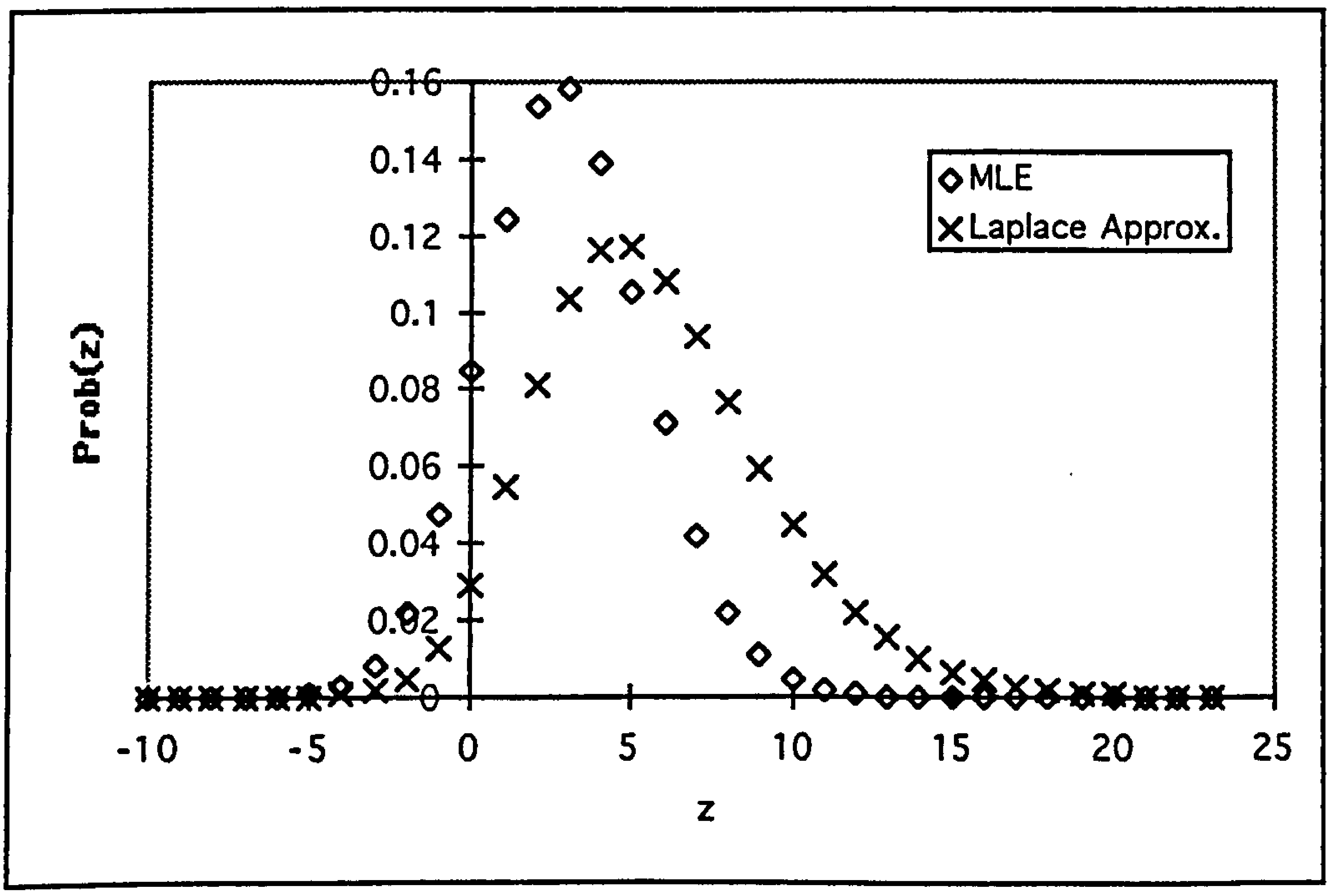

Figure 3.7: The predictive distribution of $Z = Y_{212} - Y_{211}$ (the plug-in estimate and the Laplace approximation)

Finally, Figure 3.7 shows the plug-in estimate and the Laplace approximation for the predictive distribution of $Z = Y_{21,2} - Y_{21,1}$, derived, respectively, in sections 3.1 and 3.2.4.5 (see Table A3.5, Appendix 3). This figure shows the unsatisfactory result of a plug-in approach.

3.4. Generalisation for the case of $k \geq 2$ treatments

In this section we will generalise the problem to the case where there are $k \geq 2$ possible treatments, $T_1, T_2, \ldots, T_k$, available to give to an individual. We assume that X_i and Y_i, $i=1, 2, \ldots, n$, are random variables representing, respectively, the observation on the i-th individual before any treatment is used and after one of the treatments is applied. We also assume that

- given θ_i, $X_i \sim Po(\exp(\theta_i))$,
- given θ_i, $\alpha_1, \alpha_2, \ldots, \alpha_k$ and x_i, $Y_i \sim Po(\exp(\alpha_j + \theta_i))$ if treatment T_j is used.

The parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ are used to model the multiplicative effects of the treatments and θ_i is used to model the particular characteristics of the i-th individual.

As in the former sections, the design of this problem supposes that the choice of the treatment to be used is based on the observed value x_i. Let Ω denote the set of all possible values for x_i and let us define $C_1, C_2, \ldots, C_k$ as subsets of Ω such that

$$\bigcup_{i=1}^{k} C_i = \Omega \qquad \text{and} \qquad C_i \cap C_j = \varnothing, \quad i \neq j, \quad i, j = 1, 2, \ldots, k.$$

Now, the choice of the treatment to use is based on the rule: if $x_i \in C_j$ we give treatment T_j to individual i.

Given a new individual, with observed value x_{n+1}, we want to predict its outcome after one of the treatments is used. For that, we define the random variables $Y_{n+1,j}$, $j=1, 2,$

..., k, representing the outcome on the $(n+1)$-th individual after T_j is applied, and we assume that

$$X_{n+1} \sim Po\left(\exp(\theta_{n+1})\right),$$
$$Y_{n+1,j} \sim Po\left(\exp\left(\alpha_j + \theta_{n+1}\right)\right), \quad j = 1, 2, \ldots, k.$$

For $j=1, 2, \ldots, k$, the predictive distribution of $Y_{n+1,j}$ will be given by

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) = \int_{\Re^2} p\left(y_{n+1,j} \mid \theta_{n+1}, \alpha_j\right) p\left(\theta_{n+1}, \alpha_j \mid D^n, x_{n+1}\right) d\theta_{n+1} d\alpha_j,$$

where D^n is the data set formed by the observations recorded on the n individuals, that is

$$D^n = \left\{(x_i, y_i),\, i = 1, 2, \ldots, n\right\}.$$

Let us define a treatment indicator function

$$\delta_{ij} = \begin{cases} 1, & \text{if treatment } T_j \text{ is given to individual } i \\ 0, & \text{otherwise} \end{cases}$$

Note that $\sum_{j=1}^{k} \delta_{ij} = 1$, for $i = 1, 2, \ldots, n$.

Robbins and Zhang's (1991) model can be extended to the case where we have $k \geq 2$ possible treatments. Such a model will be

- given ψ_i, $X_i \sim Po(\psi_i)$,
- given $\psi_i, \gamma_1, \gamma_2, \ldots, \gamma_k$ and x_i, $Y_i \sim Po\left(\gamma_j \psi_i\right)$ if treatment T_j is used,
- T_j is used if $x_i \in C_j$,

and the consistent estimates for γ_j, $j = 1, 2, \ldots, k$, are

$$\hat{\gamma}_j = \frac{\sum_{i=1}^{n} y_i I\left(x_i \in C_j\right)}{\sum_{i=1}^{n} x_i I\left(\left(x_i - 1\right) \in C_j\right)}.$$

3.4.1. A Classical Approach

Let $\theta^n = (\theta_1, \theta_2, \ldots, \theta_n)$ and $\alpha^k = (\alpha_1, \alpha_2, \ldots, \alpha_k)$. Generalising the results in section 3.1, we find that the maximum likelihood estimates are given by

$$\hat{\theta}_i = \ln\left(\frac{x_i + y_i}{1 + \sum_{j=1}^{k} \delta_{ij} e^{\hat{\alpha}_j}}\right), \quad i = 1, 2, \ldots, n, \qquad \hat{\theta}_{n+1} = \ln(x_{n+1}),$$

$$\hat{\alpha}_j = \ln\left(\frac{S_{yj}}{S_{xj}}\right), \quad j = 1, 2, \ldots, k,$$

where

$$S_{xj} = \sum_{i=1}^{n} x_i \delta_{ij}, \quad S_{yj} = \sum_{i=1}^{n} y_i \delta_{ij}, \quad j = 1, 2, \ldots, k.$$

These maximum likelihood estimates can now be used to obtain a simple plug-in estimate for the predictive distribution of $Y_{n+1,j}$, which will be given by

$$\hat{p}(y_{n+1,j} \mid D^n, x_{n+1}) = Po\left(\exp\left(\hat{\alpha}_j + \hat{\theta}_{n+1}\right)\right), \quad j = 1, 2, \ldots, k.$$

3.4.2. A Bayesian Approach

We again consider a hierarchical prior structure. At the first stage we take

$$p(\theta^n, \theta_{n+1}, \alpha^k \mid \xi, \eta^k) = \prod_{i=1}^{n+1}\{p(\theta_i \mid \xi)\}\prod_{j=1}^{k}\{p(\alpha_j \mid \eta_j)\}$$

where $\eta^k = (\eta_1, \eta_2, \ldots, \eta_k)$, and at the second stage we take

$$p(\xi, \eta^k) = p(\xi)\prod_{j=1}^{k}\{p(\eta_j)\},$$

with

$$e^{\theta_i} \sim Ga\left(h, e^{\xi}\right), i = 1, 2, \ldots, n+1 \qquad e^{\alpha_j} \sim Ga\left(g_j, e^{\eta_j}\right), j = 1, 2, \ldots, k$$

$$e^{\xi} \sim Ga(u, v) \qquad e^{\eta_j} \sim Ga\left(r_j, s_j\right), j = 1, 2, \ldots, k.$$

By analyses similar to section 3.2.1 we find, after eliminating $\theta_1, \theta_2, \ldots, \theta_n, \eta_1, \eta_2, \ldots, \eta_k$, that the marginal posterior distribution for the parameter vector $\left(\theta_{n+1}, \alpha^k, \xi\right)$ is given by

$$p\left(\theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, \xi \mid D^n, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\xi+\theta_{n+1}} - ve^{\xi} - e^{\theta_{n+1}}\right\}}{\prod_{j=1}^{k}\left\{\left(1+e^{\alpha_j}+e^{\xi}\right)^{W_j}\right\}} \times$$

$$\times \frac{\exp\left\{((n+1)h+u)\xi + \theta_{n+1}\left(x_{n+1}+h\right) + \sum_{j=1}^{k}\alpha_j\left(S_{yj}+g_j\right)\right\}}{\prod_{j=1}^{k}\left\{\left(e^{\alpha_j}+s_j\right)^{g_j+r_j}\right\}}$$

where

$$W_j = \sum_{i=1}^{n}\delta_{ij}\left(x_i + y_i + h\right) = S_{xj} + S_{yj} + \sum_{i=1}^{n}\delta_{ij}h,$$

for $j=1, 2, \ldots, k$. The marginal predictive distributions of $Y_{n+1,j}$ $\left(j = 1, 2, \ldots, k\right)$ are then given by

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) = \int_{\Re^{k+2}} p\left(y_{n+1,j} \mid \theta_{n+1}, \alpha_j\right) p\left(\theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, \xi \mid D^n, x_{n+1}\right) d\theta_{n+1}$$
$$d\alpha_1\, d\alpha_2 \ldots d\alpha_k\, d\xi$$

$$\propto \frac{1}{y_{n+1,j}} \int_{\Re^{k+2}} \frac{\exp\left\{-e^{\theta_{n+1}}\left(1+e^{\alpha_j}+e^{\xi}\right) - ve^{\xi}\right\}\exp\left\{((n+1)h+u)\xi\right\}}{\prod_{p=1}^{k}\left\{\left(1+e^{\alpha_p}+e^{\xi}\right)^{W_p}\right\}} \times$$

$$\times \frac{\exp\left\{\theta_{n+1}\left(h + x_{n+1} + y_{n+1,j}\right) + \alpha_j\left(g_j + S_{yj} + y_{n+1,j}\right)\right\}}{\prod_{p=1}^{k}\left\{\left(e^{\alpha_p} + s_p\right)^{g_p + r_p}\right\}} \times$$

$$\times \exp\left\{\sum_{\substack{p=1 \\ p \neq j}}^{k} \alpha_p\left(g_p + S_{yp}\right)\right\} d\theta_{n+1}\, d\alpha_1\, d\alpha_2 \ldots d\alpha_k\, d\xi.$$

Notice that these involve $(k+2)$-dimensional integrations which are very likely to lead to numerical problems, so that estimative and approximative methods must be used.

If we consider second stage vague priors $\left(u, v \to 0 \text{ and } r_j, s_j \to 0, j = 1, 2, \ldots, k\right)$, the predictive distributions of $Y_{n+1,j}$ can be obtained explicitly, and are given by

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) \propto \frac{B\left(h + x_{n+1} + y_{n+1,j}, W_j\right)}{B\left(y_{n+1,j} + 1, S_{yj} - 1\right)}.$$

3.4.2.1. Estimation Via Gibbs Sampling

For the Gibbs sampling algorithm approach (section 1.2.1) the full conditional distributions in (3.12) generalise to

$$p\left(\theta_{n+1} \mid \alpha_1, \alpha_2, \ldots, \alpha_k, \xi, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\theta_{n+1}}\left(1 + e^{\xi}\right)\right\} \exp\left\{\theta_{n+1}\left(h + x_{n+1}\right)\right\},$$

$$p\left(\xi \mid \theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, D^n, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}} + v\right)\right\} \exp\left\{\left((n+1)h + u\right)\xi\right\}}{\prod_{j=1}^{k}\left\{\left(1 + e^{\alpha_j} + e^{\xi}\right)^{W_j}\right\}},$$

$$p\left(\alpha_j \mid \theta_{n+1}, \alpha_{i \neq j}, \xi, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\alpha_j\left(g_j + S_{yj}\right)\right\}}{\left(1 + e^{\alpha_j} + e^{\xi}\right)^{W_j}\left(e^{\alpha_j} + s_j\right)^{g_j + r_j}}, \quad j = 1, 2, \ldots, k.$$

If we consider vague second stage priors, the full conditional distributions above become

$$p\left(\theta_{n+1} \mid \alpha_1, \alpha_2, \ldots, \alpha_k, \xi, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\theta_{n+1}}\left(1+e^{\xi}\right)\right\}\exp\left\{\theta_{n+1}\left(h+x_{n+1}\right)\right\},$$

$$p\left(\xi \mid \theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, D^n, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\xi+\theta_{n+1}}\right\}\exp\{(n+1)h\xi\}}{\prod_{j=1}^{k}\left\{\left(1+e^{\alpha_j}+e^{\xi}\right)^{W_j}\right\}},$$

$$p\left(\alpha_j \mid \theta_{n+1}, \alpha_{i\neq j}, \xi, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\alpha_j S_{yj}\right\}}{\left(1+e^{\alpha_j}+e^{\xi}\right)^{W_j}}, \quad j = 1, 2, \ldots, k.$$

3.4.2.2. Estimation Via Asymptotic Results

Following Bernardo & Smith (1994) and section 3.2.3.1, we obtain that θ_{n+1}, α_1, $\alpha_2, \ldots, \alpha_k$ are independent a posteriori with

$$p\left(\theta_{n+1} \mid \alpha_1, \alpha_2, \ldots, \alpha_k, D^n, x_{n+1}\right) = N\left(\hat{\theta}_{n+1}, \frac{1}{e^{\hat{\theta}_{n+1}}}\right) = N\left(\ln\left(x_{n+1}\right), \frac{1}{x_{n+1}}\right),$$

$$p\left(\alpha_j \mid \theta_{n+1}, \alpha_{i\neq j}, D^n, x_{n+1}\right) = N\left(\hat{\alpha}_j, \left(\sum_{i=1}^{n}\left\{\delta_{ij}\, e^{\hat{\alpha}_j+\hat{\theta}_i}\left(1-\frac{e^{\hat{\alpha}_j}}{1+\sum_{p=1}^{k}\delta_{ip}e^{\hat{\alpha}_p}}\right)\right\}\right)^{-1}\right),$$

$j=1, 2, \ldots, k$.

Problems arise if $x_{n+1} = 0$ or if at least one of the summations S_{xj}, S_{yj} ($j=1, 2, \ldots, k$) is null or if $x_i = y_i = 0$ for at least one $i = 1, 2, \ldots, n$.

In a similar fashion for the O'Hagan's (1994) approximation, the full conditional distributions in (3.16) generalise to

$$p\left(\theta_{n+1} \mid \alpha_1, \alpha_2, \ldots, \alpha_k, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_{n+1} - \frac{\tilde{b}}{\tilde{a}}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{a}}\right),$$

$$p\left(\alpha_j \mid \theta_{n+1}, \alpha_{i\neq j}, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\alpha}_j - \frac{\tilde{d}_j}{\tilde{c}_j}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{c}_j}\right), \; j=1, 2, \ldots, k,$$

$$p\left(\xi \mid \theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, D^n, x_{n+1}\right) = N\left(\tilde{\xi} - \frac{\left(\theta_{n+1} - \tilde{\theta}_{n+1}\right)\tilde{b} + \sum_{j=1}^{k}\left(\alpha_j - \tilde{\alpha}_j\right)\tilde{d}_j}{\tilde{f}}, \frac{1}{\tilde{f}}\right),$$

where $\tilde{\theta}_{n+1}, \tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_k, \tilde{\xi}$ result from the solution of the system formed by the equations

$$e^{\theta_{n+1}}\left(1 + e^{\xi}\right) - h - x_{n+1} = 0,$$

$$e^{\xi}\left(e^{\theta_{n+1}} + v + \sum_{j=1}^{k} \frac{W_j}{1 + e^{\alpha_j} + e^{\xi}}\right) - (n+1)h - u = 0, \tag{3.26}$$

$$S_{yj} + g_j - \frac{W_j e^{\alpha_j}}{1 + e^{\alpha_j} + e^{\xi}} - \frac{\left(g_j + r_j\right)e^{\alpha_j}}{e^{\alpha_j} + s_j} = 0 \quad , \quad j = 1, 2, \ldots, k,$$

and

$$\tilde{a} = e^{\tilde{\theta}_{n+1}}\left(1 + e^{\tilde{\xi}}\right), \qquad \tilde{b} = e^{\tilde{\xi} + \tilde{\theta}_{n+1}},$$

$$\tilde{c}_j = \frac{W_j e^{\tilde{\alpha}_j}\left(1 + e^{\tilde{\xi}}\right)}{\left(1 + e^{\tilde{\alpha}_j} + e^{\tilde{\xi}}\right)^2} + \frac{\left(g_j + r_j\right)s_j e^{\tilde{\alpha}_j}}{\left(e^{\tilde{\alpha}_j} + s_j\right)^2}, \quad \tilde{d}_j = -\frac{W_j e^{\tilde{\alpha}_j + \tilde{\xi}}}{\left(1 + e^{\tilde{\alpha}_j} + e^{\tilde{\xi}}\right)^2}, \; j = 1, 2, \ldots, k,$$

$$\tilde{f} = e^{\tilde{\xi}}\left(e^{\tilde{\theta}_{n+1}} + v + \sum_{j=1}^{k} \frac{W_j\left(1 + e^{\tilde{\alpha}_j}\right)}{\left(1 + e^{\tilde{\alpha}_j} + e^{\tilde{\xi}}\right)^2}\right). \tag{3.27}$$

When we consider second stage vague priors, the resulting simplifications are mainly due to the possibility of obtaining explicitly the solution for the system of equations (3.26), namely

$$\tilde{\theta}_{n+1} = \ln\left\{\frac{(T_x + x_{n+1})(x_{n+1} + h)}{T_x + x_{n+1} + (n+1)h}\right\}, \quad \tilde{\xi} = \ln\left\{\frac{(n+1)h}{T_x + x_{n+1}}\right\},$$

$$\tilde{\alpha}_j = \ln\left\{\frac{S_{yj}(T_x + x_{n+1} + (n+1)h)}{(T_x + x_{n+1})(W_j - S_{yj})}\right\} \quad , \quad j = 1, 2, \ldots, k, \qquad (3.28)$$

with

$$T_x = \sum_{i=1}^{n} x_i .$$

3.4.2.3. Laplace Approximation

The results obtained in section 3.2.4 can be generalised easily. The Laplace approximation for the predictive distribution of $Y_{n+1,j}$ is

$$p(y_{n+1,j} \mid D^n, x_{n+1}) \;\tilde{\propto}\; \frac{1}{y_{n+1,j}!}\left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\{-nh^*(\theta^*_{n+1}, \alpha^*_1, \alpha^*_2, \ldots, \alpha^*_k, \xi^*) +$$

$$+ nh(\tilde{\theta}_{n+1}, \tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_k, \tilde{\xi})\},$$

with the functions $h(\theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, \xi)$ and $h^*(\theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, \xi)$ being defined such that

$$-nh(\theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, \xi) = -e^{\xi+\theta_{n+1}} - ve^{\xi} - e^{\theta_{n+1}} + ((n+1)h + u)\xi +$$

$$+ \theta_{n+1}(h + x_{n+1}) + \sum_{p=1}^{k} \alpha_p (S_{yp} + g_p) - \sum_{p=1}^{k} \{W_p \ln(1 + e^{\alpha_p} + e^{\xi})\} -$$

$$- \sum_{p=1}^{k} \{(g_p + r_p)\ln(e^{\alpha_p} + s_p)\}$$

and

$$-nh^*(\theta_{n+1}, \alpha_1, \alpha_2, \ldots, \alpha_k, \xi) = -e^{\alpha_j+\theta_{n+1}} + (\alpha_j + \theta_{n+1})y_{n+1,j} - e^{\xi+\theta_{n+1}} - ve^{\xi} - e^{\theta_{n+1}} +$$

$$+((n+1)h+u)\xi+\theta_{n+1}(h+x_{n+1})+\sum_{p=1}^{k}\alpha_p(S_{yp}+g_p)-$$

$$-\sum_{p=1}^{k}\left\{W_p\ln\left(1+e^{\alpha_p}+e^{\xi}\right)\right\}-\sum_{p=1}^{k}\left\{(g_p+r_p)\ln\left(e^{\alpha_p}+s_p\right)\right\}.$$

The vector $\left(\tilde{\theta}_{n+1},\tilde{\alpha}_1,\tilde{\alpha}_2,\ldots,\tilde{\alpha}_k,\tilde{\xi}\right)$ is the solution of (3.26), obtained through a numerical method and

$$\tilde{\sigma}=\left\{\left(\frac{\tilde{b}^2}{\tilde{a}}+\sum_{p=1}^{k}\frac{\tilde{d}_p^2}{\tilde{c}_p}\right)\tilde{a}\prod_{p=1}^{k}\{\tilde{c}_p\}\right\}^{-1/2},$$

with $\tilde{a}$, $\tilde{b}$, $\tilde{c}_p$ and $\tilde{d}_p$ being defined as in (3.27). Solving numerically the system formed by the equations

$$e^{\theta_{n+1}}\left(1+e^{\alpha_j}+e^{\xi}\right)-h-x_{n+1}-y_{n+1,j}=0,$$

$$e^{\alpha_j+\theta_{n+1}}+\frac{W_je^{\alpha_j}}{1+e^{\alpha_j}+e^{\xi}}+\frac{(g_j+r_j)e^{\alpha_j}}{e^{\alpha_j}+s_j}-g_j-S_{yj}-y_{n+1,j}=0, \tag{3.29}$$

$$\frac{W_pe^{\alpha_p}}{1+e^{\alpha_p}+e^{\xi}}+\frac{(g_p+r_p)e^{\alpha_p}}{e^{\alpha_p}+s_p}-g_p-S_{yp}=0,\quad p=1,2,\ldots,k;\quad p\neq j,$$

$$e^{\xi}\left(e^{\theta_{n+1}}+v+\sum_{p=1}^{k}\frac{W_p}{1+e^{\alpha_p}+e^{\xi}}\right)-(n+1)h-u=0,$$

we obtain $\left(\theta^*_{n+1},\alpha^*_1,\alpha^*_2,\ldots,\alpha^*_k,\xi^*\right)$ and σ^* is defined by

$$\sigma^*=\left\{q^*\prod_{\substack{p=1\\p\neq j}}^{k}\{c^*_p\}\left[\left(f^*-\frac{d_j^{*2}}{q^*}-\sum_{\substack{p=1\\p\neq j}}^{k}\frac{d_p^{*2}}{c^*_p}\right)\left(a^*-\frac{t^{*2}}{q^*}\right)-\left(b^*-\frac{d^*_jt^*}{q^*}\right)^2\right]\right\}^{-1/2},$$

where

$$t^*=e^{\alpha^*_j+\theta^*_{n+1}},\qquad b^*=e^{\xi^*+\theta^*_{n+1}},\qquad a^*=e^{\theta^*_{n+1}}+t^*+b^*,$$

$$c_p^* = \frac{W_p e^{\alpha_p^*}\left(1+e^{\xi^*}\right)}{\left(1+e^{\alpha_p^*}+e^{\xi^*}\right)^2} + \frac{\left(g_p+r_p\right)s_p e^{\alpha_p^*}}{\left(e^{\alpha_p^*}+s_p\right)^2}, \quad p = 1, 2, \ldots, k,$$

$$q^* = t^* + c_j^*, \qquad d_p^* = -\frac{W_p e^{\alpha_p^*+\xi^*}}{\left(1+e^{\alpha_p^*}+e^{\xi^*}\right)^2}, \quad p = 1, 2, \ldots, k,$$

$$f^* = e^{\xi^*}\left(e^{\theta_{n+1}^*} + v + \sum_{p=1}^{k} \frac{W_p\left(1+e^{\alpha_p^*}\right)}{\left(1+e^{\alpha_p^*}+e^{\xi^*}\right)^2}\right).$$

When considering second stage vague priors, the most significant simplifications are due to the fact that we are able to solve explicitly the systems of equations (3.26) and (3.29), which originate $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_k, \tilde{\xi}\right)$ and $\left(\theta_{n+1}^*, \alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*, \xi^*\right)$, respectively. In such a case, $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_k, \tilde{\xi}\right)$ is given by (3.28) and the components of the vector $\left(\theta_{n+1}^*, \alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*, \xi^*\right)$ become

$$\theta_{n+1}^* = \ln\left\{\frac{\left(T_x + x_{n+1}\right)\left(x_{n+1} + y_{n+1,j} + h\right)\left(W_j - S_{yj} + x_{n+1} + h\right)}{\left(W_j + x_{n+1} + y_{n+1,j} + h\right)\left(T_x + x_{n+1} + (n+1)h\right)}\right\},$$

$$\alpha_j^* = \ln\left\{\frac{\left(T_x + x_{n+1} + (n+1)h\right)\left(S_{yj} + y_{n+1,j}\right)}{\left(T_x + x_{n+1}\right)\left(W_j - S_{yj} + x_{n+1} + h\right)}\right\},$$

$$\alpha_p^* = \ln\left\{\frac{\left(T_x + x_{n+1} + (n+1)h\right)S_{yp}}{\left(T_x + x_{n+1}\right)\left(W_p - S_{yp}\right)}\right\}, \quad p = 1, 2, \ldots, k; \quad p \neq j,$$

$$\xi^* = \ln\left\{\frac{(n+1)h}{T_x + x_{n+1}}\right\}.$$

CHAPTER 4

A CROSSOVER DESIGN TO COMPARE TWO TREATMENTS EFFECTS: PREDICTION IN A POISSON ERRORS IN VARIABLES MODEL

Let us suppose that we have two treatments, T_1 and T_2, both with multiplicative effects, and that we aim to compare their effects upon an individual, within a crossover situation. We assume that a two-period design with treatment orderings of the form T_1T_1, T_1T_2, T_2T_1 and T_2T_2 is used. We also suppose that for individual i ($i=1, 2, ..., n$), the responses in the two periods, represented by the random variables W_{i1} and W_{i2}, are Poisson counts given from the appropriate distributions as follows:

Treatment	Period 1	Period 2
T_1	$Po(\exp(\theta_i))$	$Po(\exp(\beta+\theta_i))$
T_2	$Po(\exp(\alpha+\theta_i))$	$Po(\exp(\alpha+\beta+\theta_i))$

Note that, in this model, we have to consider two types of effects: one caused by the treatment and another caused by the period of time.

The multiplicative treatment effect is modelled by the parameter α and the period effect, which we also suppose to be of a multiplicative form, is modelled by the parameter β. We also have to consider the existent dependence between W_{i1} and W_{i2} ($i=1, 2, ..., n$), due to the particular characteristics of individual i. Such dependence is modelled by the parameters $\theta_1, \theta_2, ..., \theta_n$, which can be seen as nuisance parameters.

Our aim is to derive the predictive distributions for the counts Z_1 and Z_2 for a future individual from treatment T_1 and T_2, respectively.

In order to simplify the development, we define dummy covariates δ_{i1} and δ_{i2} ($i=1, 2, \ldots, n$) to identify the treatment given to individual i in the first and second periods, respectively, through

$$\delta_{i1} = \begin{cases} 0, & \text{if treatment } T_1 \text{ is used in period 1} \\ 1, & \text{if treatment } T_2 \text{ is used in period 1} \end{cases}$$

$$\delta_{i2} = \begin{cases} 0, & \text{if treatment } T_1 \text{ is used in period 2} \\ 1, & \text{if treatment } T_2 \text{ is used in period 2} \end{cases} \tag{4.1}$$

Using these dummy covariates, the model we are considering can be written in the alternative form

$$\text{Period 1:} \quad W_{i1} \sim Po\left(\exp\left(\theta_i + \delta_{i1}\alpha\right)\right)$$

$$\text{Period 2:} \quad W_{i2} \sim Po\left(\exp\left(\theta_i + \beta + \delta_{i2}\alpha\right)\right)$$

The data available, D^n, is formed by the responses from the individuals in the two periods and by the information about which treatment each individual received in each period, that is,

$$D^n = \left\{\left(w_{i1}, w_{i2}, \delta_{i1}, \delta_{i2}\right), i = 1, 2, \ldots, n\right\}.$$

Based on this data D^n, we will derive the predictive distributions for the counts on a future individual when subjected to the two treatments. We are interested in the joint predictive distribution of $\left(Z_1, Z_2\right)$, and in both marginal predictive distributions. Note that, conditional on θ_{n+1}, Z_1 and Z_2 are independent random variables such that

$$Z_1 \sim Po\left(\exp\left(\theta_{n+1}\right)\right),$$

$$Z_2 \sim Po\left(\exp\left(\alpha + \theta_{n+1}\right)\right).$$

Dunsmore & Robson (1997) consider this model (in an alternative parameterisation). Here we develop in addition an asymptotic normal approximation for the posterior distribution and the Laplace approximation.

4.1. A Classical Approach

The likelihood function is given by

$$L(\theta_1, \theta_2, \ldots, \theta_n, \alpha, \beta; D^n) = \prod_{i=1}^{n} \left\{ \frac{\exp\left\{-e^{\theta_i + \delta_{i1}\alpha}\right\} \exp\left\{(\theta_i + \delta_{i1}\alpha) w_{i1}\right\}}{w_{i1}!} \right\} \times$$

$$\times \prod_{i=1}^{n} \left\{ \frac{\exp\left\{-e^{\theta_i + \beta + \delta_{i2}\alpha}\right\} \exp\left\{(\theta_i + \beta + \delta_{i2}\alpha) w_{i2}\right\}}{w_{i2}!} \right\}$$

$$\propto \exp\left\{ -\sum_{i=1}^{n} e^{\theta_i + \delta_{i1}\alpha} - \sum_{i=1}^{n} e^{\theta_i + \beta + \delta_{i2}\alpha} \right\} \times$$

$$\times \exp\left\{ \sum_{i=1}^{n} (\theta_i + \delta_{i1}\alpha) w_{i1} + \sum_{i=1}^{n} (\theta_i + \beta + \delta_{i2}\alpha) w_{i2} \right\}.$$

Thus, to obtain the maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$, $\hat{\alpha}$ and $\hat{\beta}$, we have to solve, using a numerical technique, a system formed by $(n+2)$ equations

$$e^{\theta_i + \delta_{i1}\alpha} + e^{\theta_i + \beta + \delta_{i2}\alpha} - w_{i1} - w_{i2} = 0, \quad i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n} \delta_{i1} e^{\theta_i + \delta_{i1}\alpha} + \sum_{i=1}^{n} \delta_{i2} e^{\theta_i + \beta + \delta_{i2}\alpha} - \sum_{i=1}^{n} \delta_{i1} w_{i1} - \sum_{i=1}^{n} \delta_{i2} w_{i2} = 0,$$

$$\sum_{i=1}^{n} e^{\theta_i + \beta + \delta_{i2}\alpha} - \sum_{i=1}^{n} w_{i2} = 0.$$

To get a plug-in estimate of the predictive distributions we would need to know the maximum likelihood estimates of θ_{n+1} and α. However, the system of equations above will not allow us to obtain $\hat{\theta}_{n+1}$, since the available data, D^n, does not provide any information about the future individual. Therefore, this kind of approach cannot be used at all to estimate in a classical framework the predictive distributions we seek.

4.2. A Bayesian Approach

Within a Bayesian framework, the predictive distributions for (Z_1, Z_2), Z_1 and Z_2 will be given, respectively, by

$$p(z_1, z_2 \mid D^n) = \int_{\Re^2} p(z_1 \mid \theta_{n+1}) p(z_2 \mid \theta_{n+1}, \alpha) p(\theta_{n+1}, \alpha \mid D^n) d\theta_{n+1} d\alpha,$$

$$p(z_1 \mid D^n) = \int_{\Re} p(z_1 \mid \theta_{n+1}) p(\theta_{n+1} \mid D^n) d\theta_{n+1},$$

$$p(z_2 \mid D^n) = \int_{\Re^2} p(z_2 \mid \theta_{n+1}, \alpha) p(\theta_{n+1}, \alpha \mid D^n) d\theta_{n+1} d\alpha .$$

We will again consider a hierarchical prior structure for the parameters. At the first stage we take

$$p(\theta^n, \theta_{n+1}, \alpha, \beta \mid \xi, \eta, \zeta) = \prod_{i=1}^{n} \{p(\theta_i \mid \xi)\} p(\alpha \mid \eta) p(\beta \mid \zeta),$$

and at the second stage we take

$$p(\xi, \eta, \zeta) = p(\xi) p(\eta) p(\zeta),$$

with

$$e^{\theta_i} \sim Ga(k, e^{\xi}) \qquad e^{\alpha} \sim Ga(h, e^{\eta}) \qquad e^{\beta} \sim Ga(g, e^{\zeta})$$

$$e^{\xi} \sim Ga(l, m) \qquad e^{\eta} \sim Ga(u, v) \qquad e^{\zeta} \sim Ga(r, s)$$

where k, h, g, l, m, u, v, r and s are assumed to be known.

4.2.1. The Exact Predictive Distributions

Using the hierarchical prior structure defined above with the equivalent forms in Table A1.1, the joint prior distribution for the parameter vector $(\theta^n, \theta_{n+1}, \alpha, \beta, \xi, \eta, \zeta)$ will be given by

$$p\left(\theta^n, \theta_{n+1}, \alpha, \beta, \xi, \eta, \zeta\right) = \propto \exp\left\{-\sum_{i=1}^{n+1} e^{\xi+\theta_i} - e^{\eta+\alpha} - e^{\zeta+\beta} - me^{\xi} - ve^{\eta} - se^{\zeta}\right\}\times$$

$$\times\exp\left\{k\sum_{i=1}^{n+1}\theta_i + h(\eta+\alpha) + g(\zeta+\beta) + u\eta + r\zeta + \right.$$

$$\left. +((n+1)k+l)\xi \right\}. \quad (4.3)$$

The joint posterior distribution, which results from (4.2) and (4.3), is given by

$$p\left(\theta^n, \theta_{n+1}, \alpha, \beta, \xi, \eta, \zeta \mid D^n\right) \propto \exp\left\{-\sum_{i=1}^{n+1} e^{\xi+\theta_i} - e^{\eta+\alpha} - e^{\zeta+\beta} - me^{\xi} - ve^{\eta} - se^{\zeta}\right\}\times$$

$$\times\exp\left\{-\sum_{i=1}^{n} e^{\theta_i+\delta_{i1}\alpha} - \sum_{i=1}^{n} e^{\theta_i+\beta+\delta_{i2}\alpha}\right\}\exp\left\{k\sum_{i=1}^{n+1}\theta_i\right\}\times$$

$$\times\exp\left\{h(\eta+\alpha) + g(\zeta+\beta) + u\eta + r\zeta + ((n+1)k+l)\xi\right\}\times$$

$$\times\exp\left\{\sum_{i=1}^{n}(\theta_i + \delta_{i1}\alpha)w_{i1} + \sum_{i=1}^{n}(\theta_i + \beta + \delta_{i2}\alpha)w_{i2}\right\}.$$

However, notice that Z_1 and Z_2 only depend on θ_{n+1} and α. Therefore, those are the parameters whose posterior distribution is required to derive the predictive distributions. Unfortunately, it is not possible to eliminate all the remaining parameters; the best we can do is to eliminate $\theta_1, \theta_2, \ldots, \theta_n, \eta$ and ζ to obtain the marginal posterior distribution of $(\theta_{n+1}, \alpha, \beta, \xi)$, which is

$$p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n\right) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}} + m\right)\right\}\exp\left\{((n+1)k+l)\xi + k\theta_{n+1}\right\}}{\prod_{i=1}^{n}\left\{\left(e^{\xi} + e^{\delta_{i1}\alpha} + e^{\beta+\delta_{i2}\alpha}\right)^{k+w_{i1}+w_{i2}}\right\}}\times$$

$$\times\frac{\exp\left\{\alpha\left(h + \sum_{i=1}^{n}\delta_{i1}w_{i1} + \sum_{i=1}^{n}\delta_{i2}w_{i2}\right) + \beta\left(g + \sum_{i=1}^{n}w_{i2}\right)\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}}. \quad (4.4)$$

Future results can be written in a much simpler form if we define

$$n_{11} = \sum_{i=1}^{n} (1-\delta_{i1})(1-\delta_{i2}) \qquad n_{12} = \sum_{i=1}^{n} (1-\delta_{i1})\delta_{i2}$$

$$n_{21} = \sum_{i=1}^{n} \delta_{i1}(1-\delta_{i2}) \qquad n_{22} = \sum_{i=1}^{n} \delta_{i1}\delta_{i2}$$

$$S_1 = \sum_{i=1}^{n} \delta_{i1} w_{i1} \qquad S_2 = \sum_{i=1}^{n} \delta_{i2} w_{i2}$$

$$T_{11} = \sum_{i=1}^{n} (1-\delta_{i1})(1-\delta_{i2})(k + w_{i1} + w_{i2}) \qquad T_{12} = \sum_{i=1}^{n} (1-\delta_{i1})\delta_{i2}(k + w_{i1} + w_{i2})$$

$$T_{21} = \sum_{i=1}^{n} \delta_{i1}(1-\delta_{i2})(k + w_{i1} + w_{i2}) \qquad T_{22} = \sum_{i=1}^{n} \delta_{i1}\delta_{i2}(k + w_{i1} + w_{i2})$$

Notice that n_{ij} is the number of individuals who receive treatment T_i in period 1 and treatment T_j in period 2. Using this new notation, the posterior distribution (4.4) can be written as

$$p(\theta_{n+1}, \alpha, \beta, \xi \mid D^n) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}} + m\right)\right\} \exp\left\{((n+1)k + l)\xi + k\theta_{n+1}\right\}}{\left(e^{\alpha} + v\right)^{h+u} \left(e^{\beta} + s\right)^{g+r} \left(1 + e^{\beta} + e^{\xi}\right)^{T_{11}} \left(1 + e^{\alpha+\beta} + e^{\xi}\right)^{T_{12}}} \times$$

$$\times \frac{\exp\left\{\alpha(h + S_1 + S_2) + \beta\left(g + \sum_{i=1}^{n} w_{i2}\right)\right\}}{\left(e^{\alpha} + e^{\beta} + e^{\xi}\right)^{T_{21}} \left(e^{\alpha} + e^{\alpha+\beta} + e^{\xi}\right)^{T_{22}}}. \tag{4.5}$$

The predictive distributions of interest are then given by

$$p(z_1, z_2 \mid D^n) \propto \frac{\Gamma(k + z_1 + z_2)}{z_1! z_2!} \int_{\Re^3} \frac{\exp\left\{-m e^{\xi}\right\} \exp\left\{((n+1)k + l)\xi\right\}}{\left(e^{\alpha} + v\right)^{h+u} \left(e^{\beta} + s\right)^{g+r} \left(1 + e^{\beta} + e^{\xi}\right)^{T_{11}} \left(1 + e^{\alpha+\beta} + e^{\xi}\right)^{T_{12}}} \times$$

$$\times \frac{\exp\left\{\alpha(h + S_1 + S_2 + z_2) + \beta\left(g + \sum_{i=1}^{n} w_{i2}\right)\right\} d\alpha\, d\beta\, d\xi}{\left(e^{\alpha} + e^{\beta} + e^{\xi}\right)^{T_{21}} \left(e^{\alpha} + e^{\alpha+\beta} + e^{\xi}\right)^{T_{22}} \left(1 + e^{\alpha} + e^{\xi}\right)^{k + z_1 + z_2}},$$

$$p\left(z_1 \mid D^n\right) \propto \frac{\Gamma(k+z_1)}{z_1!} \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\} \exp\left\{((n+1)k+l)\xi\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}\left(1+e^{\beta}+e^{\xi}\right)^{T_{11}}\left(1+e^{\alpha+\beta}+e^{\xi}\right)^{T_{12}}} \times$$

$$\times \frac{\exp\left\{\alpha\left(h+S_1+S_2\right)+\beta\left(g+\sum_{i=1}^{n} w_{i2}\right)\right\} d\alpha\, d\beta\, d\xi}{\left(e^{\alpha}+e^{\beta}+e^{\xi}\right)^{T_{21}}\left(e^{\alpha}+e^{\alpha+\beta}+e^{\xi}\right)^{T_{22}}\left(1+e^{\xi}\right)^{k+z_1}},$$

$$p\left(z_2 \mid D^n\right) \propto \frac{\Gamma(k+z_2)}{z_2!} \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\} \exp\left\{((n+1)k+l)\xi\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}\left(1+e^{\beta}+e^{\xi}\right)^{T_{11}}\left(1+e^{\alpha+\beta}+e^{\xi}\right)^{T_{12}}} \times$$

$$\times \frac{\exp\left\{\alpha\left(h+S_1+S_2+z_2\right)+\beta\left(g+\sum_{i=1}^{n} w_{i2}\right)\right\} d\alpha\, d\beta\, d\xi}{\left(e^{\alpha}+e^{\beta}+e^{\xi}\right)^{T_{21}}\left(e^{\alpha}+e^{\alpha+\beta}+e^{\xi}\right)^{T_{22}}\left(e^{\alpha}+e^{\xi}\right)^{k+z_2}}.$$

These predictive distributions cannot be obtained explicitly. A numerical technique to solve three-dimensional integrals is required in order to evaluate the predictive probabilities. No real simplification follows for the case of vague second stage priors ($l, m, u, v, r, s \rightarrow 0$).

4.2.2. Estimation Via Gibbs Sampling

From the posterior distribution (4.5), the full conditional distributions are derived, namely

$$p\left(\theta_{n+1} \mid \alpha, \beta, \xi, D^n\right) \propto \exp\left\{-e^{\xi+\theta_{n+1}}\right\} \exp\left\{k\,\theta_{n+1}\right\},$$

$$p\left(\alpha \mid \theta_{n+1}, \beta, \xi, D^n\right) \propto \frac{\exp\left\{\alpha\left(h+S_1+S_2\right)\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(1+e^{\alpha+\beta}+e^{\xi}\right)^{T_{12}}\left(e^{\alpha}+e^{\beta}+e^{\xi}\right)^{T_{21}}\left(e^{\alpha}+e^{\alpha+\beta}+e^{\xi}\right)^{T_{22}}},$$

$$p\left(\beta \mid \theta_{n+1}, \alpha, \xi, D^n\right) \propto \frac{\exp\left\{\beta\left(g+\sum_{i=1}^{n} w_{i2}\right)\right\}}{\left(e^{\beta}+s\right)^{g+r}\left(1+e^{\beta}+e^{\xi}\right)^{T_{11}}\left(1+e^{\alpha+\beta}+e^{\xi}\right)^{T_{12}}\left(e^{\alpha}+e^{\beta}+e^{\xi}\right)^{T_{21}}} \times$$

$$\times \frac{1}{\left(e^{\alpha}+e^{\alpha+\beta}+e^{\xi}\right)^{T_{22}}},$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, \beta, D^{n}\right) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}}+m\right)\right\}\exp\left\{\left((n+1)k+l\right)\xi\right\}}{\left(1+e^{\beta}+e^{\xi}\right)^{T_{11}}\left(1+e^{\alpha+\beta}+e^{\xi}\right)^{T_{12}}\left(e^{\alpha}+e^{\beta}+e^{\xi}\right)^{T_{21}}} \times$$

$$\times \frac{1}{\left(e^{\alpha}+e^{\alpha+\beta}+e^{\xi}\right)^{T_{22}}}.$$

For the vague second stage priors case, the simplifications are again minimal.

Generation of values of θ_{n+1} again follows from a transformed gamma distribution (see Table A1.1); to generate values of α, β and ξ, we will have to use rejection sampling (section 1.2.3).

4.2.3. Estimation Via Asymptotic Results

The approximation presented by Bernardo & Smith (1994) cannot be used here since it is not possible to derive a maximum likelihood estimate for θ_{n+1}.

The second asymptotic approximation, presented by O'Hagan (1994), is based on characteristics of the posterior distribution itself, namely its mode and its modal dispersion matrix. Given the posterior distribution (4.5) for $\left(\theta_{n+1}, \alpha, \beta, \xi\right)$, and assuming that n is large enough, we can approximate it by

$$p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^{n}\right) \cong N_{4}(m, V),$$

with m being the posterior mode and V, defined by (1.5), being the modal dispersion matrix.

From (4.5) the posterior mode $m=\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$ is obtained from the equations

$$e^{\xi+\theta_{n+1}}-k=0,$$

$$h + S_1 + S_2 - \frac{(h+u)e^{\alpha}}{e^{\alpha} + v} - \frac{T_{12}e^{\alpha+\beta}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\alpha}}{e^{\alpha} + e^{\beta} + e^{\xi}} - \frac{T_{22}\left(e^{\alpha} + e^{\alpha+\beta}\right)}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0,$$

$$g + \sum_{i=1}^{n} w_{i2} - \frac{(g+r)e^{\beta}}{e^{\beta} + s} - \frac{T_{11}e^{\beta}}{1 + e^{\beta} + e^{\xi}} - \frac{T_{12}e^{\alpha+\beta}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\beta}}{e^{\alpha} + e^{\beta} + e^{\xi}} - \frac{T_{22}e^{\alpha+\beta}}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0, \quad (4.6)$$

$$(n+1)k + l - e^{\xi}\left(e^{\theta_{n+1}} + m\right) - \frac{T_{11}e^{\xi}}{1 + e^{\beta} + e^{\xi}} - \frac{T_{12}e^{\xi}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\xi}}{e^{\alpha} + e^{\beta} + e^{\xi}} - \frac{T_{22}e^{\xi}}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0.$$

The modal precision matrix V^{-1} takes the form

$$V^{-1} = \begin{pmatrix} \tilde{a} & 0 & 0 & \tilde{a} \\ 0 & \tilde{b} & \tilde{c} & \tilde{d} \\ 0 & \tilde{c} & \tilde{f} & \tilde{q} \\ \tilde{a} & \tilde{d} & \tilde{q} & \tilde{t} \end{pmatrix}$$

where

$$\tilde{a} = e^{\tilde{\xi}+\tilde{\theta}_{n+1}},$$

$$\tilde{b} = \frac{(h+u)ve^{\tilde{\alpha}}}{\left(e^{\tilde{\alpha}} + v\right)^2} + \frac{T_{12}e^{\tilde{\alpha}+\tilde{\beta}}\left(1 + e^{\tilde{\xi}}\right)}{\left(1 + e^{\tilde{\alpha}+\tilde{\beta}} + e^{\tilde{\xi}}\right)^2} + \frac{T_{21}e^{\tilde{\alpha}}\left(e^{\tilde{\beta}} + e^{\tilde{\xi}}\right)}{\left(e^{\tilde{\alpha}} + e^{\tilde{\beta}} + e^{\tilde{\xi}}\right)^2} + \frac{T_{22}\left(e^{\tilde{\alpha}} + e^{\tilde{\alpha}+\tilde{\beta}}\right)e^{\tilde{\xi}}}{\left(e^{\tilde{\alpha}} + e^{\tilde{\alpha}+\tilde{\beta}} + e^{\tilde{\xi}}\right)^2},$$

$$\tilde{c} = \frac{T_{12}e^{\tilde{\alpha}+\tilde{\beta}}\left(1 + e^{\tilde{\xi}}\right)}{\left(1 + e^{\tilde{\alpha}+\tilde{\beta}} + e^{\tilde{\xi}}\right)^2} - \frac{T_{21}e^{\tilde{\alpha}+\tilde{\beta}}}{\left(e^{\tilde{\alpha}} + e^{\tilde{\beta}} + e^{\tilde{\xi}}\right)^2} + \frac{T_{22}e^{\tilde{\alpha}+\tilde{\beta}+\tilde{\xi}}}{\left(e^{\tilde{\alpha}} + e^{\tilde{\alpha}+\tilde{\beta}} + e^{\tilde{\xi}}\right)^2},$$

$$\tilde{d} = -\frac{T_{12}e^{\tilde{\alpha}+\tilde{\beta}+\tilde{\xi}}}{\left(1 + e^{\tilde{\alpha}+\tilde{\beta}} + e^{\tilde{\xi}}\right)^2} - \frac{T_{21}e^{\tilde{\alpha}+\tilde{\xi}}}{\left(e^{\tilde{\alpha}} + e^{\tilde{\beta}} + e^{\tilde{\xi}}\right)^2} - \frac{T_{22}\left(e^{\tilde{\alpha}} + e^{\tilde{\alpha}+\tilde{\beta}}\right)e^{\tilde{\xi}}}{\left(e^{\tilde{\alpha}} + e^{\tilde{\alpha}+\tilde{\beta}} + e^{\tilde{\xi}}\right)^2},$$

$$\tilde{f} = \frac{(g+r)se^{\tilde{\beta}}}{\left(e^{\tilde{\beta}}+s\right)^2} + \frac{T_{11}e^{\tilde{\beta}}\left(1+e^{\tilde{\xi}}\right)}{\left(1+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} + \frac{T_{12}e^{\tilde{\alpha}+\tilde{\beta}}\left(1+e^{\tilde{\xi}}\right)}{\left(1+e^{\tilde{\alpha}+\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} + \frac{T_{21}e^{\tilde{\beta}}\left(e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)}{\left(e^{\tilde{\alpha}}+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} + \frac{T_{22}e^{\tilde{\alpha}+\tilde{\beta}}\left(e^{\tilde{\alpha}}+e^{\tilde{\xi}}\right)}{\left(e^{\tilde{\alpha}}+e^{\tilde{\alpha}+\tilde{\beta}}+e^{\tilde{\xi}}\right)^2},$$

$$\tilde{q} = -\frac{T_{11}e^{\tilde{\beta}+\tilde{\xi}}}{\left(1+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} - \frac{T_{12}e^{\tilde{\alpha}+\tilde{\beta}+\tilde{\xi}}}{\left(1+e^{\tilde{\alpha}+\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} - \frac{T_{21}e^{\tilde{\beta}+\tilde{\xi}}}{\left(e^{\tilde{\alpha}}+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} - \frac{T_{22}e^{\tilde{\alpha}+\tilde{\beta}+\tilde{\xi}}}{\left(e^{\tilde{\alpha}}+e^{\tilde{\alpha}+\tilde{\beta}}+e^{\tilde{\xi}}\right)^2},$$

$$\tilde{t} = e^{\tilde{\xi}}\left(e^{\tilde{\theta}_{n+1}}+m\right) + \frac{T_{11}e^{\tilde{\xi}}\left(1+e^{\tilde{\beta}}\right)}{\left(1+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} + \frac{T_{12}e^{\tilde{\xi}}\left(1+e^{\tilde{\alpha}+\tilde{\beta}}\right)}{\left(1+e^{\tilde{\alpha}+\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} + \frac{T_{21}e^{\tilde{\xi}}\left(e^{\tilde{\alpha}}+e^{\tilde{\beta}}\right)}{\left(e^{\tilde{\alpha}}+e^{\tilde{\beta}}+e^{\tilde{\xi}}\right)^2} + \frac{T_{22}e^{\tilde{\xi}}\left(e^{\tilde{\alpha}}+e^{\tilde{\alpha}+\tilde{\beta}}\right)}{\left(e^{\tilde{\alpha}}+e^{\tilde{\alpha}+\tilde{\beta}}+e^{\tilde{\xi}}\right)^2}.$$

Hence, the approximated posterior distribution for $(\theta_{n+1}, \alpha, \beta, \xi)$ will be

$$p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n\right) \cong N_4(m, V), \tag{4.7}$$

and the estimation of the predictive distributions of (Z_1, Z_2), Z_1 and Z_2 is done through the Gibbs sampling routine (section 1.2.1), generating random values from the following full conditional distributions,

$$p\left(\theta_{n+1} \mid \alpha, \beta, \xi, D^n\right) = N\left(\tilde{\theta}_{n+1} - \left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{a}}\right),$$

$$p\left(\alpha \mid \theta_{n+1}, \beta, \xi, D^n\right) = N\left(\tilde{\alpha} - \frac{\left(\beta-\tilde{\beta}\right)\tilde{c} + \left(\xi-\tilde{\xi}\right)\tilde{d}}{\tilde{b}}, \frac{1}{\tilde{b}}\right),$$

$$p\left(\beta \mid \theta_{n+1}, \alpha, \xi, D^n\right) = N\left(\tilde{\beta} - \frac{\left(\alpha-\tilde{\alpha}\right)\tilde{c} + \left(\xi-\tilde{\xi}\right)\tilde{q}}{\tilde{f}}, \frac{1}{\tilde{f}}\right),$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, \beta, D^n\right) = N\left(\tilde{\xi} - \frac{\left(\theta_{n+1} - \tilde{\theta}_{n+1}\right)\tilde{a} + \left(\alpha - \tilde{\alpha}\right)\tilde{d} + \left(\beta - \tilde{\beta}\right)\tilde{q}}{\tilde{t}}, \frac{1}{\tilde{t}}\right),$$

which are easily derived from (4.7). Again, no significant simplification occurs with second stage vague priors.

4.2.4. Laplace Approximation

Given the posterior distribution (4.5) for $(\theta_{n+1}, \alpha, \beta, \xi)$, the predictive distributions of (Z_1, Z_2), Z_1 and Z_2 are

$$p\left(z_1, z_2 \mid D^n\right) \propto \frac{1}{z_1!\, z_2!} \int_{\Re^4} \exp\left\{-e^{\theta_{n+1}} - e^{\alpha+\theta_{n+1}}\right\} \exp\left\{\theta_{n+1} z_1 + \left(\alpha + \theta_{n+1}\right) z_2\right\} \times$$

$$\times\, p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n\right) d\theta_{n+1} d\alpha\, d\beta\, d\xi,$$

$$p\left(z_1 \mid D^n\right) \propto \frac{1}{z_1!} \int_{\Re^4} \exp\left\{-e^{\theta_{n+1}}\right\} \exp\left\{\theta_{n+1} z_1\right\} p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n\right) d\theta_{n+1} d\alpha\, d\beta d\xi,$$

$$p\left(z_2 \mid D^n\right) \propto \frac{1}{z_2!} \int_{\Re^4} \exp\left\{-e^{\alpha+\theta_{n+1}}\right\} \exp\left\{\left(\alpha + \theta_{n+1}\right) z_2\right\} \times$$

$$\times\, p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n\right) d\theta_{n+1} d\alpha\, d\beta\, d\xi.$$

The Laplace approximation (section 1.2.4) can be used to obtain approximations of these distributions.

Let us consider the function $h\left(\theta_{n+1}, \alpha, \beta, \xi\right)$ defined by (1.13), that is, such that

$$-n h\left(\theta_{n+1}, \alpha, \beta, \xi\right) = \ln p\left(\theta_{n+1}, \alpha, \beta, \xi\right) + \ln p\left(D^n \mid \theta_{n+1}, \alpha, \beta, \xi\right)$$

$$\propto \ln p\left(\theta_{n+1}, \alpha, \beta, \xi \mid D^n\right)$$

and let us also define $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$ and $\tilde{\sigma}$ as in (1.16). Analysing the definitions for $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$ and $\tilde{\sigma}$, we conclude that $\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)=m$, the posterior mode, solution of the system of equations (4.6), and that the matrix $n\nabla^2 h\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)$ is equal to the modal precision matrix V^{-1} defined in the former section. Thus,

$$\tilde{\sigma} = \left\{\tilde{a}\tilde{b}\tilde{f}\tilde{t} - \tilde{a}\tilde{b}\tilde{q}^2 - \tilde{a}\tilde{t}\tilde{c}^2 + 2\tilde{a}\tilde{c}\tilde{d}\tilde{q} - \tilde{a}\tilde{f}\tilde{d}^2 - \tilde{b}\tilde{f}\tilde{a}^2 + \tilde{a}^2\tilde{c}^2\right\}^{-1/2}.$$

4.2.4.1. Joint Predictive Distribution of (Z_1, Z_2)

We have that

$$\begin{aligned}-n h_c^*(\theta_{n+1}, \alpha, \beta, \xi) = &-e^{\theta_{n+1}} - e^{\alpha+\theta_{n+1}} + \theta_{n+1} z_1 + (\alpha + \theta_{n+1}) z_2 - e^{\xi}\left(e^{\theta_{n+1}} + m\right) + \\ &+ ((n+1)k + l)\xi + k\theta_{n+1} + \alpha(h + S_1 + S_2) + \beta\left(g + \sum_{i=1}^{n} w_{i2}\right) - \\ &- (h+u)\ln\left(e^{\alpha} + v\right) - (g + r)\ln\left(e^{\beta} + s\right) - T_{11}\ln\left(1 + e^{\beta} + e^{\xi}\right) - \\ &- T_{12}\ln\left(1 + e^{\alpha+\beta} + e^{\xi}\right) - T_{21}\ln\left(e^{\alpha} + e^{\beta} + e^{\xi}\right) - \\ &- T_{22}\ln\left(e^{\alpha} + e^{\alpha+\beta} + e^{\xi}\right),\end{aligned}$$

according to definition (1.14). To derive $\left(\theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c, \xi^*_c\right)$, defined in (1.16), we have to solve, using a numerical technique, the system formed by the equations

$$e^{\theta_{n+1}}\left(1 + e^{\alpha} + e^{\xi}\right) - z_1 - z_2 - k = 0,$$

$$\begin{aligned}h + S_1 + S_2 + z_2 - e^{\alpha+\theta_{n+1}} - \frac{(h+u)e^{\alpha}}{e^{\alpha} + v} - \frac{T_{12}e^{\alpha+\beta}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\alpha}}{e^{\alpha} + e^{\beta} + e^{\xi}} - \\ - \frac{T_{22}\left(e^{\alpha} + e^{\alpha+\beta}\right)}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0,\end{aligned}$$

$$g+\sum_{i=1}^{n} w_{i2}-\frac{(g+r)e^{\beta}}{e^{\beta}+s}-\frac{T_{11}e^{\beta}}{1+e^{\beta}+e^{\xi}}-\frac{T_{12}e^{\alpha+\beta}}{1+e^{\alpha+\beta}+e^{\xi}}-\frac{T_{21}e^{\beta}}{e^{\alpha}+e^{\beta}+e^{\xi}}-$$

$$-\frac{T_{22}e^{\alpha+\beta}}{e^{\alpha}+e^{\alpha+\beta}+e^{\xi}}=0,$$

$$(n+1)k+l-e^{\xi}\left(e^{\theta_{n+1}}+m\right)-\frac{T_{11}e^{\xi}}{1+e^{\beta}+e^{\xi}}-\frac{T_{12}e^{\xi}}{1+e^{\alpha+\beta}+e^{\xi}}-\frac{T_{21}e^{\xi}}{e^{\alpha}+e^{\beta}+e^{\xi}}-$$

$$-\frac{T_{22}e^{\xi}}{e^{\alpha}+e^{\alpha+\beta}+e^{\xi}}=0.$$

If we define the constants

$$b_c^*=e^{\theta_{n+1(c)}^*+\alpha_c^*},\qquad c_c^*=e^{\theta_{n+1(c)}^*+\xi_c^*},\qquad a_c^*=e^{\theta_{n+1(c)}^*}+b_c^*+c_c^*,$$

$$d_c^*=e^{\alpha_c^*+\theta_{n+1(c)}^*}+\frac{(h+u)ve^{\alpha_c^*}}{\left(e^{\alpha_c^*}+v\right)^2}+\frac{T_{12}e^{\alpha_c^*+\beta_c^*}\left(1+e^{\xi_c^*}\right)}{\left(1+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2}+\frac{T_{21}e^{\alpha_c^*}\left(e^{\beta_c^*}+e^{\xi_c^*}\right)}{\left(e^{\alpha_c^*}+e^{\beta_c^*}+e^{\xi_c^*}\right)^2}+$$

$$+\frac{T_{22}\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}\right)e^{\xi_c^*}}{\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2},$$

$$f_c^*=\frac{T_{12}e^{\alpha_c^*+\beta_c^*}\left(1+e^{\xi_c^*}\right)}{\left(1+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2}-\frac{T_{21}e^{\alpha_c^*+\beta_c^*}}{\left(e^{\alpha_c^*}+e^{\beta_c^*}+e^{\xi_c^*}\right)^2}+\frac{T_{22}e^{\alpha_c^*+\beta_c^*+\xi_c^*}}{\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2},$$

$$q_c^*=-\frac{T_{12}e^{\alpha_c^*+\beta_c^*+\xi_c^*}}{\left(1+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2}-\frac{T_{21}e^{\alpha_c^*+\xi_c^*}}{\left(e^{\alpha_c^*}+e^{\beta_c^*}+e^{\xi_c^*}\right)^2}-\frac{T_{22}\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}\right)e^{\xi_c^*}}{\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2},$$

$$t_c^*=\frac{(g+r)se^{\beta_c^*}}{\left(e^{\beta_c^*}+s\right)^2}+\frac{T_{11}e^{\beta_c^*}\left(1+e^{\xi_c^*}\right)}{\left(1+e^{\beta_c^*}+e^{\xi_c^*}\right)^2}+\frac{T_{12}e^{\alpha_c^*+\beta_c^*}\left(1+e^{\xi_c^*}\right)}{\left(1+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2}+$$

$$+\frac{T_{21}e^{\beta_c^*}\left(e^{\alpha_c^*}+e^{\xi_c^*}\right)}{\left(e^{\alpha_c^*}+e^{\beta_c^*}+e^{\xi_c^*}\right)^2}+\frac{T_{22}e^{\alpha_c^*+\beta_c^*}\left(e^{\alpha_c^*}+e^{\xi_c^*}\right)}{\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2},$$

$$o_c^* = -\frac{T_{11}e^{\beta_c^*+\xi_c^*}}{\left(1+e^{\beta_c^*}+e^{\xi_c^*}\right)^2} - \frac{T_{12}e^{\alpha_c^*+\beta_c^*+\xi_c^*}}{\left(1+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2} - \frac{T_{21}e^{\beta_c^*+\xi_c^*}}{\left(e^{\alpha_c^*}+e^{\beta_c^*}+e^{\xi_c^*}\right)^2} -$$

$$-\frac{T_{22}e^{\alpha_c^*+\beta_c^*+\xi_c^*}}{\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2},$$

$$p_c^* = e^{\xi_c^*}\left(e^{\theta_{n+1(c)}^*}+m\right) + \frac{T_{11}e^{\xi_c^*}\left(1+e^{\beta_c^*}\right)}{\left(1+e^{\beta_c^*}+e^{\xi_c^*}\right)^2} + \frac{T_{12}e^{\xi_c^*}\left(1+e^{\alpha_c^*+\beta_c^*}\right)}{\left(1+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2}$$

$$+\frac{T_{21}e^{\xi_c^*}\left(e^{\alpha_c^*}+e^{\beta_c^*}\right)}{\left(e^{\alpha_c^*}+e^{\beta_c^*}+e^{\xi_c^*}\right)^2} + \frac{T_{22}e^{\xi_c^*}\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}\right)}{\left(e^{\alpha_c^*}+e^{\alpha_c^*+\beta_c^*}+e^{\xi_c^*}\right)^2},$$

we can write

$$n\nabla^2 h_c^*\left(\theta_{n+1(c)}^*, \alpha_c^*, \beta_c^*, \xi_c^*\right) = \begin{pmatrix} a_c^* & b_c^* & 0 & c_c^* \\ b_c^* & d_c^* & f_c^* & q_c^* \\ 0 & f_c^* & t_c^* & o_c^* \\ c_c^* & q_c^* & o_c^* & p_c^* \end{pmatrix},$$

which leads to

$$\sigma_c^* = \left\{a_c^* d_c^* t_c^* p_c^* - a_c^* d_c^* o_c^{*2} - a_c^* p_c^* f_c^{*2} + 2a_c^* f_c^* q_c^* o_c^* - a_c^* t_c^* q_c^{*2} - t_c^* p_c^* b_c^{*2} +\right.$$

$$\left. + b_c^{*2} o_c^2 - 2b_c^* f_c^* c_c^* o_c^* + 2b_c^* q_c^* c_c^* t_c^* - d_c^* t_c^* c_c^{*2} + f_c^{*2} c_c^{*2}\right\}^{-1/2},$$

and the Laplace approximation for the joint predictive distribution will be given by

$$p\left(z_1, z_2 \mid D^n\right) \,\tilde{\propto}\, \frac{1}{z_1! z_2!}\left(\frac{\sigma_c^*}{\tilde{\sigma}}\right) \exp\left\{-nh_c^*\left(\theta_{n+1(c)}^*, \alpha_c^*, \beta_c^*, \xi_c^*\right) + nh\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\right\},$$

following (1.15).

4.2.4.2. Marginal Predictive Distribution of Z_1

To evaluate the predictive distribution of Z_1, we take

$$-n h_1^*(\theta_{n+1}, \alpha, \beta, \xi) = -e^{\theta_{n+1}} + \theta_{n+1} z_1 - e^{\xi}\left(e^{\theta_{n+1}} + m\right) + ((n+1)k + l)\xi + k\theta_{n+1} +$$

$$+\alpha(h + S_1 + S_2) + \beta\left(g + \sum_{i=1}^{n} w_{i2}\right) - (h+u)\ln\left(e^{\alpha} + v\right) -$$

$$-(g+r)\ln\left(e^{\beta} + s\right) - T_{11}\ln\left(1 + e^{\beta} + e^{\xi}\right) - T_{12}\ln\left(1 + e^{\alpha+\beta} + e^{\xi}\right) -$$

$$-T_{21}\ln\left(e^{\alpha} + e^{\beta} + e^{\xi}\right) - T_{22}\ln\left(e^{\alpha} + e^{\alpha+\beta} + e^{\xi}\right),$$

so that $\left(\theta^*_{n+1(1)}, \alpha_1^*, \beta_1^*, \xi_1^*\right)$ is obtained from

$$e^{\theta_{n+1}}\left(1 + e^{\xi}\right) - z_1 - k = 0,$$

$$h + S_1 + S_2 - \frac{(h+u)e^{\alpha}}{e^{\alpha} + v} - \frac{T_{12}e^{\alpha+\beta}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\alpha}}{e^{\alpha} + e^{\beta} + e^{\xi}} - \frac{T_{22}\left(e^{\alpha} + e^{\alpha+\beta}\right)}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0,$$

$$g + \sum_{i=1}^{n} w_{i2} - \frac{(g+r)e^{\beta}}{e^{\beta} + s} - \frac{T_{11}e^{\beta}}{1 + e^{\beta} + e^{\xi}} - \frac{T_{12}e^{\alpha+\beta}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\beta}}{e^{\alpha} + e^{\beta} + e^{\xi}} -$$

$$-\frac{T_{22}e^{\alpha+\beta}}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0,$$

$$(n+1)k + l - e^{\xi}\left(e^{\theta_{n+1}} + m\right) - \frac{T_{11}e^{\xi}}{1 + e^{\beta} + e^{\xi}} - \frac{T_{12}e^{\xi}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\xi}}{e^{\alpha} + e^{\beta} + e^{\xi}} -$$

$$-\frac{T_{22}e^{\xi}}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0.$$

The matrix $n\nabla^2 h_1^*\left(\theta^*_{n+1(1)}, \alpha_1^*, \beta_1^*, \xi_1^*\right)$ takes the form

$$n\nabla^2 h_1^*\left(\theta^*_{n+1(1)}, \alpha_1^*, \beta_1^*, \xi_1^*\right) = \begin{pmatrix} a_1^* & 0 & 0 & b_1^* \\ 0 & c_1^* & d_1^* & f_1^* \\ 0 & d_1^* & q_1^* & t_1^* \\ b_1^* & f_1^* & t_1^* & p_1^* \end{pmatrix}$$

where

$$a_1^* = e^{\theta_{n+1(1)}^*}\left(1+e^{\xi_1^*}\right), \qquad b_1^* = e^{\theta_{n+1(1)}^*+\xi_1^*},$$

$$c_1^* = \frac{(h+u)ve^{\alpha_1^*}}{\left(e^{\alpha_1^*}+v\right)^2} + \frac{T_{12}e^{\alpha_1^*+\beta_1^*}\left(1+e^{\xi_1^*}\right)}{\left(1+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2} + \frac{T_{21}e^{\alpha_1^*}\left(e^{\beta_1^*}+e^{\xi_1^*}\right)}{\left(e^{\alpha_1^*}+e^{\beta_1^*}+e^{\xi_1^{**}}\right)^2} + \frac{T_{22}\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}\right)e^{\xi_1^*}}{\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2},$$

$$d_1^* = \frac{T_{12}e^{\alpha_1^*+\beta_1^*}\left(1+e^{\xi_1^*}\right)}{\left(1+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2} - \frac{T_{21}e^{\alpha_1^*+\beta_1^*}}{\left(e^{\alpha_1^*}+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} + \frac{T_{22}e^{\alpha_1^*+\beta_1^*+\xi_1^*}}{\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2},$$

$$f_1^* = -\frac{T_{12}e^{\alpha_1^*+\beta_1^*+\xi_1^*}}{\left(1+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2} - \frac{T_{21}e^{\alpha_1^*+\xi_1^*}}{\left(e^{\alpha_1^*}+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} - \frac{T_{22}\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}\right)e^{\xi_1^*}}{\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2},$$

$$q_1^* = \frac{(g+r)se^{\beta_1^*}}{\left(e^{\beta_1^*}+s\right)^2} + \frac{T_{11}e^{\beta_1^*}\left(1+e^{\xi_1^*}\right)}{\left(1+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} + \frac{T_{12}e^{\alpha_1^*+\beta_1^*}\left(1+e^{\xi_1^*}\right)}{\left(1+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2} +$$

$$+ \frac{T_{21}e^{\beta_1^*}\left(e^{\alpha_1^*}+e^{\xi_1^*}\right)}{\left(e^{\alpha_1^*}+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} + \frac{T_{22}e^{\alpha_1^*+\beta_1^*}\left(e^{\alpha_1^*}+e^{\xi_1^*}\right)}{\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2},$$

$$t_1^* = -\frac{T_{11}e^{\beta_1^*+\xi_1^*}}{\left(1+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} - \frac{T_{12}e^{\alpha_1^*+\beta_1^*+\xi_1^*}}{\left(1+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2} - \frac{T_{21}e^{\beta_1^*+\xi_1^*}}{\left(e^{\alpha_1^*}+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} - \frac{T_{22}e^{\alpha_1^*+\beta_1^*+\xi_1^*}}{\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2},$$

$$p_1^* = e^{\xi_1^*}\left(e^{\theta_{n+1(1)}^*}+m\right) + \frac{T_{11}e^{\xi_1^*}\left(1+e^{\beta_1^*}\right)}{\left(1+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} + \frac{T_{12}e^{\xi_1^*}\left(1+e^{\alpha_1^*+\beta_1^*}\right)}{\left(1+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2} +$$

$$+ \frac{T_{21}e^{\xi_1^*}\left(e^{\alpha_1^*}+e^{\beta_1^*}\right)}{\left(e^{\alpha_1^*}+e^{\beta_1^*}+e^{\xi_1^*}\right)^2} + \frac{T_{22}e^{\xi_1^*}\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}\right)}{\left(e^{\alpha_1^*}+e^{\alpha_1^*+\beta_1^*}+e^{\xi_1^*}\right)^2},$$

and, therefore,

$$\sigma_1^* = \left\{a_1^*c_1^*q_1^*p_1^* - a_1^*c_1^*t_1^{*2} - a_1^*p_1^*d_1^{*2} + 2a_1^*d_1^*f_1^*t_1^* - a_1^*q_1^*f_1^{*2} -\right.$$

$$-c_1^* q_1^* b_1^{*2} + d_1^{*2} b_1^{*2}\Big\}^{1/2}.$$

Finally, the predictive distribution of Z_1 can be approximated by

$$p\left(z_1 \mid D^n\right) \propto \frac{1}{z_1!}\left(\frac{\sigma_1^*}{\tilde{\sigma}}\right) \exp\left\{-n h_1^*\left(\theta_{n+1(1)}^*, \alpha_1^*, \beta_1^*, \xi_1^*\right) + nh\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\right\}.$$

4.2.4.3. Marginal Predictive Distribution of Z_2

In a similar way, for Z_2, we take

$$\begin{aligned}-n h_2^*\left(\theta_{n+1}, \alpha, \beta, \xi\right) = &-e^{\alpha+\theta_{n+1}} + \left(\alpha + \theta_{n+1}\right)z_2 - e^{\xi}\left(e^{\theta_{n+1}} + m\right) + \left((n+1)k + l\right)\xi + \\ &+ k\theta_{n+1} + \alpha\left(h + S_1 + S_2\right) + \beta\left(g + \sum_{i=1}^{n} w_{i2}\right) - (h+u)\ln\left(e^{\alpha} + v\right) - \\ &-(g+r)\ln\left(e^{\beta} + s\right) - T_{11}\ln\left(1 + e^{\beta} + e^{\xi}\right) - T_{12}\ln\left(1 + e^{\alpha+\beta} + e^{\xi}\right) - \\ &-T_{21}\ln\left(e^{\alpha} + e^{\beta} + e^{\xi}\right) - T_{22}\ln\left(e^{\alpha} + e^{\alpha+\beta} + e^{\xi}\right).\end{aligned}$$

Following the definition in (1.16), $\left(\theta_{n+1(2)}^*, \alpha_2^*, \beta_2^*, \xi_2^*\right)$ will be the solution of the system formed by the equations

$$e^{\theta_{n+1}}\left(e^{\alpha} + e^{\xi}\right) - z_2 - k = 0,$$

$$h + S_1 + S_2 + z_2 - e^{\alpha+\theta_{n+1}} - \frac{(h+u)e^{\alpha}}{e^{\alpha} + v} - \frac{T_{12}e^{\alpha+\beta}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\alpha}}{e^{\alpha} + e^{\beta} + e^{\xi}} - \frac{T_{22}\left(e^{\alpha} + e^{\alpha+\beta}\right)}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0,$$

$$g + \sum_{i=1}^{n} w_{i2} - \frac{(g+r)e^{\beta}}{e^{\beta} + s} - \frac{T_{11}e^{\beta}}{1 + e^{\beta} + e^{\xi}} - \frac{T_{12}e^{\alpha+\beta}}{1 + e^{\alpha+\beta} + e^{\xi}} - \frac{T_{21}e^{\beta}}{e^{\alpha} + e^{\beta} + e^{\xi}} - \frac{T_{22}e^{\alpha+\beta}}{e^{\alpha} + e^{\alpha+\beta} + e^{\xi}} = 0,$$

$$(n+1)k+l-e^{\xi}\left(e^{\theta_{n+1}}+m\right)-\frac{T_{11}e^{\xi}}{1+e^{\beta}+e^{\xi}}-\frac{T_{12}e^{\xi}}{1+e^{\alpha+\beta}+e^{\xi}}-\frac{T_{21}e^{\xi}}{e^{\alpha}+e^{\beta}+e^{\xi}}-$$

$$-\frac{T_{22}e^{\xi}}{e^{\alpha}+e^{\alpha+\beta}+e^{\xi}}=0,$$

Now, let us define

$$b_2^* = e^{\theta^*_{n+1(2)}+\alpha_2^*}, \qquad c_2^* = e^{\theta^*_{n+1(2)}+\xi_2^*}, \qquad a_2^* = b_2^* + c_2^*,$$

$$d_2^* = e^{\alpha_2^*+\theta^*_{n+1(2)}}+\frac{(h+u)ve^{\alpha_2^*}}{\left(e^{\alpha_2^*}+v\right)^2}+\frac{T_{12}e^{\alpha_2^*+\beta_2^*}\left(1+e^{\xi_2^*}\right)}{\left(1+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2}+\frac{T_{21}e^{\alpha_2^*}\left(e^{\beta_2^*}+e^{\xi_2^*}\right)}{\left(e^{\alpha_2^*}+e^{\beta_2^*}+e^{\xi_2^*}\right)^2}+$$

$$+\frac{T_{22}\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}\right)e^{\xi_2^*}}{\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2},$$

$$f_2^* = \frac{T_{12}e^{\alpha_2^*+\beta_2^*}\left(1+e^{\xi_2^*}\right)}{\left(1+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2}-\frac{T_{21}e^{\alpha_2^*+\beta_2^*}}{\left(e^{\alpha_2^*}+e^{\beta_2^*}+e^{\xi_2^*}\right)^2}+\frac{T_{22}e^{\alpha_2^*+\beta_2^*+\xi_2^*}}{\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2},$$

$$q_2^* = -\frac{T_{12}e^{\alpha_2^*+\beta_2^*+\xi_2^*}}{\left(1+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2}-\frac{T_{21}e^{\alpha_2^*+\xi_2^*}}{\left(e^{\alpha_2^*}+e^{\beta_2^*}+e^{\xi_2^*}\right)^2}-\frac{T_{22}\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}\right)e^{\xi_2^*}}{\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2},$$

$$t_2^* = \frac{(g+r)se^{\beta_2^*}}{\left(e^{\beta_2^*}+s\right)^2}+\frac{T_{11}e^{\beta_2^*}\left(1+e^{\xi_2^*}\right)}{\left(1+e^{\beta_2^*}+e^{\xi_2^*}\right)^2}+\frac{T_{12}e^{\alpha_2^*+\beta_2^*}\left(1+e^{\xi_2^*}\right)}{\left(1+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2}+\frac{T_{21}e^{\beta_2^*}\left(e^{\alpha_2^*}+e^{\xi_2^*}\right)}{\left(e^{\alpha_2^*}+e^{\beta_2^*}+e^{\xi_2^*}\right)^2}+$$

$$+\frac{T_{22}e^{\alpha_2^*+\beta_2^*}\left(e^{\alpha_2^*}+e^{\xi_2^*}\right)}{\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2},$$

$$o_2^* = -\frac{T_{11}e^{\beta_2^*+\xi_2^*}}{\left(1+e^{\beta_2^*}+e^{\xi_2^*}\right)^2}-\frac{T_{12}e^{\alpha_2^*+\beta_2^*+\xi_2^*}}{\left(1+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2}-\frac{T_{21}e^{\beta_2^*+\xi_2^*}}{\left(e^{\alpha_2^*}+e^{\beta_2^*}+e^{\xi_2^*}\right)^2}-$$

$$-\frac{T_{22}e^{\alpha_2^*+\beta_2^*+\xi_2^*}}{\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2},$$

$$p_2^* = e^{\xi_2^*}\left(e^{\theta_{n+1(2)}^*} + m\right) + \frac{T_{11}e^{\xi_2^*}\left(1+e^{\beta_2^*}\right)}{\left(1+e^{\beta_2^*}+e^{\xi_2^*}\right)^2} + \frac{T_{12}e^{\xi_2^*}\left(1+e^{\alpha_2^*+\beta_2^*}\right)}{\left(1+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2} + \frac{T_{21}e^{\xi_2^*}\left(e^{\alpha_2^*}+e^{\beta_2^*}\right)}{\left(e^{\alpha_2^*}+e^{\beta_2^*}+e^{\xi_2^*}\right)^2} +$$

$$+ \frac{T_{22}e^{\xi_2^*}\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}\right)}{\left(e^{\alpha_2^*}+e^{\alpha_2^*+\beta_2^*}+e^{\xi_2^*}\right)^2},$$

so that we can write

$$n\nabla^2 h_2^*\left(\theta_{n+1(2)}^*, \alpha_2^*, \beta_2^*, \xi_2^*\right) = \begin{pmatrix} a_2^* & b_2^* & 0 & c_2^* \\ b_2^* & d_2^* & f_2^* & q_2^* \\ 0 & f_2^* & t_2^* & o_2^* \\ c_2^* & q_2^* & o_2^* & p_2^* \end{pmatrix},$$

which leads to

$$\sigma_2^* = \left\{a_2^* d_2^* t_2^* p_2^* - a_2^* d_2^* o_2^{*2} - a_2^* p_2^* f_2^{*2} + 2a_2^* f_2^* q_2^* o_2^* - a_2^* t_2^* q_2^{*2} - t_2^* p_2^* b_2^{*2} + \right.$$

$$\left. + b_2^{*2} o_2^{2} - 2b_2^* f_2^* c_2^* o_2^* + 2b_2^* q_2^* c_2^* t_2^* - d_2^* t_2^* c_2^{*2} + f_2^{*2} c_2^{*2}\right\}^{-1/2},$$

and the Laplace approximation for the predictive distribution of Z_2 will be given by

$$p\left(z_2 \mid D^n\right) \propto \frac{1}{z_2!}\left(\frac{\sigma_2^*}{\tilde{\sigma}}\right) \exp\left\{-n h_2^*\left(\theta_{n+1(2)}^*, \alpha_2^*, \beta_2^*, \xi_2^*\right) + nh\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\right\}.$$

If we consider second stage vague priors, we do not get relevant simplifications in the evaluation of the above Laplace approximations.

CHAPTER 5

A MODEL TO PREDICT THE NUMBER OF ACCIDENTS BASED ON TWO TRAFFIC FLOWS

We now consider models which can be used to predict the number of accidents occurring at a road junction in a given period of time, based on a measurement of the traffic flows at the junction as well as on covariates describing important features of the junction. Barnett & Wright (1990) and Dunsmore & Robson (1992) considered this problem within a classical framework. Here, we develop a Bayesian predictive approach.

In this chapter we consider a very simplified model in which the predictions are based only on measurements of just two traffic flows. In Chapter 6 we consider incorporation of the covariates and the extension to multi-flow situations.

With respect to the i-th road junction, let Y_i be a random variable representing the number of accidents which occur at that junction in a given period of time, and let X_{i1} and X_{i2} be two random variables representing the countings which are used to quantify, respectively, the two traffic flows at the junction. Typically these counts are taken over short periods of time, and so act as assessments for the real traffic flows. We suppose that X_{i1}, X_{i2} and Y_i $(i = 1, 2, \ldots, n+1)$ are independent random variables such that

$$X_{i1} \sim Po\left(\exp\left(a_{i1} + k_{i1}\right)\right),$$
$$X_{i2} \sim Po\left(\exp\left(a_{i2} + k_{i2}\right)\right),$$
$$Y_i \sim Po\left(\exp\left(\lambda_1 a_{i1} + \lambda_2 a_{i2}\right)\right).$$

Here, a_{i1} and a_{i2} $(i = 1, 2, \ldots, n+1)$ are unknown parameters which model the characteristics of the traffic flows, λ_1 and λ_2 are unknown parameters modelling, respectively, the effect of the first and the second flow upon the number of accidents and k_{i1} and k_{i2} $(i = 1, 2, \ldots, n+1)$ are known constants which relate to the length of the observational period and the time of the day and year that observations were made. Further discussions of the model and background are given in Barnett & Wright (1990, 1992).

After observing n road junctions we obtain the data set

$$D^n = \left\{ \left(x_{i1}, x_{i2}, y_i, k_{i1}, k_{i2}\right),\ i = 1, 2, \dots, n \right\}.$$

Based on D^n and on observed values $x_{n+1,1}$, $x_{n+1,2}$, $k_{n+1,1}$ and $k_{n+1,2}$, for a new junction, our aim is to derive the predictive distribution for Y_{n+1}, the number of accidents at the $(n+1)$-th junction, during the period of time considered.

5.1. A Classical Approach

In order to simplify future notation, let us define a parameter vector $\theta_{red} = (a_{11}, a_{21}, \dots, a_{n1}, a_{12}, a_{22}, \dots, a_{n2}, a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2)$. We also define a vector d_{n+1} whose components are the observations for the new junction, that is, $d_{n+1} = (x_{n+1,1}, x_{n+1,2}, k_{n+1,1}, k_{n+1,2})$.

The likelihood function is

$$L\left(\theta_{red};\, D^n,\, d_{n+1}\right) \propto \exp\left\{ -\sum_{i=1}^{n+1}\sum_{j=1}^{2} \exp\left\{a_{ij} + k_{ij}\right\} - \sum_{i=1}^{n} \exp\left\{\lambda_1 a_{i1} + \lambda_2 a_{i2}\right\} \right\} \times$$

$$\times \exp\left\{ \sum_{i=1}^{n+1}\sum_{j=1}^{2} \left(a_{ij} + k_{ij}\right) x_{ij} + \sum_{i=1}^{n} \left(\lambda_1 a_{i1} + \lambda_2 a_{i2}\right) y_i \right\}.$$

The maximum likelihood estimates for $a_{n+1,1}$ and $a_{n+1,2}$ can be obtained explicitly, being

$$\begin{aligned} \hat{a}_{n+1,1} &= \ln\left(x_{n+1,1}\right) - k_{n+1,1}, \\ \hat{a}_{n+1,2} &= \ln\left(x_{n+1,2}\right) - k_{n+1,2}, \end{aligned} \qquad (5.1)$$

whilst the maximum likelihood estimates for the remaining parameters, $\hat{a}_{i1}$, $\hat{a}_{i2}$ $(i = 1, 2, \ldots, n)$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$, are obtained solving numerically the system formed by the $2(n+1)$ equations

$$\exp\{a_{ij} + k_{ij}\} + \lambda_j \exp\{\lambda_1 a_{i1} + \lambda_2 a_{i2}\} - x_{ij} - \lambda_j y_i = 0 \quad , i = 1, 2, \ldots, n \quad ; j = 1, 2,$$

$$\sum_{i=1}^{n} a_{ij}\left(y_i - \exp\{\lambda_1 a_{i1} + \lambda_2 a_{i2}\}\right) = 0 \quad , j = 1, 2. \tag{5.2}$$

Using these maximum likelihood estimates, a simple approximation for the predictive distribution of Y_{n+1} would be, in a classical framework, given by

$$Y_{n+1} \sim Po\left(\exp\left\{\hat{\lambda}_1 \hat{a}_{n+1,1} + \hat{\lambda}_2 \hat{a}_{n+1,2}\right\}\right).$$

From (5.1) we note that problems arise if $x_{n+1,1} = 0$ or $x_{n+1,2} = 0$.

5.2. A Bayesian Approach

In order to derive the predictive distribution of Y_{n+1} in a Bayesian framework, we consider a hierarchical prior structure as follows. At the first stage we take

$$p(\theta_{red} \mid \xi_1, \xi_2, \eta_1, \eta_2) = \prod_{i=1}^{n+1}\left\{p(a_{i1} \mid \xi_1)\, p(a_{i2} \mid \xi_2)\right\} \prod_{j=1}^{2}\left\{p(\lambda_j \mid \eta_j)\right\},$$

and at the second stage we take

$$p(\xi_1, \xi_2, \eta_1, \eta_2) = \prod_{j=1}^{2}\left\{p(\xi_j)\, p(\eta_j)\right\},$$

with

$$\exp(a_{ij} + k_{ij}) \sim Ga\left(b_j, e^{\xi_j}\right) \quad , i = 1, 2, \ldots, n+1 \quad ; j = 1, 2,$$

$$\exp(\lambda_j) \sim Ga\left(d_j, e^{\eta_j}\right) \quad , j = 1, 2,$$

$$\exp(\xi_j) \sim Gd(u_j, v_j) \quad , j = 1, 2,$$

$$\exp(\eta_j) \sim Gd(r_j, s_j) \quad , j = 1, 2.$$

The joint prior distribution for the parameter vector $\theta = (a_{11}, a_{21}, ..., a_{n1}, a_{12}, a_{22}, ..., a_{n2}, a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2, \xi_1, \xi_2, \eta_1, \eta_2)$ is

$$p(\theta \mid D^n, d_{n+1}) \propto \exp\left\{-\sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(1+e^{\xi_j}\right)\exp\left\{a_{ij}+k_{ij}\right\} - \sum_{i=1}^{n}\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}\right\}\times$$

$$\times\exp\left\{-\sum_{j=1}^{2} v_j e^{\xi_j} - \sum_{j=1}^{2} e^{\eta_j}\left(s_j + e^{\lambda_j}\right)\right\}\exp\left\{\sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(x_{ij}+b_j\right)\left(a_{ij}+k_{ij}\right)\right\}\times$$

$$\times\exp\left\{\sum_{j=1}^{2}\left((n+1)b_j + u_j\right)\xi_j + \sum_{j=1}^{2}\eta_j\left(d_j + r_j\right) + \sum_{j=1}^{2}\lambda_j\left(d_j + \sum_{i=1}^{n} a_{ij}y_i\right)\right\}. \tag{5.3}$$

5.2.1. The Exact Predictive Distribution

The exact predictive distribution of Y_{n+1} is given by

$$p(y_{n+1} \mid D^n, d_{n+1}) = \frac{1}{y_{n+1}!}\int_{\Re^4}\exp\left\{-\exp\left\{\lambda_1 a_{n+1,1} + \lambda_2 a_{n+1,2}\right\}\right\}\times$$

$$\times\exp\left\{\left(\lambda_1 a_{n+1,1} + \lambda_2 a_{n+1,2}\right)y_{n+1}\right\}p\left(a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2 \mid D^n, d_{n+1}\right)\times$$

$$\times\, da_{n+1,1}\, da_{n+1,2}\, d\lambda_1\, d\lambda_2.$$

Although we just need the posterior distribution of $(a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2)$, it is not possible to do all the necessary integrations of (5.3) to remove the remaining parameters. After eliminating the hyperparameters ξ_1, ξ_2, η_1 and η_2, we obtain the posterior distribution

$$p\left(\theta_{red} \mid D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\sum_{i=1}^{n+1}\sum_{j=1}^{2}\exp\left\{a_{ij}+k_{ij}\right\}-\sum_{i=1}^{n}\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}\right\}}{\prod_{j=1}^{2}\left\{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}\right\}} \times$$

$$\times \frac{\exp\left\{\sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(a_{ij}+k_{ij}\right)\left(x_{ij}+b_j\right)+\sum_{j=1}^{2}\lambda_j\left(d_j+\sum_{i=1}^{n}a_{ij}y_i\right)\right\}}{\prod_{j=1}^{2}\left\{\left(s_j+e^{\lambda_j}\right)^{d_j+r_j}\right\}}, \quad (5.4)$$

and the exact predictive distribution of Y_{n+1} will then be given by

$$p\left(y_{n+1} \mid D^n, d_{n+1}\right) \propto \frac{1}{y_{n+1}!}\int_{\Re^{2n+4}} \frac{\exp\left\{-\sum_{i=1}^{n+1}\sum_{j=1}^{2}\exp\left\{a_{ij}+k_{ij}\right\}-\sum_{i=1}^{n+1}\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}\right\}}{\prod_{j=1}^{2}\left\{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}\right\}} \times$$

$$\times \frac{\exp\left\{\sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(a_{ij}+k_{ij}\right)\left(x_{ij}+b_j\right)+\sum_{j=1}^{2}\lambda_j\left(d_j+\sum_{i=1}^{n+1}a_{ij}y_i\right)\right\}}{\prod_{j=1}^{2}\left\{\left(s_j+e^{\lambda_j}\right)^{d_j+r_j}\right\}} \, d\theta_{red}.$$

A numerical method is then required to evaluate these predictive probabilities. Due to the high dimensionality of the integral here involved, the numerical techniques are not reliable and we might expect problems. If we consider a vague second stage prior structure $\left(u_1, u_2, v_1, v_2, r_1, r_2, s_1, s_2 \rightarrow 0\right)$, the solution of the problem does not simplify significantly.

5.2.2. Estimation Via Gibbs Sampling

Since the exact predictive distribution of Y_{n+1}, derived in the previous section, involves a $(2n+4)$-dimensional integral, we consider estimates or approximations as in previous chapters. To implement the Gibbs sampling algorithm, we have to know the full conditional distributions, which are derived from (5.4), as being

$$p\left(a_{ij} \mid \theta_{red}^{(a_{ij})}, D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\exp\left\{a_{ij}+k_{ij}\right\}-\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}\right\}}{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}} \times$$

$$\times \exp\left\{a_{ij}\left(b_j+x_{ij}+\lambda_j y_i\right)\right\} \quad , i=1,2,\dots,n \quad ; j=1,2,$$

$$p\left(a_{n+1,j} \mid \theta_{red}^{(a_{n+1,j})}, D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\exp\left\{a_{n+1,j}+k_{n+1,j}\right\}\right\}\exp\left\{a_{n+1,j}\left(b_j+x_{n+1,j}\right)\right\}}{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}},$$

$$j=1,2, \quad (5.5)$$

$$p\left(\lambda_j \mid \theta_{red}^{(\lambda_j)}, D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\sum_{i=1}^{n}\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}\right\}\exp\left\{\lambda_j\left(d_j+\sum_{i=1}^{n}a_{ij}y_i\right)\right\}}{\left(s_j+e^{\lambda_j}\right)^{d_j+r_j}},$$

$$j=1,2,$$

where $\theta_{red}^{(\theta_i)}$ represents the vector θ_{red} with the component θ_i removed.

The assumption of vague second stage priors does not lead to simpler distributions.

Following the algorithm described in (1.1) with the above distributions, after t iterations we obtain a sample

$$\left(a_{11}^{(t)}, a_{21}^{(t)}, \dots, a_{n1}^{(t)}, a_{12}^{(t)}, a_{22}^{(t)}, \dots, a_{n2}^{(t)}, a_{n+1,1}^{(t)}, a_{n+1,2}^{(t)}, \lambda_1^{(t)}, \lambda_2^{(t)}\right)$$

and repeating the procedure M times we will have M samples

$$\left(a_{11(j)}^{(t)}, a_{21(j)}^{(t)}, \dots, a_{n1(j)}^{(t)}, a_{12(j)}^{(t)}, a_{22(j)}^{(t)}, \dots, a_{n2(j)}^{(t)}, a_{n+1,1(j)}^{(t)}, a_{n+1,2(j)}^{(t)}, \lambda_{1(j)}^{(t)}, \lambda_{2(j)}^{(t)}\right),$$

for $j=1, 2, \dots, M$. Note that although we have to generate, in each cycle, values for all the parameters, we are just interested in some of them, namely

$$\left(a_{n+1,1(j)}^{(t)}, a_{n+1,2(j)}^{(t)}, \lambda_{1(j)}^{(t)}, \lambda_{2(j)}^{(t)}\right) \quad , j=1,2\dots M.$$

Now, defining $\mu_j = \exp\left\{\lambda_{1(j)}^{(t)} a_{n+1,1(j)}^{(t)} + \lambda_{2(j)}^{(t)} a_{n+1,2(j)}^{(t)}\right\}$, for $j=1, 2, ..., M$, the predictive probabilities of Y_{n+1} are estimated, according to (1.2), by

$$p\left(y_{n+1} \mid D^n, d_{n+1}\right) = \frac{1}{M}\sum_{j=1}^{M} \frac{\exp\left\{-\mu_j\right\}\mu_j^{y_{n+1}}}{y_{n+1}!} .$$

Although we obtained very good results in Chapters 2 and 3 using the Gibbs sampling approach, it was then pointed out that the implementation of such a method could be difficult if the sampling from the full conditional distributions could not be done directly. In the present situation, sampling from the distributions in (5.5) must be done using a sampling technique such as the rejection sampling algorithm (section 1.2.3). Because of the forms of the full conditional distributions in (5.5), the adaptive rejection sampling algorithm should be considered. Furthermore, we noticed that when the number of parameters increases, we have to increase hugely the number of iterations in each cycle of the Gibbs routine in order to make it accurate. Since in the present problem we need to generate values for $2n+4$ parameters, this may be a drawback for the use of the Gibbs sampling method.

5.2.3. Estimation Via Asymptotic Results

If n is large enough, the posterior distributions (5.3) or (5.4) can be approximated by a multivariate normal distribution (section 1.2.2) and then the Gibbs routine can be easily used to estimate the predictive distribution of Y_{n+1}.

5.2.3.1. Posterior Normality Based on the Likelihood Function

Following the asymptotic result suggested by Bernardo & Smith (1994) and summarised in section 1.2.2.1, we conclude that the Gibbs routine can be applied to estimate the predictive probabilities of Y_{n+1}, using the conditional distributions

$$p\left(a_{n+1,j} \mid D^n, d_{n+1}\right) = N\left(\hat{a}_{n+1,j}, \frac{1}{H_j}\right) \quad , j = 1, 2,$$

$$p\left(\lambda_j \mid \lambda_{i\neq j}, D^n, d_{n+1}\right) = N\left(\hat{\lambda}_j - \frac{\left(\lambda_i - \hat{\lambda}_i\right)Q}{T_j}, \frac{1}{T_j}\right) \quad, i, j = 1, 2 \quad ; i \neq j,$$

where

$$H_j = \exp\left\{\hat{a}_{n+1,j} + k_{n+1,j}\right\} = x_{n+1,j} \quad , j = 1, 2,$$

$$T_1 = S_1 - \sum_{i=1}^{n} \frac{C_{i1}^2 A_{i2} + D_{i21}^2 A_{i1} - 2 B_i C_{i1} D_{i21}}{A_{i1} A_{i2} - B_i^2} \;,$$

$$T_2 = S_2 - \sum_{i=1}^{n} \frac{D_{i12}^2 A_{i2} + C_{i2}^2 A_{i1} - 2 B_i D_{i12} C_{i2}}{A_{i1} A_{i2} - B_i^2} \;,$$

$$Q = R - \sum_{i=1}^{n} \frac{C_{i1}\left(D_{i12} A_{i2} - B_i C_{i2}\right) + D_{i21}\left(A_{i1} C_{i2} - B_i D_{i12}\right)}{A_{i1} A_{i2} - B_i^2} \;,$$

with

$$A_{ij} = \exp\left\{\hat{a}_{ij} + k_{ij}\right\} + \hat{\lambda}_j^2 \exp\left\{\hat{\lambda}_1 \hat{a}_{i1} + \hat{\lambda}_j \hat{a}_{i2}\right\} \quad , i = 1, 2, \ldots, n \quad ; j = 1, 2,$$

$$B_i = \hat{\lambda}_1 \hat{\lambda}_2 \exp\left\{\hat{\lambda}_1 \hat{a}_{i1} + \hat{\lambda}_2 \hat{a}_{i2}\right\} \quad , i = 1, 2, \ldots, n,$$

$$C_{ij} = \left(1 + \hat{\lambda}_j \hat{a}_{ij}\right) \exp\left\{\hat{\lambda}_1 \hat{a}_{i1} + \hat{\lambda}_2 \hat{a}_{i2}\right\} - y_i \quad , i = 1, 2, \ldots, n \quad ; j = 1, 2,$$

$$D_{ijk} = \hat{\lambda}_j \hat{a}_{ik} \exp\left\{\hat{\lambda}_1 \hat{a}_{i1} + \hat{\lambda}_2 \hat{a}_{i2}\right\} \quad , i = 1, 2, \ldots, n \quad ; j, k = 1, 2 \quad ; j \neq k,$$

$$S_j = \sum_{i=1}^{n} \hat{a}_{ij}^2 \exp\left\{\hat{\lambda}_1 \hat{a}_{i1} + \hat{\lambda}_2 \hat{a}_{i2}\right\} \quad , j = 1, 2, \qquad R = \sum_{i=1}^{n} \hat{a}_{i1} \hat{a}_{i2} \exp\left\{\hat{\lambda}_1 \hat{a}_{i1} + \hat{\lambda}_2 \hat{a}_{i2}\right\},$$

and $\hat{a}_{ij}$ $(i = 1, 2, \ldots, n+1$ and $j = 1, 2)$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ being the maximum likelihood estimates given by (5.1) and (5.2).

Note that in contrast with the solution obtained in Chapters 2, 3 and 4, where the parameters were independent, now λ_1 and λ_2 depend on each other.

5.2.3.2. Posterior Normality Based on Characteristics of the Posterior Distribution

In section 1.2.2.2 we summarised another asymptotic normal approximation for the posterior distribution which is based on the posterior mode and on the modal dispersion matrix (O'Hagan (1994)). The solution of the present problem turns out to be easier and simpler to implement if we approximate the posterior distribution of θ instead of θ_{red} by a multivariate normal distribution. According to (1.6), the posterior distribution (5.3) is asymptotically approximated by

$$p\left(\theta \mid D^n, d_{n+1}\right) \cong N_{2n+8}(m, V),$$

where $m = \left(\tilde{a}_{11}, \tilde{a}_{21}, \ldots, \tilde{a}_{n1}, \tilde{a}_{12}, \tilde{a}_{22}, \ldots, \tilde{a}_{n2}, \tilde{a}_{n+1,1}, \tilde{a}_{n+1,2}, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\xi}_1, \tilde{\xi}_2, \tilde{\eta}_1, \tilde{\eta}_2\right)$ is the posterior mode and V, defined by (1.5) is the modal dispersion matrix.

Solving numerically the system formed by the $(2n+8)$ equations

$$\left(1+e^{\xi_j}\right)\exp\left\{a_{ij}+k_{ij}\right\}+\lambda_j\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}-b_j-x_{ij}-\lambda_j y_i=0, \quad i=1,2,\ldots,n\ ;\ j=1,2,$$

$$\left(1+e^{\xi_j}\right)\exp\left\{a_{n+1,j}+k_{n+1,j}\right\}-b_j-x_{n+1,j}=0, \quad j=1,2,$$

$$\sum_{i=1}^{n} a_{ij}y_i-\sum_{i=1}^{n} a_{ij}\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}+d_j-e^{\eta_j+\lambda_j}=0, \quad j=1,2, \tag{5.6}$$

$$e^{\xi_j}\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)-(n+1)b_j-u_j=0, \quad j=1,2,$$

$$e^{\eta_j}\left(s_j+e^{\lambda_j}\right)-d_j-r_j=0, \quad j=1,2,$$

we obtain m. Then, defining the constants

$$\tilde{A}_{ij}=\left(1+e^{\tilde{\xi}_j}\right)\exp\left\{\tilde{a}_{ij}+k_{ij}\right\}+\tilde{\lambda}_j^2\exp\left\{\tilde{\lambda}_1\tilde{a}_{i1}+\tilde{\lambda}_2\tilde{a}_{i2}\right\}, \quad i=1,2,\ldots,n\ ;\ j=1,2,$$

$$\tilde{B}_i=\tilde{\lambda}_1\tilde{\lambda}_2\exp\left\{\tilde{\lambda}_1\tilde{a}_{i1}+\tilde{\lambda}_2\tilde{a}_{i2}\right\}, \quad i=1,2,\ldots,n,$$

$$\tilde{C}_{ij} = \left(1 + \tilde{\lambda}_j \tilde{a}_{ij}\right) \exp\left\{\tilde{\lambda}_1 \tilde{a}_{i1} + \tilde{\lambda}_2 \tilde{a}_{i2}\right\} - y_i, \quad i = 1, 2, \ldots, n \ ; \ j = 1, 2,$$

$$\tilde{D}_{ijk} = \tilde{\lambda}_j \tilde{a}_{ik} \exp\left\{\tilde{\lambda}_1 \tilde{a}_{i1} + \tilde{\lambda}_2 \tilde{a}_{i2}\right\}, \quad i = 1, 2, \ldots, n \ ; \ j = 1, 2 \ ; \ k = 1, 2,$$

$$\tilde{E}_{ij} = \exp\left\{\tilde{\xi}_j + \tilde{a}_{ij} + k_{ij}\right\}, \quad i = 1, 2, \ldots, n+1 \ ; \ j = 1, 2,$$

$$\tilde{F}_j = \left(1 + e^{\tilde{\xi}_j}\right) \exp\left\{\tilde{a}_{n+1,j} + k_{n+1,j}\right\}, \quad j = 1, 2, \tag{5.7}$$

$$\tilde{G}_j = \sum_{i=1}^{n} \tilde{a}_{ij}^2 \exp\left\{\tilde{\lambda}_1 \tilde{a}_{i1} + \tilde{\lambda}_2 \tilde{a}_{i2}\right\} + e^{\tilde{\eta}_j + \tilde{\lambda}_j}, \quad j = 1, 2,$$

$$\tilde{H} = \sum_{i=1}^{n} \tilde{a}_{i1} \tilde{a}_{i2} \exp\left\{\tilde{\lambda}_1 \tilde{a}_{i1} + \tilde{\lambda}_2 \tilde{a}_{i2}\right\}, \qquad \tilde{L}_j = e^{\tilde{\eta}_j + \tilde{\lambda}_j}, \quad j = 1, 2,$$

$$\tilde{M}_j = e^{\tilde{\xi}_j}\left(v_j + \sum_{i=1}^{n+1} \exp\left\{\tilde{a}_{ij} + k_{ij}\right\}\right), \quad j = 1, 2, \qquad \tilde{R}_j = e^{\tilde{\eta}_j}\left(s_j + e^{\tilde{\lambda}_j}\right), \quad j = 1, 2,$$

we can write the inverse of the modal dispersion matrix in a partitioned form as

$$V^{-1} = \begin{pmatrix} \tilde{V}_{11} & \tilde{V}_{12} & 0_{n2} & \tilde{V}_{14} & \tilde{V}_{15} & 0_{n2} \\ \tilde{V}_{12} & \tilde{V}_{22} & 0_{n2} & \tilde{V}_{24} & \tilde{V}_{25} & 0_{n2} \\ 0_{2n} & 0_{2n} & \tilde{V}_{33} & 0_{22} & \tilde{V}_{35} & 0_{22} \\ \tilde{V}_{14}^T & \tilde{V}_{24}^T & 0_{22} & \tilde{V}_{44} & 0_{22} & \tilde{V}_{46} \\ \tilde{V}_{15}^T & \tilde{V}_{25}^T & \tilde{V}_{35} & 0_{22} & \tilde{V}_{55} & 0_{22} \\ 0_{2n} & 0_{2n} & 0_{22} & \tilde{V}_{46} & 0_{22} & \tilde{V}_{66} \end{pmatrix}, \tag{5.8}$$

where

$$\tilde{V}_{11} = diag\left(\tilde{A}_{i1}, \quad i = 1, 2, \ldots, n\right), \qquad \tilde{V}_{12} = diag\left(\tilde{B}_i, \quad i = 1, 2, \ldots, n\right),$$

$$\tilde{V}_{14}^T = \begin{pmatrix} \tilde{C}_{11} & \tilde{C}_{21} & \cdots & \tilde{C}_{n1} \\ \tilde{D}_{112} & \tilde{D}_{212} & \cdots & \tilde{D}_{n12} \end{pmatrix}, \qquad \tilde{V}_{15}^T = \begin{pmatrix} \tilde{E}_{11} & \tilde{E}_{21} & \cdots & \tilde{E}_{n1} \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$\tilde{V}_{22} = diag\left(\tilde{A}_{i2}, \quad i = 1, 2, \ldots, n\right), \qquad \tilde{V}_{24}^T = \begin{pmatrix} \tilde{D}_{121} & \tilde{D}_{221} & \cdots & \tilde{D}_{n21} \\ \tilde{C}_{12} & \tilde{C}_{22} & \cdots & \tilde{C}_{n2} \end{pmatrix},$$

$$\tilde{V}_{25}^T = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \tilde{E}_{12} & \tilde{E}_{22} & \cdots & \tilde{E}_{n2} \end{pmatrix}, \qquad \tilde{V}_{33} = diag\left(\tilde{F}_1, \tilde{F}_2\right),$$

$$\tilde{V}_{35} = diag\left(\tilde{E}_{n+1,1}, \tilde{E}_{n+1,2}\right), \qquad \tilde{V}_{44} = \begin{pmatrix} \tilde{G}_1 & \tilde{H} \\ \tilde{H} & \tilde{G}_2 \end{pmatrix},$$

$$\tilde{V}_{46} = diag\left(\tilde{L}_1, \tilde{L}_2\right), \qquad \tilde{V}_{55} = diag\left(\tilde{M}_1, \tilde{M}_2\right),$$

$$\tilde{V}_{66} = diag\left(\tilde{R}_1, \tilde{R}_2\right).$$

Note that the matrix V^{-1} in (5.8) is quite sparse and therefore it is simple to derive the approximate marginal posterior distribution for $\left(a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2, \xi_1, \xi_2\right)$, which is

$$p\left(a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2, \xi_1, \xi_2 \mid D^n, d_{n+1}\right) \cong N_6\left(m_p, V_p\right) \tag{5.9}$$

with $m_p = \left(\tilde{a}_{n+1,1}, \tilde{a}_{n+1,2}, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\xi}_1, \tilde{\xi}_2\right)$ and V_p defined such that

$$V_p^{-1} = \begin{pmatrix} \tilde{F}_1 & 0 & 0 & 0 & \tilde{E}_{n+1,1} & 0 \\ 0 & \tilde{F}_2 & 0 & 0 & 0 & \tilde{E}_{n+1,2} \\ 0 & 0 & \tilde{S}_1 & \tilde{S}_2 & \tilde{S}_3 & \tilde{S}_4 \\ 0 & 0 & \tilde{S}_2 & \tilde{S}_5 & \tilde{S}_6 & \tilde{S}_7 \\ \tilde{E}_{n+1,1} & 0 & \tilde{S}_3 & \tilde{S}_6 & \tilde{S}_8 & \tilde{S}_9 \\ 0 & \tilde{E}_{n+1,2} & \tilde{S}_4 & \tilde{S}_7 & \tilde{S}_9 & \tilde{S}_{10} \end{pmatrix},$$

where

$$\tilde{S}_1 = \tilde{G}_1 - \frac{\tilde{L}_1^2}{\tilde{R}_1} - \sum_{i=1}^{n} \frac{\tilde{A}_{i2}\tilde{C}_{i1}^2 + \tilde{A}_{i1}\tilde{D}_{i21}^2 - 2\tilde{B}_i\tilde{C}_{i1}\tilde{D}_{i21}}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2},$$

$$\tilde{S}_2 = \tilde{H} - \sum_{i=1}^{n} \frac{\tilde{D}_{i12}\left(\tilde{A}_{i2}\tilde{C}_{i1} - \tilde{B}_i\tilde{D}_{i21}\right) + \tilde{C}_{i2}\left(\tilde{A}_{i1}\tilde{D}_{i21} - \tilde{B}_i\tilde{C}_{i1}\right)}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2},$$

$$\tilde{S}_3 = \sum_{i=1}^{n} \frac{\tilde{E}_{i1}\left(\tilde{B}_i\tilde{D}_{i21} - \tilde{A}_{i2}\tilde{C}_{i1}\right)}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2}, \qquad \tilde{S}_4 = \sum_{i=1}^{n} \frac{\tilde{E}_{i2}\left(\tilde{B}_i\tilde{C}_{i1} - \tilde{A}_{i1}\tilde{D}_{i21}\right)}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2},$$

$$\tilde{S}_5 = \tilde{G}_2 - \frac{\tilde{L}_2^2}{\tilde{R}_2} - \sum_{i=1}^{n} \frac{\tilde{A}_{i2}\tilde{D}_{i12}^2 + \tilde{A}_{i1}\tilde{C}_{i2}^2 - 2\tilde{B}_i\tilde{C}_{i2}\tilde{D}_{i12}}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2}, \tag{5.10}$$

$$\tilde{S}_6 = \sum_{i=1}^{n} \frac{\tilde{E}_{i1}\left(\tilde{B}_i\tilde{C}_{i2} - \tilde{A}_{i2}\tilde{D}_{i12}\right)}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2}, \qquad \tilde{S}_7 = \sum_{i=1}^{n} \frac{\tilde{E}_{i2}\left(\tilde{B}_i\tilde{D}_{i12} - \tilde{A}_{i1}\tilde{C}_{i2}\right)}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2},$$

$$\tilde{S}_8 = \tilde{M}_1 - \sum_{i=1}^{n} \frac{\tilde{A}_{i2}\tilde{E}_{i1}^2}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2}, \quad \tilde{S}_9 = \sum_{i=1}^{n} \frac{\tilde{B}_i\tilde{E}_{i1}\tilde{E}_{i2}}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2}, \quad \tilde{S}_{10} = \tilde{M}_2 - \sum_{i=1}^{n} \frac{\tilde{A}_{i1}\tilde{E}_{i2}^2}{\tilde{A}_{i1}\tilde{A}_{i2} - \tilde{B}_i^2}.$$

From (5.9) we derive the full conditional distributions to be used in the Gibbs sampling algorithm in order to estimate the predictive distribution of Y_{n+1}. They are

$$p\left(a_{n+1,1} \mid a_{n+1,2}, \lambda_1, \lambda_2, \xi_1, \xi_2, D^n, d_{n+1}\right) = N\left(\tilde{a}_{n+1,1} - \frac{\left(\xi_1 - \tilde{\xi}_1\right)\tilde{E}_{n+1,1}}{\tilde{F}_1}, \frac{1}{\tilde{F}_1}\right),$$

$$p\left(a_{n+1,2} \mid a_{n+1,1}, \lambda_1, \lambda_2, \xi_1, \xi_2, D^n, d_{n+1}\right) = N\left(\tilde{a}_{n+1,2} - \frac{\left(\xi_2 - \tilde{\xi}_2\right)\tilde{E}_{n+1,2}}{\tilde{F}_2}, \frac{1}{\tilde{F}_2}\right),$$

$$p\left(\lambda_1 \mid a_{n+1,1}, a_{n+1,2}, \lambda_2, \xi_1, \xi_2, D^n, d_{n+1}\right) =$$

$$= N\left(\tilde{\lambda}_1 - \frac{\left(\lambda_2 - \tilde{\lambda}_2\right)\tilde{S}_2 + \left(\xi_1 - \tilde{\xi}_1\right)\tilde{S}_3 + \left(\xi_2 - \tilde{\xi}_2\right)\tilde{S}_4}{\tilde{S}_1}, \frac{1}{\tilde{S}_1}\right),$$

$$p\left(\lambda_2 \mid a_{n+1,1}, a_{n+1,2}, \lambda_1, \xi_1, \xi_2, D^n, d_{n+1}\right) =$$

$$= N\left(\tilde{\lambda}_2 - \frac{\left(\lambda_1 - \tilde{\lambda}_1\right)\tilde{S}_2 + \left(\xi_1 - \tilde{\xi}_1\right)\tilde{S}_6 + \left(\xi_2 - \tilde{\xi}_2\right)\tilde{S}_7}{\tilde{S}_5}, \frac{1}{\tilde{S}_5}\right),$$

$$p\left(\xi_1 \mid a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2, \xi_2, D^n, d_{n+1}\right) =$$

$$= N\left(\tilde{\xi}_1 - \frac{\left(a_{n+1,1} - \tilde{a}_{n+1,1}\right)\tilde{E}_{n+1,1} + \left(\lambda_1 - \tilde{\lambda}_1\right)\tilde{S}_3 + \left(\lambda_2 - \tilde{\lambda}_2\right)\tilde{S}_6 + \left(\xi_2 - \tilde{\xi}_2\right)\tilde{S}_9}{\tilde{S}_8}, \frac{1}{\tilde{S}_8}\right),$$

$$p\left(\xi_2 \mid a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2, \xi_1, D^n, d_{n+1}\right) =$$

$$= N\left(\tilde{\xi}_2 - \frac{\left(a_{n+1,2} - \tilde{a}_{n+1,2}\right)\tilde{E}_{n+1,2} + \left(\lambda_1 - \tilde{\lambda}_1\right)\tilde{S}_4 + \left(\lambda_2 - \tilde{\lambda}_2\right)\tilde{S}_7 + \left(\xi_1 - \tilde{\xi}_1\right)\tilde{S}_9}{\tilde{S}_{10}}, \frac{1}{\tilde{S}_{10}}\right).$$

No relevant simplifications are achieved if we consider vague second stage priors.

5.2.4. Laplace Approximation

The predictive distribution of Y_{n+1} can be regarded as a function of the posterior expectation of

$$g\left(a_{n+1,1}, a_{n+1,2}, \lambda_1, \lambda_2\right) = \exp\left\{-\exp\left\{\lambda_1 a_{n+1,1} + \lambda_2 a_{n+1,2}\right\}\right\}\exp\left\{\left(\lambda_1 a_{n+1,1} + \lambda_2 a_{n+1,2}\right)y_{n+1}\right\}$$

with respect to the posterior distribution (5.3).

Following the summary presented in section 1.2.4, let us define two functions $h(\theta)$ and $h^*(\theta)$ satisfying (1.13) and (1.14), that is, such that

$$-n h(\theta) = -\sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(1 + e^{\xi_j}\right)\exp\left\{a_{ij} + k_{ij}\right\} - \sum_{i=1}^{n}\exp\left\{\lambda_1 a_{i1} + \lambda_2 a_{i2}\right\} - \sum_{j=1}^{2} e^{\eta_j}\left(s_j + e^{\lambda_j}\right) -$$

$$-\sum_{j=1}^{2} v_j e^{\xi_j} + \sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(x_{ij} + b_j\right)\left(a_{ij} + k_{ij}\right) + \sum_{j=1}^{2}\left((n+1)b_j + u_j\right)\xi_j +$$

$$+\sum_{j=1}^{2}\left(d_j + r_j\right)\eta_j + \sum_{j=1}^{2}\lambda_j\left(d_j + \sum_{i=1}^{n} a_{ij} y_i\right)$$

and

$$-n h^*(\theta) = -\sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(1 + e^{\xi_j}\right)\exp\left\{a_{ij} + k_{ij}\right\} - \sum_{i=1}^{n+1}\exp\left\{\lambda_1 a_{i1} + \lambda_2 a_{i2}\right\} - \sum_{j=1}^{2} e^{\eta_j}\left(s_j + e^{\lambda_j}\right) -$$

$$-\sum_{j=1}^{2} v_j e^{\xi_j} + \sum_{i=1}^{n+1}\sum_{j=1}^{2}\left(x_{ij}+b_j\right)\left(a_{ij}+k_{ij}\right)+\sum_{j=1}^{2}\left((n+1)b_j+u_j\right)\xi_j +$$

$$+\sum_{j=1}^{2}\left(d_j+r_j\right)\eta_j+\sum_{j=1}^{2}\lambda_j\left(d_j+\sum_{i=1}^{n+1}a_{ij}y_i\right).$$

From (1.16), we conclude that $\tilde{\theta} = \boldsymbol{m}$, the solution of the system of equations (5.6) and that

$$\tilde{\sigma} = \left\{\tilde{T}\tilde{R}_1\tilde{R}_2\tilde{F}_1\tilde{F}_2\prod_{i=1}^{n}\left\{\tilde{A}_{i1}\tilde{A}_{i2}-\tilde{B}_i^2\right\}\right\}^{-1/2}$$

with

$$\tilde{T} = \tilde{S}_1\tilde{S}_5\tilde{S}_8\tilde{S}_{10} - \tilde{S}_1\tilde{S}_5\tilde{S}_9^2 - \tilde{S}_1\tilde{S}_{10}\tilde{S}_6^2 + 2\tilde{S}_1\tilde{S}_6\tilde{S}_7\tilde{S}_9 - \tilde{S}_1\tilde{S}_8\tilde{S}_7^2 - \tilde{S}_8\tilde{S}_{10}\tilde{S}_2^2 + \tilde{S}_2^2\tilde{S}_9^2 +$$

$$+2\tilde{S}_2\tilde{S}_3\tilde{S}_6\tilde{S}_{10} - 2\tilde{S}_2\tilde{S}_4\tilde{S}_6\tilde{S}_9 - 2\tilde{S}_2\tilde{S}_3\tilde{S}_7\tilde{S}_9 + 2\tilde{S}_2\tilde{S}_4\tilde{S}_7\tilde{S}_8 - \tilde{S}_5\tilde{S}_{10}\tilde{S}_3^2 +$$

$$+2\tilde{S}_3\tilde{S}_4\tilde{S}_5\tilde{S}_9 + \tilde{S}_3^2\tilde{S}_7^2 - 2\tilde{S}_3\tilde{S}_4\tilde{S}_6\tilde{S}_7 - \tilde{S}_5\tilde{S}_8\tilde{S}_4^2 + \tilde{S}_4^2\tilde{S}_6^2$$

and all constants involved defined as in (5.7) and (5.10), except $\tilde{S}_8$ and $\tilde{S}_{10}$ which are now defined as

$$\tilde{S}_8 = \tilde{M}_1 - \sum_{i=1}^{n}\frac{\tilde{A}_{i2}\tilde{E}_{i1}^2}{\tilde{A}_{i1}\tilde{A}_{i2}-\tilde{B}_i^2} - \frac{\tilde{E}_{n+1,1}^2}{\tilde{F}_1} \quad \text{and} \quad \tilde{S}_{10} = \tilde{M}_2 - \sum_{i=1}^{n}\frac{\tilde{A}_{i1}\tilde{E}_{i2}^2}{\tilde{A}_{i1}\tilde{A}_{i2}-\tilde{B}_i^2} - \frac{\tilde{E}_{n+1,2}^2}{\tilde{F}_2}.$$

Still from the definitions in (1.16), we define θ^* as the numerical solution of the system formed by the equations

$$\left(1+e^{\xi_j}\right)\exp\left\{a_{ij}+k_{ij}\right\}+\lambda_j\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}-b_j-x_{ij}-\lambda_j y_i = 0,$$
$$i = 1, 2, \ldots, n+1 \;;\; j = 1, 2,$$

$$\sum_{i=1}^{n+1}a_{ij}y_i - \sum_{i=1}^{n+1}a_{ij}\exp\left\{\lambda_1 a_{i1}+\lambda_2 a_{i2}\right\}+d_j-e^{\eta_j+\lambda_j} = 0, \quad j = 1, 2,$$

$$e^{\xi_j}\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)-(n+1)b_j-u_j = 0, \quad j = 1, 2,$$

$$e^{\eta_j}\left(s_j + e^{\lambda_j}\right) - d_j - r_j = 0, \quad j = 1, 2,$$

and

$$\sigma^* = \left\{T^* R_1^* R_2^* \prod_{i=1}^{n+1} \left\{A_{i1}^* A_{i2}^* - B_i^{*2}\right\}\right\}^{-1/2}$$

with

$$A_{ij}^* = \left(1 + e^{\xi_j^*}\right)\exp\left\{a_{ij}^* + k_{ij}\right\} - \lambda_j^{*2}\exp\left\{\lambda_1^* a_{i1}^* + \lambda_2^* a_{i2}^*\right\}, \quad i = 1, 2, \ldots, n+1 \;\; ; \;\; j = 1, 2,$$

$$B_i^* = \lambda_1^* \lambda_2^* \exp\left\{\lambda_1^* a_{i1}^* + \lambda_2^* a_{i2}^*\right\}, \quad i = 1, 2, \ldots, n+1, \qquad R_j^* = e^{\eta_j^*}\left(s_j + e^{\lambda_j^*}\right), \quad j = 1, 2,$$

$$T^* = S_1^* S_5^* S_8^* S_{10}^* - S_1^* S_5^* S_9^{*2} - S_1^* S_{10}^* S_6^{*2} + 2S_1^* S_6^* S_7^* S_9^* - S_1^* S_8^* S_7^{*2} - S_8^* S_{10}^* S_2^{*2} +$$

$$+ S_2^{*2} S_9^{*2} + 2S_2^* S_3^* S_6^* S_{10}^* - 2S_2^* S_4^* S_6^* S_9^* - 2S_2^* S_3^* S_7^* S_9^* + 2S_2^* S_4^* S_7^* S_8^* -$$

$$- S_5^* S_{10}^* S_3^{*2} + 2S_3^* S_4^* S_5^* S_9^* + S_3^{*2} S_7^{*2} - 2S_3^* S_4^* S_6^* S_7^* - S_5^* S_8^* S_4^{*2} + S_4^{*2} S_6^{*2},$$

where

$$S_1^* = G_1^* - \frac{L_1^{*2}}{R_1^*} - \sum_{i=1}^{n+1} \frac{A_{i2}^* C_{i1}^{*2} + A_{i1}^* D_{i21}^{*2} - 2B_i^* C_{i1}^* D_{i21}^*}{A_{i1}^* A_{i2}^* - B_i^{*2}},$$

$$S_2^* = H^* - \sum_{i=1}^{n+1} \frac{D_{i12}^*\left(A_{i2}^* C_{i1}^* - B_i^* D_{i21}^*\right) + C_{i2}^*\left(A_{i1}^* D_{i21}^* - B_i^* C_{i1}^*\right)}{A_{i1}^* A_{i2}^* - B_i^{*2}},$$

$$S_3^* = \sum_{i=1}^{n+1} \frac{E_{i1}^*\left(B_i^* D_{i21}^* - A_{i2}^* C_{i1}^*\right)}{A_{i1}^* A_{i2}^* - B_i^{*2}}, \qquad S_4^* = \sum_{i=1}^{n+1} \frac{E_{i2}^*\left(B_i^* C_{i1}^* - A_{i1}^* D_{i21}^*\right)}{A_{i1}^* A_{i2}^* - B_i^{*2}},$$

$$S_5^* = G_2^* - \frac{L_2^{*2}}{R_2^*} - \sum_{i=1}^{n+1} \frac{A_{i2}^* D_{i12}^{*2} + A_{i1}^* C_{i2}^{*2} - 2B_i^* C_{i2}^* D_{i12}^*}{A_{i1}^* A_{i2}^* - B_i^{*2}},$$

$$S_6^* = \sum_{i=1}^{n+1} \frac{E_{i1}^*\left(B_i^* C_{i2}^* - A_{i2}^* D_{i12}^*\right)}{A_{i1}^* A_{i2}^* - B_i^{*2}}, \qquad S_7^* = \sum_{i=1}^{n+1} \frac{E_{i2}^*\left(B_i^* D_{i12}^* - A_{i1}^* C_{i2}^*\right)}{A_{i1}^* A_{i2}^* - B_i^{*2}},$$

$$S_8^* = M_1^* - \sum_{i=1}^{n+1} \frac{A_{i2}^* E_{i1}^{*2}}{A_{i1}^* A_{i2}^* - B_i^{*2}}, \quad S_9^* = \sum_{i=1}^{n+1} \frac{B_i^* E_{i1}^* E_{i2}^*}{A_{i1}^* A_{i2}^* - B_i^{*2}}, \quad S_{10}^* = M_2^* - \sum_{i=1}^{n+1} \frac{A_{i1}^* E_{i2}^{*2}}{A_{i1}^* A_{i2}^* - B_i^{*2}},$$

and

$$C_{ij}^* = \left(1 + \lambda_j^* a_{ij}^*\right) \exp\left\{\lambda_1^* a_{i1}^* + \lambda_2^* a_{i2}^*\right\} - y_i, \quad i = 1, 2, \dots, n+1 \ ; \ j = 1, 2,$$

$$D_{ijk}^* = \lambda_j^* a_{ik}^* \exp\left\{\lambda_1^* a_{i1}^* + \lambda_2^* a_{i2}^*\right\}, \quad i = 1, 2, \dots, n+1 \ ; \ j, k = 1, 2 \ ; \ j \neq k,$$

$$E_{ij}^* = \exp\left\{\xi_j^* + a_{ij}^* + k_{ij}\right\}, \quad i = 1, 2, \dots, n+1 \ ; \ j = 1, 2,$$

$$G_j^* = \sum_{i=1}^{n+1} a_{ij}^{*2} \exp\left\{\lambda_1^* a_{i1}^* + \lambda_2^* a_{i2}^*\right\} + e^{\eta_j^* + \lambda_j^*}, \quad j = 1, 2, \qquad L_j^* = e^{\eta_j^* + \lambda_j^*}, \quad j = 1, 2,$$

$$H^* = \sum_{i=1}^{n+1} a_{i1}^* a_{i2}^* \exp\left\{\lambda_1^* a_{i1}^* + \lambda_2^* a_{i2}^*\right\}, \quad M_j^* = e^{\xi_j^*}\left(v_j + \sum_{i=1}^{n+1} \exp\left\{a_{ij}^* + k_{ij}\right\}\right), \quad j = 1, 2.$$

Finally, by (1.17), the predictive distribution of Y_{n+1} will be approximately given by

$$p\left(y_{n+1} \mid D^n, d_{n+1}\right) \tilde{\propto} \frac{1}{y_{n+1}!} \left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\left\{-nh^*\left(\theta^*\right) + n h\left(\tilde{\theta}\right)\right\}.$$

CHAPTER 6

A GENERAL MODEL TO PREDICT THE NUMBER OF ACCIDENTS AT A ROAD JUNCTION

In this chapter we will generalise the model studied in the previous chapter. The predictions about the number of accidents occurring at a road junction in a given period of time will be based on the measurement of f traffic flows and on c covariates describing characteristics of the junction.

Let $X_{i1}, X_{i2}, \ldots, X_{if}$ and Y_i $(i = 1, 2, \ldots, n+1)$ be random variables representing, respectively, the measurements of the f traffic flows and the number of accidents which occurred at the i-th junction. We assume that, for $i = 1, 2, \ldots, n+1$,

$$
\begin{aligned}
X_{i1} &\sim Po\left(\exp\left(a_{i1} + k_{i1}\right)\right),\\
X_{i2} &\sim Po\left(\exp\left(a_{i2} + k_{i2}\right)\right),\\
&\vdots\\
X_{if} &\sim Po\left(\exp\left(a_{if} + k_{if}\right)\right),\\
Y_i &\sim Po\left(\exp\left(\lambda_1 a_{i1} + \lambda_2 a_{i2} + \ldots + \lambda_f a_{if} + \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_c z_{ic}\right)\right),
\end{aligned}
$$

or equivalently,

$$
\begin{aligned}
X_{ij} &\sim Po\left(\exp\left(a_{ij} + k_{ij}\right)\right), \quad j = 1, 2, \ldots, f,\\
Y_i &\sim Po\left(\exp\left(\sum_{j=1}^{f} \lambda_j a_{ij} + \sum_{t=1}^{c} \beta_t z_{it}\right)\right).
\end{aligned}
$$

Here, a_{ij} $(i = 1, 2, \ldots, n+1$ and $j = 1, 2, \ldots, f)$ are unknown parameters modelling the j-th traffic flow at the i-th junction, λ_j $(j = 1, 2, \ldots, f)$ are also unknown parameters used to model the effect of the j-th traffic flow upon the number of accidents, $z_{i\ell}$ $(i = 1, 2, \ldots, n+1$ and $\ell = 1, 2, \ldots, c)$ are covariates whose values, assumed to be known, describe the ℓ-th characteristic of the i-th junction and β_ℓ $(\ell = 1, 2, \ldots, c)$ are parameters used to model the influence of the ℓ-th characteristic upon the number of accidents. As in Chapter 5, k_{ij} $(i = 1, 2, \ldots, n+1$ and $j = 1, 2, \ldots, f)$ are supposed to be known constants which relate to the length of the observational period and the time of day and year that the observations were made.

The available data set is

$$D^n = \left\{\left(x_{i1}, x_{i2}, \ldots, x_{if}, y_i, k_{i1}, k_{i2}, \ldots, k_{if}, z_{i1}, z_{i2}, \ldots, z_{ic}\right), i = 1, 2, \ldots, n\right\},$$

formed by the observations made on n road junctions. Given a new junction with observed measurements

$$d_{n+1} = \left(x_{n+1,1}, x_{n+1,2}, \ldots, x_{n+1,f}, k_{n+1,1}, k_{n+1,2}, \ldots, k_{n+1,f}, z_{n+1,1}, z_{n+1,2}, \ldots, z_{n+1,c}\right),$$

our aim is to derive the predictive distribution of Y_{n+1}, the number of accidents that will occur at that new junction.

6.1. A Classical Approach

Let us define a parameter vector $\theta_{red} = (a_{11}, a_{21}, \ldots, a_{n1}, a_{12}, a_{22}, \ldots, a_{n2}, \ldots, a_{1f}, a_{2f}, \ldots, a_{nf}, a_{n+1,1}, a_{n+1,2}, \ldots, a_{n+1,f}, \lambda_1, \lambda_2, \ldots, \lambda_f, \beta_1, \beta_2, \ldots, \beta_c)$. The likelihood function is

$$L\left(\theta_{red}; D^n, d_{n+1}\right) \propto \exp\left\{-\sum_{i=1}^{n+1}\sum_{j=1}^{f}\exp\left\{a_{ij} + k_{ij}\right\} - \sum_{i=1}^{n}\exp\left[\sum_{j=1}^{f}\lambda_j a_{ij} + \sum_{\ell=1}^{c}\beta_\ell z_{i\ell}\right]\right\} \times$$

$$\times \exp\left\{\sum_{i=1}^{n+1}\sum_{j=1}^{f}\left(a_{ij} + k_{ij}\right)x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{f}\lambda_j a_{ij} y_i + \sum_{i=1}^{n}\sum_{\ell=1}^{c}\beta_\ell z_{i\ell} y_i\right\},$$

and we then easily conclude that the maximum likelihood estimates for $a_{n+1,j}$ ($j=1, 2, \ldots, f$) are given by

$$\hat{a}_{n+1,j} = \ln\left(x_{n+1,j}\right) - k_{n+1,j}\,, \quad j = 1, 2, \ldots, f, \tag{6.1}$$

whilst the remaining maximum likelihood estimates are obtained solving numerically the system formed by the $((n+1)f + c)$ equations

$$\begin{aligned}
&\exp\left\{a_{ij} + k_{ij}\right\} + \lambda_j \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} - x_{ij} - \lambda_j y_i = 0, \\
&\qquad\qquad\qquad\qquad\qquad\qquad i = 1, 2, \ldots, n \;;\; j = 1, 2, \ldots, f, \\
&\sum_{i=1}^{n} a_{ij} y_i - \sum_{i=1}^{n} a_{ij} \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} = 0, \quad j = 1, 2, \ldots, f, \\
&\sum_{i=1}^{n} z_{i\ell} y_i - \sum_{i=1}^{n} z_{i\ell} \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} = 0, \quad \ell = 1, 2, \ldots, c.
\end{aligned} \tag{6.2}$$

A simple plug-in estimative approximation for the predictive distribution of Y_{n+1} would then be

$$Y_{n+1} \sim Po\left(\exp\left(\sum_{j=1}^{f} \hat{\lambda}_j \hat{a}_{n+1,j} + \sum_{\ell=1}^{c} \hat{\beta}_\ell z_{n+1,\ell}\right)\right).$$

6.2. A Bayesian Approach

Once again we consider a hierarchical prior structure. At the first stage we take

$$p\left(\theta_{red} \mid \xi_1, \xi_2, \ldots, \xi_f, \eta_1, \eta_2, \ldots, \eta_f, \zeta_1, \zeta_2, \ldots, \zeta_c\right) =$$

$$= \prod_{i=1}^{n+1} \prod_{j=1}^{f} \left\{p\left(a_{ij} \mid \xi_j\right)\right\} \prod_{j=1}^{f} \left\{p\left(\lambda_j \mid \eta_j\right)\right\} \prod_{\ell=1}^{c} \left\{p\left(\beta_\ell \mid \zeta_\ell\right)\right\},$$

and at the second stage we take

$$p(\xi_1, \xi_2, \dots, \xi_f, \eta_1, \eta_2, \dots, \eta_f, \zeta_1, \zeta_2, \dots, \zeta_c) = \prod_{j=1}^{f}\{p(\xi_j)p(\eta_j)\}\prod_{\ell=1}^{c}\{p(\zeta_\ell)\},$$

with

$$\exp(a_{ij} + k_{ij}) \sim Ga(b_j, e^{\xi_j}), \quad i = 1, 2, \dots, n+1 \;;\; j = 1, 2, \dots, f,$$

$$\exp(\lambda_j) \sim Ga(d_j, e^{\eta_j}), \quad j = 1, 2, \dots, f, \qquad \exp(\beta_\ell) \sim Ga(g_\ell, e^{\zeta_\ell}), \quad \ell = 1, 2, \dots, c,$$

$$\exp(\xi_j) \sim Ga(u_j, v_j), \qquad \exp(\eta_j) \sim Ga(r_j, s_j), \quad j = 1, 2, \dots, f,$$

$$\exp(\zeta_\ell) \sim Ga(t_\ell, w_\ell), \quad \ell = 1, 2, \dots, c.$$

The joint posterior distribution for the parameter vector $\theta = (\theta_{red}, \xi_1, \xi_2, \dots, \xi_f, \eta_1, \eta_2, \dots, \eta_f, \zeta_1, \zeta_2, \dots, \zeta_c)$ is

$$\begin{aligned}
p(\theta \mid D^n, d_{n+1}) \propto\ & \exp\left\{-\sum_{i=1}^{n+1}\sum_{j=1}^{f}\left(1 + e^{\xi_j}\right)\exp\{a_{ij} + k_{ij}\} - \sum_{i=1}^{n}\exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip} + \sum_{m=1}^{c}\beta_m z_{im}\right\}\right\}\times \\
& \times\exp\left\{-\sum_{j=1}^{f} v_j e^{\xi_j} - \sum_{j=1}^{f} e^{\eta_j}\left(s_j + e^{\lambda_j}\right) - \sum_{\ell=1}^{c} e^{\zeta_\ell}\left(w_\ell + e^{\beta_\ell}\right)\right\}\times \\
& \times\exp\left\{\sum_{i=1}^{n+1}\sum_{j=1}^{f}(a_{ij} + k_{ij})(b_j + x_{ij}) + \sum_{j=1}^{f}((n+1)b_j + u_j)\xi_j\right\}\times \\
& \times\exp\left\{\sum_{j=1}^{f}\eta_j(d_j + r_j) + \sum_{\ell=1}^{c}\zeta_\ell(g_\ell + t_\ell) + \sum_{j=1}^{f}\lambda_j\left(d_j + \sum_{i=1}^{n} a_{ij} y_i\right)\right\}\times \\
& \times\exp\left\{\sum_{\ell=1}^{c}\beta_\ell\left(g_\ell + \sum_{i=1}^{n} z_{i\ell} y_i\right)\right\}. \qquad (6.3)
\end{aligned}$$

Eliminating the hyperparameters $\xi_1, \xi_2, \dots, \xi_f, \eta_1, \eta_2, \dots, \eta_f, \zeta_1, \zeta_2, \dots, \zeta_c$, we obtain the marginal posterior distribution for θ_{red}, which is given by

$$p\left(\theta_{red} \mid D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\sum_{i=1}^{n+1}\sum_{j=1}^{f}\exp\left\{a_{ij}+k_{ij}\right\}-\sum_{i=1}^{n}\exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip}+\sum_{m=1}^{c}\beta_m z_{im}\right\}\right\}}{\prod_{j=1}^{f}\left\{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}\right\}} \times$$

$$\times \frac{\exp\left\{\sum_{i=1}^{n+1}\sum_{j=1}^{f}\left(a_{ij}+k_{ij}\right)\left(b_j+x_{ij}\right)+\sum_{j=1}^{f}\lambda_j\left(d_j+\sum_{i=1}^{n}a_{ij}y_i\right)+\sum_{\ell=1}^{c}\beta_\ell\left(g_\ell+\sum_{i=1}^{n}z_{i\ell}y_i\right)\right\}}{\prod_{j=1}^{f}\left\{\left(s_j+e^{\lambda_j}\right)^{d_j+r_j}\right\}\prod_{\ell=1}^{c}\left\{\left(w_\ell+e^{\beta_\ell}\right)^{g_\ell+t_\ell}\right\}} . \quad (6.4)$$

6.2.1. The Exact Predictive Distribution

In a Bayesian framework, the probability function of Y_{n+1} combined with the posterior distribution (6.4), provides the predictive distribution of Y_{n+1}

$$p\left(y_{n+1} \mid D^n, d_{n+1}\right) \propto \frac{1}{y_{n+1}!}\int_{\Re^{(n+2)f+c}} \frac{\exp\left\{-\sum_{i=1}^{n+1}\sum_{j=1}^{f}\exp\left\{a_{ij}+k_{ij}\right\}\right\}}{\prod_{j=1}^{f}\left\{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}\right\}} \times$$

$$\times \frac{\exp\left\{-\sum_{i=1}^{n}\exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip}+\sum_{m=1}^{c}\beta_m z_{im}\right\}\right\}\exp\left\{\sum_{i=1}^{n+1}\sum_{j=1}^{f}\left(a_{ij}+k_{ij}\right)\left(b_j+x_{ij}\right)\right\}}{\prod_{j=1}^{f}\left\{\left(s_j+e^{\lambda_j}\right)^{d_j+r_j}\right\}} \times$$

$$\times \frac{\exp\left\{\sum_{j=1}^{f}\lambda_j\left(d_j+\sum_{i=1}^{n+1}a_{ij}y_i\right)+\sum_{\ell=1}^{c}\beta_\ell\left(g_\ell+\sum_{i=1}^{n+1}z_{i\ell}y_i\right)\right\}}{\prod_{\ell=1}^{c}\left\{\left(w_\ell+e^{\beta_\ell}\right)^{g_\ell+t_\ell}\right\}} d\theta_{red} .$$

To evaluate these predictive probabilities, a numerical integration technique is required. However, the high dimensionality of the integral is very likely to cause numerical problems.

6.2.2. Estimation Via Gibbs Sampling

The implementation of the Gibbs sampling algorithm (section 1.2.1) requires the full conditional distributions which can easily be derived from (6.4). In order to simplify future notation, let us define, as in Chapter 5, a vector $\theta_{red}^{(\theta_i)}$ as being the parameter vector θ_{red} with the component θ_i removed. The full conditional distributions from which we have to generate random values are

$$p\left(a_{ij} \mid \theta_{red}^{(a_{ij})}, D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\exp\left\{a_{ij}+k_{ij}\right\}-\exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip}+\sum_{m=1}^{c}\beta_m z_{im}\right\}\right\}}{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}}\times$$

$$\times\exp\left\{a_{ij}\left(b_j+x_{ij}+\lambda_j y_i\right)\right\},\quad i=1,2,\ldots,n\ ;\ j=1,2,\ldots,f,$$

$$p\left(a_{n+1,j} \mid \theta_{red}^{(a_{n+1,j})}, D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\exp\left\{a_{n+1,j}+k_{n+1,j}\right\}\right\}\exp\left\{a_{n+1,j}\left(b_j+x_{n+1,j}\right)\right\}}{\left(v_j+\sum_{i=1}^{n+1}\exp\left\{a_{ij}+k_{ij}\right\}\right)^{(n+1)b_j+u_j}},$$

$$j=1,2,\ldots,f,$$

$$p\left(\lambda_j \mid \theta_{red}^{(\lambda_j)}, D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\sum_{i=1}^{n}\exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip}+\sum_{m=1}^{c}\beta_m z_{im}\right\}\right\}\exp\left\{\lambda_j\left(d_j+\sum_{i=1}^{n}a_{ij}y_i\right)\right\}}{\left(s_j+e^{\lambda_j}\right)^{d_j+r_j}},$$

$$j=1,2,\ldots,f,$$

$$p\left(\beta_\ell \mid \theta_{red}^{(\beta_\ell)}, D^n, d_{n+1}\right) \propto \frac{\exp\left\{-\sum_{i=1}^{n}\exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip}+\sum_{m=1}^{c}\beta_m z_{im}\right\}\right\}\exp\left\{\beta_\ell\left(g_\ell+\sum_{i=1}^{n}z_{i\ell}y_i\right)\right\}}{\left(w_\ell+e^{\beta_\ell}\right)^{g_\ell+t_\ell}},$$

$$\ell=1,2,\ldots,c.$$

Note that the generation of random values from these distributions requires a sampling technique such as the rejection sampling algorithm (section 1.2.3). The use of the adaptive rejection sampling algorithm (section 1.2.3.2) should be preferred.

6.2.3. Estimation Via Asymptotic Results

Supposing n is large enough, the posterior distribution can be approximated by a multivariate normal distribution (section 1.2.2), which makes the implementation of the Gibbs routine to be quite easy and efficient when used to estimate the predictive distribution of Y_{n+1}.

6.2.3.1. Posterior Normality Based on the Likelihood Function

The asymptotic result presented by Bernardo & Smith (1994) and summarised in section 1.2.2.1 is based on the maximum likelihood estimates given by (6.1) and (6.2) and it is independent of the prior structure we consider. Following that result and using the maximum likelihood estimates, we define the constants

$$A_{ij} = \exp\left\{\hat{a}_{ij} + k_{ij}\right\} + \hat{\lambda}_j^2 \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad i = 1, 2, \ldots, n \ ; \ j = 1, 2, \ldots, f,$$

$$B_{ijk} = \hat{\lambda}_j \hat{\lambda}_k \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad i = 1, 2, \ldots, n \ ; \ j, k = 1, 2, \ldots, f \ ; \ j \neq k,$$

$$C_{ij} = \left(1 + \hat{\lambda}_j \hat{a}_{ij}\right) \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\} - y_i, \quad i = 1, 2, \ldots, n \ ; \ j = 1, 2, \ldots, f,$$

$$D_{ijk} = \hat{\lambda}_j \hat{a}_{ik} \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad i = 1, 2, \ldots, n \ ; \ j, k = 1, 2, \ldots, f \ ; \ j \neq k,$$

$$E_{ijk} = \hat{\lambda}_j z_{ik} \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, f \ ; k = 1, 2, \ldots, c,$$

$$F_j = \exp\left\{\hat{a}_{n+1,j} + k_{n+1,j}\right\}, \quad j = 1, 2, \ldots, f, \tag{6.5}$$

$$G_j = \sum_{i=1}^{n} \hat{a}_{ij}^2 \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad j = 1, 2, \ldots, f,$$

$$H_{jk} = \sum_{i=1}^{n} \hat{a}_{ij}\hat{a}_{ik} \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad j, k = 1, 2, \ldots, f \;\; ; \;\; j \neq k,$$

$$L_{jk} = \sum_{i=1}^{n} \hat{a}_{ij}\, z_{ik} \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad j = 1, 2, \ldots, f \;\; ; \;\; k = 1, 2, \ldots, c,$$

$$M_{\ell} = \sum_{i=1}^{n} z_{i\ell}^{2} \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad \ell = 1, 2, \ldots, c,$$

$$R_{\ell k} = \sum_{i=1}^{n} z_{i\ell}\, z_{ik} \exp\left\{\sum_{p=1}^{f} \hat{\lambda}_p \hat{a}_{ip} + \sum_{m=1}^{c} \hat{\beta}_m z_{im}\right\}, \quad \ell, k = 1, 2, \ldots, c \;\; ; \;\; \ell \neq k,$$

and, considering an indicator function

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

we also define

$$T_{1p} = G_p - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n} \left(C_{ki}\delta_{ip} + D_{kip}\left(1-\delta_{ip}\right)\right)\left(C_{kj}\delta_{jp} + D_{kjp}\left(1-\delta_{jp}\right)\right) W_{kij}, \; p = 1, 2, \ldots, f,$$

$$S_{1pq} = H_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n} \left(C_{ki}\delta_{ip} + D_{kip}\left(1-\delta_{ip}\right)\right)\left(C_{kj}\delta_{jq} + D_{kjq}\left(1-\delta_{jq}\right)\right) W_{kij},$$

$$p, q = 1, 2, \ldots, f \;\; ; \;\; p \neq q,$$

$$Q_{pq} = L_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n} \left(C_{ki}\delta_{ip} + D_{kip}\left(1-\delta_{ip}\right)\right) E_{kjq} W_{kij}, \; p = 1, 2, \ldots, f \; ; \; q = 1, 2, \ldots, c,$$

$$T_{2p} = M_p - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n} E_{kip}\, E_{kjp}\, W_{kij}, \quad p = 1, 2, \ldots, c,$$

$$S_{2pq} = R_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n} E_{kip} E_{kjq} W_{kij}, \quad p, q = 1, 2, \ldots, c \;\; ; p \neq q,$$

where the constants involved are defined in (6.5), except W_{kij} ($k=1, 2, \ldots, n$; $i=1, 2, \ldots, f$; $j=1, 2, \ldots, f$) which are obtained numerically inverting the matrix

$$\begin{pmatrix} A_1 & B_{12} & B_{13} & \cdots & B_{1f} \\ B_{12} & A_2 & B_{23} & \cdots & B_{2f} \\ B_{13} & B_{23} & A_3 & \cdots & B_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{1f} & B_{2f} & B_{3f} & \cdots & A_f \end{pmatrix} \tag{6.6}$$

whose blocks are defined by

$$A_j = diag\left(A_{ij}, i = 1, 2, \dots, n\right), \quad j = 1, 2, \dots, f,$$

$$B_{ij} = diag\left(B_{kij}, k = 1, 2, \dots, n\right), \quad i, j = 1, 2, \dots, f \;\; ; \; i \neq j.$$

We can take advantage of the special form of the matrix in (6.6) to obtain its inverse. Such an inverse will be a symmetric matrix with diagonal blocks, that is, the inverse of (6.6) will be of the form

$$\begin{pmatrix} W_{11} & W_{12} & W_{13} & \cdots & W_{1f} \\ W_{12} & W_{22} & W_{23} & \cdots & W_{2f} \\ W_{13} & W_{23} & W_{33} & \cdots & W_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_{1f} & W_{2f} & W_{3f} & \cdots & W_{ff} \end{pmatrix}$$

with

$$W_{ij} = diag\left(W_{kij}, k = 1, 2, \dots, n\right), \quad i, j = 1, 2, \dots, f.$$

A numerical algorithm is required to invert (6.6) and hence derive W_{kij} ($k=1, 2, \dots, n$; $i=1, 2, \dots, f$; $j=1, 2, \dots, f$).

Finally, the predictive distribution of Y_{n+1} can be estimated through the Gibbs routine (section 1.2.1) using the full conditional distributions

$$p\left(a_{n+1,j} \mid a_{n+1,k(k\neq j)}, \lambda_1, \lambda_2, \dots, \lambda_f, \beta_1, \beta_2, \dots, \beta_c, D^n, d_{n+1}\right) = N\left(\hat{a}_{n+1,j}, \frac{1}{F_j}\right),$$

$$j = 1, 2, \dots, f,$$

$$p\left(\lambda_j \mid a_{n+1,1}, a_{n+1,2}, \dots, a_{n+1,f}, \lambda_{k(k\neq j)}, \beta_1, \beta_2, \dots, \beta_c, D^n, d_{n+1}\right) =$$

$$= N\left(\hat{\lambda}_j - \frac{\sum_{\substack{i=1 \\ i \neq j}}^{f} \left(\lambda_i - \hat{\lambda}_i\right) S_{1ij} + \sum_{i=1}^{c} \left(\beta_i - \hat{\beta}_i\right) Q_{ji}}{T_{1j}}, \frac{1}{T_{1j}} \right), \quad j = 1, 2, \ldots, f,$$

$$p\left(\beta_\ell \mid a_{n+1,1}, a_{n+1,2}, \ldots, a_{n+1,f}, \lambda_1, \lambda_2, \ldots, \lambda_f, \beta_{k(k \neq \ell)}, D^n, d_{n+1}\right) =$$

$$= N\left(\hat{\beta}_\ell - \frac{\sum_{\substack{i=1 \\ i \neq \ell}}^{c} \left(\beta_i - \hat{\beta}_i\right) S_{2i\ell} + \sum_{i=1}^{f} \left(\lambda_i - \hat{\lambda}_i\right) Q_{i\ell}}{T_{2\ell}}, \frac{1}{T_{2\ell}} \right), \quad \ell = 1, 2, \ldots, c.$$

6.2.3.2. Posterior Normality Based on Characteristics of the Posterior Distribution

We now consider the posterior normal approximation suggested by O'Hagan (1994) and summarised in section 1.2.2.2. Following that result, the posterior distribution (6.3) will be approximately

$$p\left(\theta \mid D^n, d_{n+1}\right) \cong N(m, V),$$

where m is the posterior mode and V is the modal dispersion matrix, defined by (1.4) and (1.5).

The posterior mode $m = \tilde{\theta}$ is obtained solving numerically the system formed by the $((n+4)f+2c)$ equations

$$\left(1 + e^{\xi_j}\right)\exp\left\{a_{ij} + k_{ij}\right\} + \lambda_j \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} - b_j - x_{ij} - \lambda_j y_i = 0,$$

$$i = 1, 2, \ldots, n \;;\; j = 1, 2, \ldots, f,$$

$$\left(1 + e^{\xi_j}\right)\exp\left\{a_{n+1,j} + k_{n+1,j}\right\} - b_j - x_{n+1,j} = 0, \quad j = 1, 2, \ldots, f,$$

$$e^{\eta_j+\lambda_j} + \sum_{i=1}^{n} a_{ij} \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} - \sum_{i=1}^{n} a_{ij} y_i - d_j = 0, \quad j = 1, 2, \dots, f, \qquad (6.7)$$

$$e^{\zeta_\ell+\beta_\ell} + \sum_{i=1}^{n} z_{i\ell} \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} - \sum_{i=1}^{n} z_{i\ell} y_i - g_\ell = 0, \quad \ell = 1, 2, \dots, c,$$

$$e^{\xi_j}\left(v_j + \sum_{i=1}^{n+1} \exp\{a_{ij} + k_{ij}\}\right) - (n+1) b_j - u_j = 0, \quad j = 1, 2, \dots, f,$$

$$e^{\eta_j}\left(s_j + e^{\lambda_j}\right) - d_j - r_j = 0, \quad j = 1, 2, \dots, f,$$

$$e^{\zeta_\ell}\left(w_\ell + e^{\beta_\ell}\right) - g_\ell - t_\ell = 0, \quad \ell = 1, 2, \dots, c.$$

Then, based on the solution m of this system, let us define

$$\tilde{A}_{ij} = \left(1 + e^{\tilde{\xi}_j}\right) \exp\{\tilde{a}_{ij} + k_{ij}\} + \tilde{\lambda}_j^2 \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\},$$
$$i = 1, 2, \dots, n \;;\; j = 1, 2, \dots, f,$$

$$\tilde{B}_{ijk} = \tilde{\lambda}_j \tilde{\lambda}_k \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, \quad i = 1, 2, \dots, n \;;\; j, k = 1, 2, \dots, f \;;\; j \neq k,$$

$$\tilde{C}_{ij} = \left(1 + \tilde{\lambda}_j \tilde{a}_{ij}\right) \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\} - y_i, \quad i = 1, 2, \dots, n \;;\; j = 1, 2, \dots, f,$$

$$\tilde{D}_{ijk} = \tilde{\lambda}_j \tilde{a}_{ik} \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, \quad i = 1, 2, \dots, n \;;\; j, k = 1, 2, \dots, f \;;\; j \neq k,$$

$$\tilde{E}_{ijk} = \tilde{\lambda}_j z_{ik} \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, i = 1, 2, \dots, n \;;\; j = 1, 2, \dots, f \;;\; k = 1, 2, \dots, c,$$

$$\tilde{F}_{ij} = \exp\left\{\tilde{\xi}_j + \tilde{a}_{ij} + k_{ij}\right\}, \quad i = 1, 2, \dots, n+1 \;;\; j = 1, 2, \dots, f,$$

$$\tilde{G}_j = \left(1 + e^{\tilde{\xi}_j}\right) \exp\left\{\tilde{a}_{n+1,j} + k_{n+1,j}\right\}, \quad j = 1, 2, \dots, f,$$

$$\tilde{H}_j = e^{\tilde{\eta}_j+\tilde{\lambda}_j} + \sum_{i=1}^{n} \tilde{a}_{ij}^2 \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, \quad j = 1, 2, \ldots, f,$$

$$\tilde{L}_{jk} = \sum_{i=1}^{n} \tilde{a}_{ij}\,\tilde{a}_{ik} \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, \quad j, k = 1, 2, \ldots, f \;;\; j \neq k, \tag{6.8}$$

$$\tilde{M}_{jk} = \sum_{i=1}^{n} \tilde{a}_{ij}\, z_{ik} \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, \quad j = 1, 2, \ldots, f \;;\; k = 1, 2, \ldots, c,$$

$$\tilde{P}_j = e^{\tilde{\eta}_j+\tilde{\lambda}_j}, \qquad \tilde{U}_j = e^{\tilde{\eta}_j}\left(s_j + e^{\tilde{\lambda}_j}\right), \qquad j = 1, 2, \ldots, f,$$

$$\tilde{Q}_\ell = e^{\tilde{\zeta}_\ell+\tilde{\beta}_\ell} + \sum_{i=1}^{n} z_{i\ell}^2 \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, \quad \ell = 1, 2, \ldots, c,$$

$$\tilde{R}_{\ell k} = \sum_{i=1}^{n} z_{i\ell} z_{ik} \exp\left\{\sum_{p=1}^{f} \tilde{\lambda}_p \tilde{a}_{ip} + \sum_{m=1}^{c} \tilde{\beta}_m z_{im}\right\}, \quad \ell, k = 1, 2, \ldots, c \;;\; \ell \neq k,$$

$$\tilde{S}_\ell = e^{\tilde{\zeta}_\ell+\tilde{\beta}_\ell}, \qquad \tilde{W}_\ell = e^{\tilde{\zeta}_\ell}\left(w_\ell + e^{\tilde{\beta}_\ell}\right), \qquad \ell = 1, 2, \ldots, c,$$

$$\tilde{T}_j = e^{\tilde{\xi}_j}\left(v_j + \sum_{i=1}^{n+1} \exp\left\{\tilde{a}_{ij} + k_{ij}\right\}\right), \quad j = 1, 2, \ldots, f.$$

If we now define the matrices

$$\tilde{\boldsymbol{A}}_j = diag\left(\tilde{A}_{ij}, i = 1, 2, \ldots, n\right), \quad j = 1, 2, \ldots, f,$$

$$\tilde{\boldsymbol{B}}_{ij} = diag\left(\tilde{B}_{kij}, k = 1, 2, \ldots, n\right), \quad i, j = 1, 2, \ldots, f \;;\; i \neq j,$$

$$\tilde{\boldsymbol{U}} = diag\left(\tilde{U}_j, j = 1, 2, \ldots, f\right), \qquad \tilde{\boldsymbol{W}} = diag\left(\tilde{W}_\ell, \ell = 1, 2, \ldots, c\right),$$

$$\tilde{\boldsymbol{G}} = diag\left(\tilde{G}_j, j = 1, 2, \ldots, f\right), \qquad \tilde{\boldsymbol{L}} = \begin{pmatrix} \tilde{H}_1 & \tilde{L}_{12} & \tilde{L}_{13} & \cdots & \tilde{L}_{1f} \\ \tilde{L}_{12} & \tilde{H}_2 & \tilde{L}_{23} & \cdots & \tilde{L}_{2f} \\ \tilde{L}_{13} & \tilde{L}_{23} & \tilde{H}_3 & \cdots & \tilde{L}_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{L}_{1f} & \tilde{L}_{2f} & \tilde{L}_{3f} & \cdots & \tilde{H}_f \end{pmatrix},$$

$$\tilde{M} = \begin{pmatrix} \tilde{M}_{11} & \tilde{M}_{12} & \tilde{M}_{13} & \cdots & \tilde{M}_{1c} \\ \tilde{M}_{21} & \tilde{M}_{22} & \tilde{M}_{23} & \cdots & \tilde{M}_{2c} \\ \tilde{M}_{31} & \tilde{M}_{32} & \tilde{M}_{33} & \cdots & \tilde{M}_{3c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{M}_{f1} & \tilde{M}_{f2} & \tilde{M}_{f3} & \cdots & \tilde{M}_{fc} \end{pmatrix}, \qquad \tilde{R} = \begin{pmatrix} \tilde{Q}_{1} & \tilde{R}_{12} & \tilde{R}_{13} & \cdots & \tilde{R}_{1c} \\ \tilde{R}_{12} & \tilde{Q}_{2} & \tilde{R}_{23} & \cdots & \tilde{R}_{2c} \\ \tilde{R}_{13} & \tilde{R}_{23} & \tilde{Q}_{3} & \cdots & \tilde{R}_{3c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{R}_{1c} & \tilde{R}_{2c} & \tilde{R}_{3c} & \cdots & \tilde{Q}_{c} \end{pmatrix},$$

$$\tilde{T} = diag\left(\tilde{T}_j,\ j = 1, 2, \ldots, f\right), \qquad \tilde{D} = diag\left(\tilde{F}_{n+1,j},\ j = 1, 2, \ldots, f\right),$$

$$\tilde{P} = diag\left(\tilde{P}_j,\ j = 1, 2, \ldots, f\right), \qquad \tilde{S} = diag\left(\tilde{S}_\ell,\ \ell = 1, 2, \ldots, c\right),$$

$$\tilde{E}_i = \begin{pmatrix} \tilde{E}_{1i1} & \tilde{E}_{2i1} & \tilde{E}_{3i1} & \cdots & \tilde{E}_{ni1} \\ \tilde{E}_{1i2} & \tilde{E}_{2i2} & \tilde{E}_{3i2} & \cdots & \tilde{E}_{ni2} \\ \tilde{E}_{1i3} & \tilde{E}_{2i3} & \tilde{E}_{3i3} & \cdots & \tilde{E}_{ni3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{E}_{1ic} & \tilde{E}_{2ic} & \tilde{E}_{3ic} & \cdots & \tilde{E}_{nic} \end{pmatrix}, \quad i = 1, 2, \ldots, f,$$

$$\tilde{F}_i = \begin{pmatrix} \tilde{F}_{11}\delta_{i1} & \tilde{F}_{21}\delta_{i1} & \tilde{F}_{31}\delta_{i1} & \cdots & \tilde{F}_{n1}\delta_{i1} \\ \tilde{F}_{12}\delta_{i2} & \tilde{F}_{22}\delta_{i2} & \tilde{F}_{32}\delta_{i2} & \cdots & \tilde{F}_{n2}\delta_{i2} \\ \tilde{F}_{13}\delta_{i3} & \tilde{F}_{23}\delta_{i3} & \tilde{F}_{33}\delta_{i3} & \cdots & \tilde{F}_{n3}\delta_{i3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{F}_{1f}\delta_{if} & \tilde{F}_{2f}\delta_{if} & \tilde{F}_{3f}\delta_{if} & \cdots & \tilde{F}_{nf}\delta_{if} \end{pmatrix}, \quad i = 1, 2, \ldots, f,$$

$$\tilde{C}_i = \begin{pmatrix} \tilde{C}_{1i}\delta_{i1} + \tilde{D}_{1i1}(1-\delta_{i1}) & \tilde{C}_{2i}\delta_{i1} + \tilde{D}_{2i1}(1-\delta_{i1}) & \cdots & \tilde{C}_{ni}\delta_{i1} + \tilde{D}_{ni1}(1-\delta_{i1}) \\ \tilde{C}_{1i}\delta_{i2} + \tilde{D}_{1i2}(1-\delta_{i2}) & \tilde{C}_{2i}\delta_{i2} + \tilde{D}_{2i2}(1-\delta_{i2}) & \cdots & \tilde{C}_{ni}\delta_{i2} + \tilde{D}_{ni2}(1-\delta_{i2}) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{C}_{1i}\delta_{if} + \tilde{D}_{1if}(1-\delta_{if}) & \tilde{C}_{2i}\delta_{if} + \tilde{D}_{2if}(1-\delta_{if}) & \cdots & \tilde{C}_{ni}\delta_{if} + \tilde{D}_{nif}(1-\delta_{if}) \end{pmatrix},$$

$$i = 1, 2, \ldots, f,$$

the inverse of the modal dispersion matrix can be written as

$$V^{-1} = \begin{pmatrix} \tilde{A}_1 & \tilde{B}_{12} & \cdots & \tilde{B}_{1f} & 0_{nf} & \tilde{C}_1^T & \tilde{E}_1^T & \tilde{F}_1^T & 0_{nf} & 0_{nc} \\ \tilde{B}_{12} & \tilde{A}_2 & \cdots & \tilde{B}_{2f} & 0_{nf} & \tilde{C}_2^T & \tilde{E}_2^T & \tilde{F}_2^T & 0_{nf} & 0_{nc} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{B}_{1f} & \tilde{B}_{2f} & \cdots & \tilde{A}_f & 0_{nf} & \tilde{C}_f^T & \tilde{E}_f^T & \tilde{F}_f^T & 0_{nf} & 0_{nc} \\ 0_{fn} & 0_{fn} & \cdots & 0_{fn} & \tilde{G} & 0_{ff} & 0_{fc} & \tilde{D} & 0_{ff} & 0_{fc} \\ \tilde{C}_1 & \tilde{C}_2 & \cdots & \tilde{C}_f & 0_{ff} & \tilde{L} & \tilde{M} & 0_{ff} & \tilde{P} & 0_{fc} \\ \tilde{E}_1 & \tilde{E}_2 & \cdots & \tilde{E}_f & 0_{cf} & \tilde{M}^T & \tilde{R} & 0_{cf} & 0_{cf} & \tilde{S} \\ \tilde{F}_1 & \tilde{F}_2 & \cdots & \tilde{F}_f & \tilde{D} & 0_{ff} & 0_{fc} & \tilde{T} & 0_{ff} & 0_{fc} \\ 0_{fn} & 0_{fn} & \cdots & 0_{fn} & 0_{ff} & \tilde{P} & 0_{fc} & 0_{ff} & \tilde{U} & 0_{fc} \\ 0_{cn} & 0_{cn} & \cdots & 0_{cn} & 0_{cf} & 0_{cf} & \tilde{S} & 0_{cf} & 0_{cf} & \tilde{W} \end{pmatrix}. \quad (6.9)$$

Therefore, we have that

$$p\left(\theta \mid D^n, d_{n+1}\right) \cong N_{(n+4)f+2c}(m, V).$$

Let us now consider a matrix

$$\tilde{H} = \begin{pmatrix} \tilde{A}_1 & \tilde{B}_{12} & \tilde{B}_{13} & \cdots & \tilde{B}_{1f} \\ \tilde{B}_{12} & \tilde{A}_2 & \tilde{B}_{23} & \cdots & \tilde{B}_{2f} \\ \tilde{B}_{13} & \tilde{B}_{23} & \tilde{A}_3 & \cdots & \tilde{B}_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{B}_{1f} & \tilde{B}_{2f} & \tilde{B}_{3f} & \cdots & \tilde{A}_f \end{pmatrix}$$

whose inverse $\tilde{H}^{-1}$, which must be evaluated numerically, is a symmetric matrix with diagonal blocks, that is,

$$\tilde{H}^{-1} = \begin{pmatrix} \tilde{H}_{11} & \tilde{H}_{12} & \tilde{H}_{13} & \cdots & \tilde{H}_{1f} \\ \tilde{H}_{12} & \tilde{H}_{22} & \tilde{H}_{23} & \cdots & \tilde{H}_{2f} \\ \tilde{H}_{13} & \tilde{H}_{23} & \tilde{H}_{33} & \cdots & \tilde{H}_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{H}_{1f} & \tilde{H}_{2f} & \tilde{H}_{3f} & \cdots & \tilde{H}_{ff} \end{pmatrix},$$

where each block is defined by

$$\tilde{H}_{ij} = diag\left(\tilde{H}_{kij}, k = 1, 2, \ldots, n\right), \quad i, j = 1, 2, \ldots, f.$$

After some algebra work we can obtain the approximate marginal posterior distribution of $\theta_p = \left(a_{n+1,1},\ a_{n+1,2},\ \ldots,\ a_{n+1,f},\ \lambda_1,\ \lambda_2,\ \ldots,\ \lambda_f,\ \beta_1,\ \beta_2,\ \ldots,\ \beta_c,\ \xi_1,\ \xi_2,\ \ldots,\ \xi_f\right)$, which is

$$p\left(\theta_p \mid D^n, d_{n+1}\right) \cong N_{3f+c}\left(m_p, V_p\right) \tag{6.10}$$

where $m_p = \tilde{\theta}_p$, whose components are obtained from (6.7) and

$$V_p = \left(V_{11} - V_{12} V_{22}^{-1} V_{12}^T\right)^{-1},$$

with

$$V_{11} = \begin{pmatrix} \tilde{G} & 0_{ff} & 0_{fc} & \tilde{D} \\ 0_{ff} & \tilde{L} & \tilde{M} & 0_{ff} \\ 0_{cf} & \tilde{M}^T & \tilde{R} & 0_{cf} \\ \tilde{D} & 0_{ff} & 0_{fc} & \tilde{T} \end{pmatrix}, \qquad V_{12} = \begin{pmatrix} 0_{ff} & 0_{fc} & 0_{fn} & 0_{fn} & 0_{fn} & \cdots & 0_{fn} \\ \tilde{P} & 0_{fc} & \tilde{C}_1 & \tilde{C}_2 & \tilde{C}_3 & \cdots & \tilde{C}_f \\ 0_{cf} & \tilde{S} & \tilde{E}_1 & \tilde{E}_2 & \tilde{E}_3 & \cdots & \tilde{E}_f \\ 0_{ff} & 0_{fc} & \tilde{F}_1 & \tilde{F}_2 & \tilde{F}_3 & \cdots & \tilde{F}_f \end{pmatrix}$$

and

$$\tilde{V}_{22} = \begin{pmatrix} \tilde{U} & 0_{fc} & 0_{fn} & 0_{fn} & 0_{fn} & \cdots & 0_{fn} \\ 0_{cf} & \tilde{W} & 0_{cn} & 0_{cn} & 0_{cn} & \cdots & 0_{cn} \\ 0_{nf} & 0_{nc} & \tilde{A}_1 & \tilde{B}_{12} & \tilde{B}_{13} & \cdots & \tilde{B}_{1f} \\ 0_{nf} & 0_{nc} & \tilde{B}_{12} & \tilde{A}_2 & \tilde{B}_{23} & \cdots & \tilde{B}_{2f} \\ 0_{nf} & 0_{nc} & \tilde{B}_{13} & \tilde{B}_{23} & \tilde{A}_3 & \cdots & \tilde{B}_{3f} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{nf} & 0_{nc} & \tilde{B}_{1f} & \tilde{B}_{2f} & \tilde{B}_{3f} & \cdots & \tilde{A}_f \end{pmatrix}.$$

Then, from (6.10) we easily derive the full conditional distributions required to perform the Gibbs routine. Defining the constants

$$\tilde{N}_p^{(1)} = \tilde{H}_p - \frac{\tilde{P}_p^2}{\tilde{U}_p} - \sum_{i=1}^{f} \sum_{j=1}^{f} \sum_{k=1}^{n} \left(\tilde{C}_{ki}\delta_{ip} + \tilde{D}_{kip}\left(1 - \delta_{ip}\right)\right)\left(\tilde{C}_{kj}\delta_{jp} + \tilde{D}_{kjp}\left(1 - \delta_{jp}\right)\right)\tilde{H}_{kij},$$

$$p = 1, 2, \ldots, f,$$

$$\tilde{N}^{(1)}_{pq} = \tilde{L}_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\left(\tilde{C}_{ki}\delta_{ip} + \tilde{D}_{kip}\left(1-\delta_{ip}\right)\right)\left(\tilde{C}_{kj}\delta_{jq} + \tilde{D}_{kjq}\left(1-\delta_{jq}\right)\right)\tilde{H}_{kij},$$

$$p, q = 1, 2, \ldots, f \;\; ; \;\; p \neq q,$$

$$\tilde{N}^{(2)}_{pq} = \tilde{M}_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\left(\tilde{C}_{ki}\delta_{ip} + \tilde{D}_{kip}\left(1-\delta_{ip}\right)\right)\tilde{E}_{kjq}\tilde{H}_{kij},$$

$$p = 1, 2, \ldots, f \;\; ; \;\; q = 1, 2, \ldots, c,$$

$$\tilde{N}^{(3)}_{pq} = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\left(\tilde{C}_{ki}\delta_{ip} + \tilde{D}_{kip}\left(1-\delta_{ip}\right)\right)\tilde{F}_{kq}\delta_{jq}\tilde{H}_{kij}, \quad p, q = 1, 2, \ldots, f,$$

$$\tilde{N}^{(4)}_{p} = \tilde{Q}_p - \frac{\tilde{S}^2_p}{\tilde{W}_p} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{E}_{kip}\tilde{E}_{kjp}\tilde{H}_{kij}, \quad p = 1, 2, \ldots, c,$$

$$\tilde{N}^{(4)}_{pq} = \tilde{R}_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{E}_{kip}\tilde{E}_{kjq}\tilde{H}_{kij}, \quad p, q = 1, 2, \ldots, c \;\; ; p \neq q,$$

$$\tilde{N}^{(5)}_{pq} = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{E}_{kip}\tilde{F}_{kq}\delta_{jq}\tilde{H}_{kij}, \quad p = 1, 2, \ldots, c \;\; ; q = 1, 2, \ldots, f,$$

$$\tilde{N}^{(6)}_{p} = \tilde{T}_p - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{F}^2_{kp}\delta_{ip}\delta_{jp}\tilde{H}_{kij}, \quad p = 1, 2, \ldots, f,$$

$$\tilde{N}^{(6)}_{pq} = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{F}_{kp}\delta_{ip}\tilde{F}_{kq}\delta_{jq}\tilde{H}_{kij}, \quad p, q = 1, 2, \ldots, f \;\; ; \;\; p \neq q,$$

the full conditional distributions are

$$p\left(a_{n+1,j} \mid \theta_p^{(a_{n+1,j})}, D^n, d_{n+1}\right) = N\left(\tilde{a}_{n+1,j} - \frac{\left(\xi_j - \tilde{\xi}_j\right)\tilde{F}_{n+1,j}}{\tilde{G}_j}, \frac{1}{\tilde{G}_j}\right), \quad j = 1, 2, \ldots, f,$$

$$p\left(\lambda_j \mid \theta_p^{(\lambda_j)}, D^n, d_{n+1}\right) = N\left(\tilde{\lambda}_j - \frac{\sum_{i=1}^{c}\left(\beta_i - \tilde{\beta}_i\right)\tilde{N}^{(2)}_{ji} + \sum_{i=1}^{f}\left(\xi_i - \tilde{\xi}_i\right)\tilde{N}^{(3)}_{ji}}{\tilde{N}^{(1)}_{j}} - \right.$$

$$\left. - \frac{\sum_{\substack{i=1 \\ i \neq j}}^{f} \left(\lambda_i - \tilde{\lambda}_i \right) \tilde{N}_{ji}^{(1)}}{\tilde{N}_j^{(1)}}, \frac{1}{\tilde{N}_j^{(1)}} \right), \quad j = 1, 2, \ldots, f,$$

$$p\left(\beta_\ell \mid \theta_p^{(\beta_\ell)}, D^n, d_{n+1} \right) = N \left(\tilde{\beta}_\ell - \frac{\sum_{i=1}^{f} \left(\lambda_i - \tilde{\lambda}_i \right) \tilde{N}_{i\ell}^{(2)} + \sum_{i=1}^{f} \left(\xi_i - \tilde{\xi}_i \right) \tilde{N}_{\ell i}^{(5)}}{\tilde{N}_\ell^{(4)}} - \right.$$

$$\left. - \frac{\sum_{\substack{i=1 \\ i \neq \ell}}^{c} \left(\beta_i - \tilde{\beta}_i \right) \tilde{N}_{\ell i}^{(4)}}{\tilde{N}_\ell^{(4)}}, \frac{1}{\tilde{N}_\ell^{(4)}} \right), \quad \ell = 1, 2, \ldots, c,$$

$$p\left(\xi_j \mid \theta_p^{(\xi_j)}, D^n, d_{n+1} \right) = N \left(\tilde{\xi}_j - \frac{\left(a_{n+1,j} - \tilde{a}_{n+1,j} \right) \tilde{F}_{n+1,j} + \sum_{i=1}^{f} \left(\lambda_i - \tilde{\lambda}_i \right) \tilde{N}_{ij}^{(3)}}{\tilde{N}_j^{(6)}} - \right.$$

$$\left. - \frac{\sum_{i=1}^{c} \left(\beta_i - \tilde{\beta}_i \right) \tilde{N}_{ij}^{(5)} + \sum_{\substack{i=1 \\ i \neq j}}^{f} \left(\xi_i - \tilde{\xi}_i \right) \tilde{N}_{ji}^{(6)}}{\tilde{N}_j^{(6)}}, \frac{1}{\tilde{N}_j^{(6)}} \right), \quad j = 1, 2, \ldots, f,$$

where $\theta_p^{(\theta_i)}$ represents the vector θ_p with the component θ_i removed.

6.2.4. Laplace Approximation

Let us define a function $h(\theta)$ satisfying (1.13), that is, such that

$$-n\,h(\theta) = -\sum_{i=1}^{n+1} \sum_{j=1}^{f} \left(1 + e^{\xi_j} \right) \exp\left\{ a_{ij} + k_{ij} \right\} - \sum_{i=1}^{n} \exp\left\{ \sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im} \right\} - \sum_{j=1}^{f} v_j e^{\xi_j} -$$

$$-\sum_{j=1}^{f} e^{\eta_j}\left(s_j + e^{\lambda_j}\right) - \sum_{\ell=1}^{c} e^{\zeta_\ell}\left(w_\ell + e^{\beta_\ell}\right) + \sum_{i=1}^{n+1}\sum_{j=1}^{f}\left(a_{ij} + k_{ij}\right)\left(b_j + x_{ij}\right) +$$

$$+\sum_{j=1}^{f}\left((n+1)b_j + u_j\right)\xi_j + \sum_{j=1}^{f}\eta_j\left(d_j + r_j\right) + \sum_{\ell=1}^{c}\zeta_\ell\left(g_\ell + t_\ell\right) +$$

$$+\sum_{j=1}^{f}\lambda_j\left(d_j + \sum_{i=1}^{n} a_{ij}y_i\right) + \sum_{\ell=1}^{c}\beta_\ell\left(g_\ell + \sum_{i=1}^{n} z_{i\ell}y_i\right).$$

Then, we define $\tilde{\theta}$ and $\tilde{\sigma}$ as in (1.16). Analysing the definition of $\tilde{\theta}$, it is obvious that $\tilde{\theta} = \boldsymbol{m}$, the posterior mode, solution of the system of equations (6.7). To evaluate $\tilde{\sigma}$, let us consider the constants defined in (6.8). It is clear that the matrix $n\nabla^2 h\left(\tilde{\theta}\right)$ is the matrix $\boldsymbol{V}^{-1}$ in (6.9). In order to evaluate its determinant we define now the matrices $\tilde{\boldsymbol{A}}_j$ $\left(j = 1, 2, \ldots, f\right)$ and $\tilde{\boldsymbol{B}}_{ij}$ $\left(i, j = 1, 2, \ldots, f\,; i \neq j\right)$ as being

$$\tilde{\boldsymbol{A}}_j = diag\left(\tilde{A}_{1j}, \tilde{A}_{2j}, \ldots, \tilde{A}_{nj}, \tilde{G}_j\right), \quad j = 1, 2, \ldots, f,$$

$$\tilde{\boldsymbol{B}}_{ij} = diag\left(\tilde{B}_{1ij}, \tilde{B}_{2ij}, \ldots, \tilde{B}_{nij}, 0\right), \quad i, j = 1, 2, \ldots, f \;\; ; \;\; i \neq j,$$

and we also define a matrix

$$\tilde{\boldsymbol{H}} = \begin{pmatrix} \tilde{\boldsymbol{A}}_1 & \tilde{\boldsymbol{B}}_{12} & \tilde{\boldsymbol{B}}_{13} & \cdots & \tilde{\boldsymbol{B}}_{1f} \\ \tilde{\boldsymbol{B}}_{12} & \tilde{\boldsymbol{A}}_2 & \tilde{\boldsymbol{B}}_{23} & \cdots & \tilde{\boldsymbol{B}}_{2f} \\ \tilde{\boldsymbol{B}}_{13} & \tilde{\boldsymbol{B}}_{23} & \tilde{\boldsymbol{A}}_3 & \cdots & \tilde{\boldsymbol{B}}_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{B}}_{1f} & \tilde{\boldsymbol{B}}_{2f} & \tilde{\boldsymbol{B}}_{3f} & \cdots & \tilde{\boldsymbol{A}}_f \end{pmatrix}, \tag{6.11}$$

and then we will have

$$\tilde{\sigma} = \left\{\left|\tilde{H}\right|\left|\tilde{N}\right|\prod_{i=1}^{f}\left\{\tilde{U}_i\right\}\prod_{i=1}^{c}\left\{\tilde{W}_i\right\}\right\}^{-1/2},$$

where $\tilde{N}$ is a $(2f + c) \times (2f + c)$ matrix which can be written in the partitioned form

$$\tilde{N} = \begin{pmatrix} \tilde{N}_{11} & \tilde{N}_{12} & \tilde{N}_{13} \\ \tilde{N}_{12}^T & \tilde{N}_{22} & \tilde{N}_{23} \\ \tilde{N}_{13}^T & \tilde{N}_{23}^T & \tilde{N}_{33} \end{pmatrix}$$

and where the components of such blocks are defined by

$$\tilde{N}_{11}(p,p) = \tilde{H}_p - \frac{\tilde{P}_p^2}{\tilde{U}_p} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\left(\tilde{C}_{kj}\delta_{jp} + \tilde{D}_{kjp}(1-\delta_{jp})\right)\left(\tilde{C}_{ki}\delta_{ip} + \tilde{D}_{kip}(1-\delta_{ip})\right)\tilde{H}_{kij},$$
$$p = 1, 2, \ldots, f,$$

$$\tilde{N}_{11}(p,q) = \tilde{L}_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\left(\tilde{C}_{kj}\delta_{jp} + \tilde{D}_{kjp}(1-\delta_{jp})\right)\left(\tilde{C}_{ki}\delta_{iq} + \tilde{D}_{kiq}(1-\delta_{iq})\right)\tilde{H}_{kij},$$
$$p, q = 1, 2, \ldots, f \;\; ; \;\; p \neq q,$$

$$\tilde{N}_{12}(p,q) = \tilde{M}_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\left(\tilde{C}_{kj}\delta_{jp} + \tilde{D}_{kjp}(1-\delta_{jp})\right)\tilde{E}_{kiq}\tilde{H}_{kij},$$
$$p = 1, 2, \ldots, f \;\; ; q = 1, 2, \ldots, c,$$

$$\tilde{N}_{13}(p,q) = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\left(\tilde{C}_{kj}\delta_{jp} + \tilde{D}_{kjp}(1-\delta_{jp})\right)\tilde{F}_{kq}\delta_{iq}\tilde{H}_{kij}, \quad p, q = 1, 2, \ldots, f,$$

$$\tilde{N}_{22}(p,p) = \tilde{Q}_p - \frac{\tilde{S}_p^2}{\tilde{W}_p} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{E}_{kjp}\tilde{E}_{kip}\tilde{H}_{kij}, \quad p = 1, 2, \ldots, c,$$

$$\tilde{N}_{22}(p,q) = \tilde{R}_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{E}_{kjp}\tilde{E}_{kiq}\tilde{H}_{kij}, \quad p, q = 1, 2, \ldots, c \;\; ; \;\; p \neq q,$$

$$\tilde{N}_{23}(p,q) = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n}\tilde{E}_{kjp}\tilde{F}_{kq}\delta_{iq}\tilde{H}_{kij}, \quad p = 1, 2, \ldots, c \;\; ; q = 1, 2, \ldots, f,$$

$$\tilde{N}_{33}(p,p) = \tilde{T}_p - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}\tilde{F}_{kp}^2\delta_{jp}\delta_{ip}\tilde{H}_{kij}, \quad p = 1, 2, \ldots, f,$$

$$\tilde{N}_{33}(p,q) = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}\tilde{F}_{kp}\delta_{jp}\tilde{F}_{kq}\delta_{iq}\tilde{H}_{kij}, \quad p, q = 1, 2, \ldots, f \;\; ; \;\; p \neq q.$$

The constants $\tilde{H}_{kij}$ $(k = 1, 2, ..., n+1 \;\; ; \;\; i, j = 1, 2, ..., f)$ are obtained inverting the matrix $\tilde{\boldsymbol{H}}$ defined in (6.11). Due to the special form of $\tilde{\boldsymbol{H}}$, its inverse will be of the form

$$\tilde{\boldsymbol{H}}^{-1} = \begin{pmatrix} \tilde{\boldsymbol{H}}_{11} & \tilde{\boldsymbol{H}}_{12} & \tilde{\boldsymbol{H}}_{13} & \cdots & \tilde{\boldsymbol{H}}_{1f} \\ \tilde{\boldsymbol{H}}_{12} & \tilde{\boldsymbol{H}}_{22} & \tilde{\boldsymbol{H}}_{23} & \cdots & \tilde{\boldsymbol{H}}_{2f} \\ \tilde{\boldsymbol{H}}_{13} & \tilde{\boldsymbol{H}}_{23} & \tilde{\boldsymbol{H}}_{33} & \cdots & \tilde{\boldsymbol{H}}_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{H}}_{1f} & \tilde{\boldsymbol{H}}_{2f} & \tilde{\boldsymbol{H}}_{3f} & \cdots & \tilde{\boldsymbol{H}}_{ff} \end{pmatrix},$$

where each block is defined as

$$\tilde{\boldsymbol{H}}_{ij} = diag\left(\tilde{H}_{kij}, k = 1, 2, ..., n+1\right), \quad i, j = 1, 2, ..., f.$$

The determinants of $\tilde{\boldsymbol{H}}$ and $\tilde{\boldsymbol{N}}$ must be evaluated through a numerical algorithm.

Now we define a function $h^*(\theta)$ as in (1.14), that is, such that

$$-n\,h^*(\theta) = -\sum_{i=1}^{n+1}\sum_{j=1}^{f}\left(1+e^{\xi_j}\right)\exp\left\{a_{ij}+k_{ij}\right\} - \sum_{i=1}^{n+1}\exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip} + \sum_{m=1}^{c}\beta_m z_{im}\right\} - \sum_{j=1}^{f} v_j e^{\xi_j} -$$

$$-\sum_{j=1}^{f} e^{\eta_j}\left(s_j + e^{\lambda_j}\right) - \sum_{t=1}^{c} e^{\zeta_t}\left(w_t + e^{\beta_t}\right) + \sum_{i=1}^{n+1}\sum_{j=1}^{f}\left(a_{ij}+k_{ij}\right)\left(b_j + x_{ij}\right) +$$

$$+\sum_{j=1}^{f}\left((n+1)b_j + u_j\right)\xi_j + \sum_{j=1}^{f}\eta_j\left(d_j + r_j\right) + \sum_{t=1}^{c}\zeta_t\left(g_t + t_t\right) +$$

$$+\sum_{j=1}^{f}\lambda_j\left(d_j + \sum_{i=1}^{n+1} a_{ij} y_i\right) + \sum_{t=1}^{c}\beta_t\left(g_t + \sum_{i=1}^{n+1} z_{it} y_i\right).$$

Then we define θ^* as in (1.16). θ^* will be obtained solving numerically the system formed by the $((n+4)f+2c)$ equations

$$\left(1+e^{\xi_j}\right)\exp\left\{a_{ij}+k_{ij}\right\} + \lambda_j \exp\left\{\sum_{p=1}^{f}\lambda_p a_{ip} + \sum_{m=1}^{c}\beta_m z_{im}\right\} - b_j - x_{ij} - \lambda_j y_i = 0,$$

$$i = 1, 2, ..., n+1 \;\; ; \;\; j = 1, 2, ..., f,$$

$$e^{\eta_j+\lambda_j} + \sum_{i=1}^{n+1} a_{ij} \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} - \sum_{i=1}^{n+1} a_{ij} y_i - d_j = 0, \quad j = 1, 2, \ldots, f,$$

$$e^{\zeta_\ell+\beta_\ell} + \sum_{i=1}^{n+1} z_{i\ell} \exp\left\{\sum_{p=1}^{f} \lambda_p a_{ip} + \sum_{m=1}^{c} \beta_m z_{im}\right\} - \sum_{i=1}^{n+1} z_{i\ell} y_i - g_\ell = 0, \quad \ell = 1, 2, \ldots, c,$$

$$e^{\xi_j}\left(v_j + \sum_{i=1}^{n+1} \exp\left\{a_{ij} + k_{ij}\right\}\right) - (n+1)b_j - u_j = 0, \quad j = 1, 2, \ldots, f,$$

$$e^{\eta_j}\left(s_j + e^{\lambda_j}\right) - d_j - r_j = 0, \quad j = 1, 2, \ldots, f,$$

$$e^{\zeta_\ell}\left(w_\ell + e^{\beta_\ell}\right) - g_\ell - t_\ell = 0, \quad \ell = 1, 2, \ldots, c.$$

Using the solution θ^* of this system of equations, we define the constants

$$A_{ij}^* = \left(1 + e^{\xi_j^*}\right) \exp\left\{a_{ij}^* + k_{ij}\right\} + \lambda_j^{*2} \exp\left\{\sum_{p=1}^{f} \lambda_p^* a_{ip}^* + \sum_{m=1}^{c} \beta_m^* z_{im}\right\},$$
$$i = 1, 2, \ldots, n+1 \;;\; j = 1, 2, \ldots, f,$$

$$B_{ijk}^* = \lambda_j^* \lambda_k^* \exp\left\{\sum_{p=1}^{f} \lambda_p^* a_{ip}^* + \sum_{m=1}^{c} \beta_m^* z_{im}\right\}, \; i = 1, 2, \ldots, n+1 \;;\; j, k = 1, 2, \ldots, f \;;\; j \neq k,$$

$$C_{ij}^* = \left(1 + \lambda_j^* a_{ij}^*\right) \exp\left\{\sum_{p=1}^{f} \lambda_p^* a_{ip}^* + \sum_{m=1}^{c} \beta_m^* z_{im}\right\}, \quad i = 1, 2, \ldots, n+1 \;;\; j = 1, 2, \ldots, f,$$

$$D_{ijk}^* = \lambda_j^* a_{ik}^* \exp\left\{\sum_{p=1}^{f} \lambda_p^* a_{ip}^* + \sum_{m=1}^{c} \beta_m^* z_{im}\right\}, \quad i = 1, 2, \ldots, n+1 \;;\; j, k = 1, 2, \ldots, f \;;\; j \neq k,$$

$$E_{ijk}^* = \lambda_j^* z_{ik} \exp\left\{\sum_{p=1}^{f} \lambda_p^* a_{ip}^* + \sum_{m=1}^{c} \beta_m^* z_{im}\right\},$$
$$i = 1, 2, \ldots, n+1 \;;\; j = 1, 2, \ldots, f \;;\; k = 1, 2, \ldots, c,$$

$$F_{ij}^* = \exp\left\{\xi_j^* + a_{ij}^* + k_{ij}\right\}, \quad i = 1, 2, \ldots, n+1 \;;\; j = 1, 2, \ldots, f,$$

$$H_j^* = e^{\eta_j^*+\lambda_j^*} + \sum_{i=1}^{n+1} a_{ij}^{*2} \exp\left\{\sum_{p=1}^{f} \lambda_p^* a_{ip}^* + \sum_{m=1}^{c} \beta_m^* z_{im}\right\}, \quad j = 1, 2, \ldots, f,$$

$$L^*_{jk} = \sum_{i=1}^{n+1} a^*_{ij} a^*_{ik} \exp\left\{\sum_{p=1}^{f} \lambda^*_p a^*_{ip} + \sum_{m=1}^{c} \beta^*_m z_{im}\right\}, \quad j, k = 1, 2, \ldots, f \ ; j \neq k,$$

$$M^*_{jk} = \sum_{i=1}^{n+1} a^*_{ij} z_{ik} \exp\left\{\sum_{p=1}^{f} \lambda^*_p a^*_{ip} + \sum_{m=1}^{c} \beta^*_m z_{im}\right\}, \quad j = 1, 2, \ldots, f \ ; k = 1, 2, \ldots, c,$$

$$P^*_j = e^{\eta^*_j + \lambda^*_j}, \qquad U^*_j = e^{\eta^*_j}\left(s_j + e^{\lambda^*_j}\right), \qquad j = 1, 2, \ldots, f,$$

$$Q^*_\ell = e^{\zeta^*_\ell + \beta^*_\ell} + \sum_{i=1}^{n+1} z^2_{i\ell} \exp\left\{\sum_{p=1}^{f} \lambda^*_p a^*_{ip} + \sum_{m=1}^{c} \beta^*_m z_{im}\right\}, \quad \ell = 1, 2, \ldots, c,$$

$$R^*_{\ell k} = \sum_{i=1}^{n+1} z_{i\ell} z_{ik} \exp\left\{\sum_{p=1}^{f} \lambda^*_p a^*_{ip} + \sum_{m=1}^{c} \beta^*_m z_{im}\right\}, \quad \ell, k = 1, 2, \ldots, c \ ; \ell \neq k,$$

$$S^*_\ell = e^{\zeta^*_\ell + \beta^*_\ell}, \ W^*_\ell = e^{\zeta^*_\ell}\left(w_\ell + e^{\beta^*_\ell}\right), \qquad \ell = 1, 2, \ldots, c,$$

$$T^*_j = e^{\xi^*_j}\left(v_j + \sum_{i=1}^{n+1} exp\left\{a^*_{ij} + k_{ij}\right\}\right), \quad j = 1, 2, \ldots, f,$$

and we also define the matrices

$$A^*_j = diag\left(A^*_{ij}, i = 1, 2, \ldots, n+1\right), \quad j = 1, 2, \ldots, f,$$

$$B^*_{ij} = diag\left(B^*_{kij}, k = 1, 2, \ldots, n+1\right), \quad i, j = 1, 2, \ldots, f \ ; \ i \neq j,$$

$$U^* = diag\left(U^*_j, j = 1, 2, \ldots, f\right), \qquad W^* = diag\left(W^*_\ell, \ell = 1, 2, \ldots, c\right),$$

$$L^* = \begin{pmatrix} H^*_1 & L^*_{12} & L^*_{13} & \cdots & L^*_{1f} \\ L^*_{12} & H^*_2 & L^*_{23} & \cdots & L^*_{2f} \\ L^*_{13} & L^*_{23} & H^*_3 & \cdots & L^*_{3f} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L^*_{1f} & L^*_{2f} & L^*_{3f} & \cdots & H^*_f \end{pmatrix}, \qquad M^* = \begin{pmatrix} M^*_{11} & M^*_{12} & M^*_{13} & \cdots & M^*_{1c} \\ M^*_{21} & M^*_{22} & M^*_{23} & \cdots & M^*_{2c} \\ M^*_{31} & M^*_{32} & M^*_{33} & \cdots & M^*_{2c} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ M^*_{f1} & M^*_{f2} & M^*_{f3} & \cdots & M^*_{fc} \end{pmatrix},$$

$$R^* = \begin{pmatrix} Q_1^* & R_{12}^* & R_{13}^* & \cdots & R_{1c}^* \\ R_{12}^* & Q_2^* & R_{23}^* & \cdots & R_{2c}^* \\ R_{13}^* & R_{23}^* & Q_3^* & \cdots & R_{3c}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{1c}^* & R_{2c}^* & R_{3c}^* & \cdots & Q_c^* \end{pmatrix}, \qquad T^* = diag\left(T_j^*, j = 1, 2, \ldots, f\right),$$

$$P^* = diag\left(P_j^*, j = 1, 2, \ldots, f\right), \qquad S^* = diag\left(S_\ell^*, \ell = 1, 2, \ldots, c\right),$$

$$E_j^* = \begin{pmatrix} E_{1j1}^* & E_{2j1}^* & E_{3j1}^* & \cdots & E_{nj1}^* & E_{n+1,j1}^* \\ E_{1j2}^* & E_{2j2}^* & E_{3j2}^* & \cdots & E_{nj2}^* & E_{n+1,j2}^* \\ E_{1j3}^* & E_{2j3}^* & E_{3j3}^* & \cdots & E_{nj3}^* & E_{n+1,j3}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ E_{1jc}^* & E_{2jc}^* & E_{3jc}^* & \cdots & E_{njc}^* & E_{n+1,jc}^* \end{pmatrix}, \quad j = 1, 2, \ldots, f,$$

$$F_j^* = \begin{pmatrix} F_{11}^*\delta_{j1} & F_{21}^*\delta_{j1} & F_{31}^*\delta_{j1} & \cdots & F_{n1}^*\delta_{j1} & F_{n+1,1}^*\delta_{j1} \\ F_{12}^*\delta_{j2} & F_{22}^*\delta_{j2} & F_{32}^*\delta_{j2} & \cdots & F_{n2}^*\delta_{j2} & F_{n+1,2}^*\delta_{j2} \\ F_{13}^*\delta_{j3} & F_{23}^*\delta_{j3} & F_{33}^*\delta_{j3} & \cdots & F_{n3}^*\delta_{j3} & F_{n+1,3}^*\delta_{j3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ F_{1f}^*\delta_{jf} & F_{2f}^*\delta_{jf} & F_{3f}^*\delta_{jf} & \cdots & F_{nf}^*\delta_{jf} & F_{n+1,f}^*\delta_{jf} \end{pmatrix}, \quad j = 1, 2, \ldots, f,$$

$$C_j^{*T} = \begin{pmatrix} C_{1j}^*\delta_{j1} + D_{1j1}^*(1-\delta_{j1}) & C_{1j}^*\delta_{j2} + D_{1j2}^*(1-\delta_{j2}) & \cdots & C_{1j}^*\delta_{jf} + D_{1jf}^*(1-\delta_{jf}) \\ C_{2j}^*\delta_{j1} + D_{2j1}^*(1-\delta_{j1}) & C_{2j}^*\delta_{j2} + D_{2j2}^*(1-\delta_{j2}) & \cdots & C_{2j}^*\delta_{jf} + D_{2jf}^*(1-\delta_{jf}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{nj}^*\delta_{j1} + D_{nj1}^*(1-\delta_{j1}) & C_{nj}^*\delta_{j2} + D_{nj2}^*(1-\delta_{j2}) & \cdots & C_{nj}^*\delta_{jf} + D_{njf}^*(1-\delta_{jf}) \\ C_{n+1,j}^*\delta_{j1} + D_{n+1,j1}^*(1-\delta_{j1}) & C_{n+1,j}^*\delta_{j2} + D_{n+1,j2}^*(1-\delta_{j2}) & \cdots & C_{n+1,j}^*\delta_{jf} + D_{n+1,jf}^*(1-\delta_{jf}) \end{pmatrix},$$

$$j = 1, 2, \ldots, f.$$

Using the matrices defined above as blocks, we build up a matrix whose determinant is equal to $\left| n\nabla^2 h^*(\theta^*) \right|$. In fact, such a matrix results from $n\nabla^2 h^*(\theta^*)$ permuting some rows and columns. Then, after some algebra work we obtain

$$\sigma^* = \left\{ |H^*||N^*| \prod_{i=1}^{f} \{U_i^*\} \prod_{i=1}^{c} \{W_i^*\} \right\}^{-1/2}. \tag{6.12}$$

The matrix H^* above is defined by

$$H^* = \begin{pmatrix} A_1^* & B_{12}^* & B_{13}^* & \cdots & B_{1f}^* \\ B_{12}^* & A_2^* & B_{23}^* & \cdots & B_{2f}^* \\ B_{13}^* & B_{23}^* & A_3^* & \cdots & B_{3f}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ B_{1f}^* & B_{2f}^* & B_{3f}^* & \cdots & A_f^* \end{pmatrix},$$

and its determinant must be evaluated numerically. Due to the special form of H^*, its inverse will be a symmetric matrix formed by diagonal blocks, that is,

$$H^{*-1} = \begin{pmatrix} H_{11}^* & H_{12}^* & H_{13}^* & \cdots & H_{1f}^* \\ H_{12}^* & H_{22}^* & H_{23}^* & \cdots & H_{2f}^* \\ H_{13}^* & H_{23}^* & H_{33}^* & \cdots & H_{3f}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_{1f}^* & H_{2f}^* & H_{3f}^* & \cdots & H_{ff}^* \end{pmatrix},$$

where

$$H_{ij}^* = diag\left(H_{kij}^*, k = 1, 2, \dots, n+1\right), \quad i, j = 1, 2, \dots, n.$$

The evaluation of H^{*-1} requires a numerical technique.

The matrix N^* in (6.12) is a $(2f+c)\times(2f+c)$ symmetric matrix of the form

$$N^* = \begin{pmatrix} N_{11}^* & N_{12}^* & N_{13}^* \\ N_{12}^{*T} & N_{22}^* & N_{23}^* \\ N_{13}^{*T} & N_{23}^{*T} & N_{33}^* \end{pmatrix},$$

and the elements of such a matrix are given by

$$N_{11}^*(p,p) = H_p^* - \frac{P_p^{*2}}{U_p^*} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}\left(C_{kj}^*\delta_{jp} + D_{kjp}^*\left(1-\delta_{jp}\right)\right)\left(C_{ki}^*\delta_{ip} + D_{kip}^*\left(1-\delta_{ip}\right)\right)H_{kij}^*,$$

$$p = 1, 2, \dots, f,$$

$$N_{11}^*(p,q) = L_{pq}^* - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}\left(C_{kj}^*\delta_{jp} + D_{kjp}^*\left(1-\delta_{jp}\right)\right)\left(C_{ki}^*\delta_{iq} + D_{kiq}^*\left(1-\delta_{iq}\right)\right)H_{kij}^*,$$

$$p, q = 1, 2, \dots, f \ ; \ p \neq q,$$

$$N^*_{12}(p,q) = M^*_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}\left(C^*_{kj}\delta_{jp} + D^*_{kjp}\left(1-\delta_{jp}\right)\right)E^*_{kiq}H^*_{kij},$$

$$p = 1, 2, \ldots, f \;\; ; \;\; q = 1, 2, \ldots, c,$$

$$N^*_{13}(p,q) = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}\left(C^*_{kj}\delta_{jp} + D^*_{kjp}\left(1-\delta_{jp}\right)\right)F^*_{kq}\delta_{iq}H^*_{kij}, \;\; p, q = 1, 2, \ldots, f,$$

$$N^*_{22}(p,p) = Q^*_p - \frac{S^{*2}_p}{W^*_p} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}E^*_{kjp}E^*_{kip}H^*_{kij}, \;\; p = 1, 2, \ldots, c,$$

$$N^*_{22}(p,q) = R^*_{pq} - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}E^*_{kjp}E^*_{kiq}H^*_{kij}, \;\; p, q = 1, 2, \ldots, c \;\; ; p \neq q,$$

$$N^*_{23}(p,q) = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}E^*_{kjp}F^*_{kq}\delta_{iq}H^*_{kij}, \;\; p = 1, 2, \ldots, c \;\; ; \;\; q = 1, 2, \ldots, f,$$

$$N^*_{33}(p,p) = T^*_p - \sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}F^{*2}_{kp}\delta_{jp}\delta_{ip}H^*_{kij}, \;\; p = 1, 2, \ldots, f,$$

$$N^*_{33}(p,q) = -\sum_{i=1}^{f}\sum_{j=1}^{f}\sum_{k=1}^{n+1}F^*_{kp}\delta_{jp}F^*_{kq}\delta_{iq}H^*_{kij}, \;\; p, q = 1, 2, \ldots, f \;\; ; \;\; p \neq q.$$

A numerical method is required to evaluate $|N^*|$.

Finally, the Laplace approximation for the predictive distribution of Y_{n+1} is given by

$$p\left(y_{n+1} \mid D^n, d_{n+1}\right) \propto \frac{1}{y_{n+1}!}\left(\frac{\sigma^*}{\tilde{\sigma}}\right)\exp\left\{-nh^*(\theta^*) + nh(\tilde{\theta})\right\}.$$

6.2.5. Example and Conclusions

A practical problem will now be considered. Values were recorded at $n = 78$ road junctions, and the data collected is presented in a table as follows:

i	x_{i1}	x_{i2}	y_i	z_{i1}	z_{i2}	z_{i3}	z_{i4}
1	69	104	5	0.0040000	11.4	2.96	0
2	76	93	6	0.0028571	6.1	2.30	-1
3	77	84	0	0.0071429	7.6	2.46	-2
4	113	101	4	0.0050000	17.6	1.26	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
78	73	55	0	0.0153846	9.0	2.89	2

The complete data set is shown in Table A2.1 (Appendix 2). For the i-th junction , two traffic flows and four covariates were considered:

x_{i1} - the first traffic flow, which was measured over a period of *15* minutes;

x_{i2} - the second traffic flow, measured over a period of *15* minutes;

y_i - the number of accidents which occurred at the junction over a period of *60* months;

z_{i1} - the entry path curvature (*1/metres*) of the junction;

z_{i2} - the entry width (*metres*) of the junction;

z_{i3} - the percentage of motorcycles at the junction (over the total flow $x_{i1} + x_{i2}$);

z_{i4} - the approach gradient which is measured in categories between *-3* and *+3* (*0* represents that the road is on level, negative values show that the road is downhill and positive values are used to indicate that it is uphill, and their absolute values characterise how steep the slope is).

Given a new road junction with an entry path curvature $z_{79,1} = 0.0012\ metres^{-1}$, an entry width $z_{79,2} = 10.4\ metres$, a percentage of motorcycles $z_{79,3} = 2.82\%$, an approach gradient $z_{79,4} = +1$ and with traffic countings $x_{79,1} = 73$ and $x_{79,2} = 116$, we aim to be able to make predictive statements about the number of accidents over a period of *60* months at that junction, that is, we want to derive the predictive distribution of Y_{79}, using the model developed in this chapter.

We assume $k_{ij} = 0, i = 1, 2, \ldots, 79$ and $j = 1, 2$, that is, we assume that the measurements of the traffic flows were made in the same conditions at all junctions.

6.2.5.1. First Stage Parameters

The choice of the first stage parameters b_1, b_2, d_1, d_2, g_1, g_2, g_3 and g_4 is undertaken as follows. Firstly, as in section 2.2.5, we estimate b_j by matching the first two marginal moments of the X_is, to obtain

$$b_j = \frac{\bar{x}_j^2}{s_j^2 - \bar{x}_j},$$

where

$$\bar{x}_j = \frac{1}{78}\sum_{i=1}^{78} x_{ij} \quad \text{and} \quad s_j^2 = \frac{1}{77}\sum_{i=1}^{78}\left(x_{ij} - \bar{x}_j\right)^2 \qquad (j = 1, 2).$$

Secondly, to choose d_j $(j = 1, 2)$ and g_ℓ $(\ell = 1, 2, 3, 4)$, we note that since

$$e^{\lambda_j} \sim Ga\left(d_j, e^{\eta_j}\right) \quad \text{and} \quad e^{\beta_\ell} \sim Ga\left(g_\ell, e^{\zeta_\ell}\right)$$

we have coefficients of variation given by

$$c.v.\left(e^{\lambda_j}\right) = d_j^{-1/2} \quad \text{and} \quad c.v.\left(e^{\beta_\ell}\right) = g_\ell^{-1/2}. \tag{6.13}$$

We have assumed that $X_{ij} \sim Po\left(\exp\left(a_{ij}\right)\right)$, so that simple estimates of a_{ij} are given by $a_{ij} = \ln\left(x_{ij}\right)$. Since

$$E(Y_i) = \exp\left\{\sum_{j=1}^{J} \lambda_j a_{ij} + \sum_{\ell=1}^{c} \beta_\ell z_{i\ell}\right\},$$

a multiple linear regression with the model

$$\ln(y_i) = \sum_{j=1}^{J} \lambda_j \ln\left(x_{ij}\right) + \sum_{\ell=1}^{c} \beta_\ell z_{i\ell} \tag{6.14}$$

provides the regression coefficients $\hat{\lambda}_j$ $(j = 1, 2)$ and $\hat{\beta}_\ell$ $(\ell = 1, 2, 3, 4)$ as well as their variances $V(\hat{\lambda}_j)$ and $V(\hat{\beta}_\ell)$. (In fact, we consider the model

$$\ln\left(y_i + 0.5\right) = \sum_{j=1}^{f} \lambda_j \ln\left(x_{ij}\right) + \sum_{\ell=1}^{c} \beta_\ell z_{i\ell}$$

instead of (6.14) in order to avoid problems with the junctions where no accidents were recorded). Considering a Taylor series for e^{λ_j}, we can write

$$e^{\lambda_j} \cong e^{\hat{\lambda}_j} + \left(\lambda_j - \hat{\lambda}_j\right) e^{\hat{\lambda}_j}$$

so that

$$E\left(e^{\lambda_j}\right) \cong e^{\hat{\lambda}_j} + \left(E\left(\lambda_j\right) - \hat{\lambda}_j\right) e^{\hat{\lambda}_j}$$

and

$$V\left(e^{\lambda_j}\right) \cong V\left(e^{\hat{\lambda}_j}\right) + V\left(\lambda_j e^{\hat{\lambda}_j}\right) + V\left(\hat{\lambda}_j e^{\hat{\lambda}_j}\right) = e^{2\hat{\lambda}_j} V\left(\lambda_j\right).$$

Estimating $E\left(\lambda_j\right)$ by $\hat{\lambda}_j$ and $V\left(\lambda_j\right)$ by $V\left(\hat{\lambda}_j\right)$, we obtain

$$c.v.\left(e^{\lambda_j}\right) \cong \sqrt{V\left(\hat{\lambda}_j\right)},$$

so that by matching it with $c.v.\left(e^{\lambda_j}\right)$ in (6.13), we obtain

$$d_j = \frac{1}{V\left(\hat{\lambda}_j\right)} \quad (j = 1, 2).$$

In a similar way, we take $g_\ell = \dfrac{1}{V\left(\hat{\beta}_\ell\right)}$, $(\ell = 1, 2, 3, 4)$.

6.2.5.2. Starting Values in Systems of Equations

When implementing both the estimative and approximate solutions of the problem we noticed that the systems of equations involved are very unstable. Therefore, good initial starting values for a_{ij}, λ_j, β_ℓ, ξ_j, η_j and ζ_ℓ $(i = 1, 2, \ldots, 79;\ j = 1, 2;\ \ell = 1, 2, 3, 4)$ must be given in order to achieve a solution. Since $X_{ij} \sim Po\left(\exp\left(a_{ij}\right)\right)$, we suggest taking

$$a_{ij}^{(0)} = \ln\left(x_{ij}\right),\quad i = 1, 2, \ldots, 79\ ;\quad j = 1, 2;$$

and $\lambda_j^{(0)}$ $(j = 1, 2)$ and $\beta_\ell^{(0)}$ $(\ell = 1, 2, 3, 4)$ as the coefficients obtained from the multiple linear regression on the model

$$\ln\left(y_i + 0.5\right) = \sum_{j=1}^{f} \lambda_j a_{ij}^{(0)} + \sum_{\ell=1}^{c} \beta_\ell z_{i\ell}.$$

Because

$$E\left(e^{\lambda_j}\right) = \frac{d_j}{e^{\eta_j}},\qquad E\left(e^{\beta_\ell}\right) = \frac{g_\ell}{e^{\zeta_\ell}}\qquad \text{and}\qquad E\left(e^{a_{ij}}\right) = \frac{b_j}{e^{\xi_j}},$$

we also suggest taking

$$\xi_j^{(0)} = \ln\left(79 b_j\right) - \ln\left(\sum_{i=1}^{79} \exp\left\{a_{ij}^{(0)}\right\}\right),\quad \eta_j^{(0)} = \ln\left(d_j\right) - \lambda_j^{(0)},\quad j = 1, 2,$$

$$\zeta_\ell^{(0)} = \ln\left(g_\ell\right) - \beta_\ell^{(0)},\ \ell = 1, 2, 3, 4.$$

Finally, we suggest the use of a multi-stage algorithm to solve the systems of equations involved in the problem.

6.2.5.3. Results

Figure 6.1. shows the plug-in estimate (MLE) for the predictive distribution of Y_{79}, its Laplace approximation and also the results obtained approximating the posterior distribution through multivariate normal distributions. Normal approximation 1 refers to the suggestion presented by Bernardo & Smith (1994) and normal approximation 2 refers to O'Hagan's (1994) suggestion. The Gibbs sampling approach developed in section 6.2.2 was not considered because it is expected to be somewhat inefficient due to the large

number of values we would have to generate. This implies that the necessary number of iterations in each cycle should be extremely large.

According to the criteria discussed earlier in section 6.2.5.1, we took the following values for the parameters of the distributions in the prior structure:

Prior 1

First stage:

$b_1 = 4.039$	$d_1 = 31.131$	$g_1 = 0.007$
$b_2 = 4.563$	$d_2 = 27.983$	$g_2 = 442.178$
		$g_3 = 78.103$
		$g_4 = 130.064$

and considered the vague prior case at the second stage.

Observation of Figure 6.1 reveals how similar the results are when the Laplace approximation and the normal approximation for the posterior distribution suggested by O'Hagan (1994) are used.

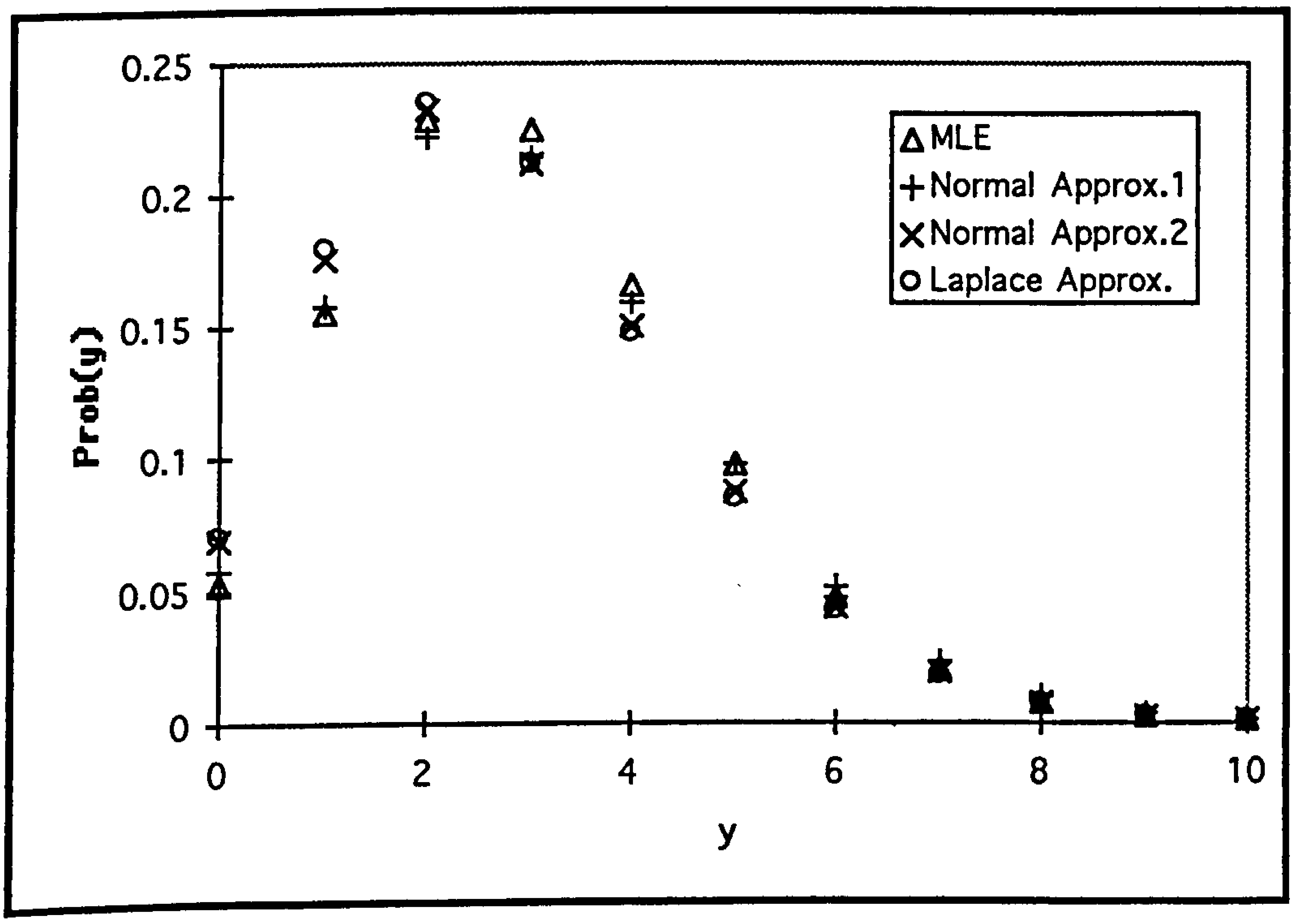

Figure 6.1: The predictive distribution of Y_{79}

We also notice that the results obtained using the normal approximation for the posterior distribution presented by Bernardo & Smith (1994) tend to diverge in the direction of the results obtained in the classical framework. These conclusions mirror the ones drawn from the examples presented in sections 2.2.5 and 3.3. When we studied those examples we also noticed that the Laplace approximation was an excellent approximation of the exact predictive distribution and that it was quite an efficient approach in the sense of the computational speed. Since now we are not able to evaluate the exact predictive distribution, we can only assume that the Laplace approximation yields very good results. This assumption is based on the conclusions drawn in earlier chapters. From now on, only the Laplace approximation will be considered.

6.2.5.4. Variations in Prior Assumptions

It would be interesting to see how the choice of the parameters of the prior distributions affect the resulting predictive distribution. We considered a range of different priors and present here a few examples.

Vague Second Stage Priors

For the case of vague second stage priors we found that the models were stable for changes in the first stage prior parameters of up to *15* to *25%*. For example, referring to the original prior in section 6.2.5.3 as Prior 1, the predictive distributions from the Laplace method for the following two priors are shown in Figure 6.2, and are indistinguishable.

Prior 2

First stage:

$b_1 = 5.87$	$d_1 = 27.6$	$g_1 = 0.004$
$b_2 = 3.98$	$d_2 = 31.01$	$g_2 = 440.67$
		$g_3 = 80.96$
		$g_4 = 129.75$

Prior 3

First stage:

$b_1 = 3.57$	$d_1 = 33.94$	$g_1 = 0.05$
$b_2 = 5.42$	$d_2 = 24.13$	$g_2 = 445.97$
		$g_3 = 74.38$
		$g_4 = 131.99$

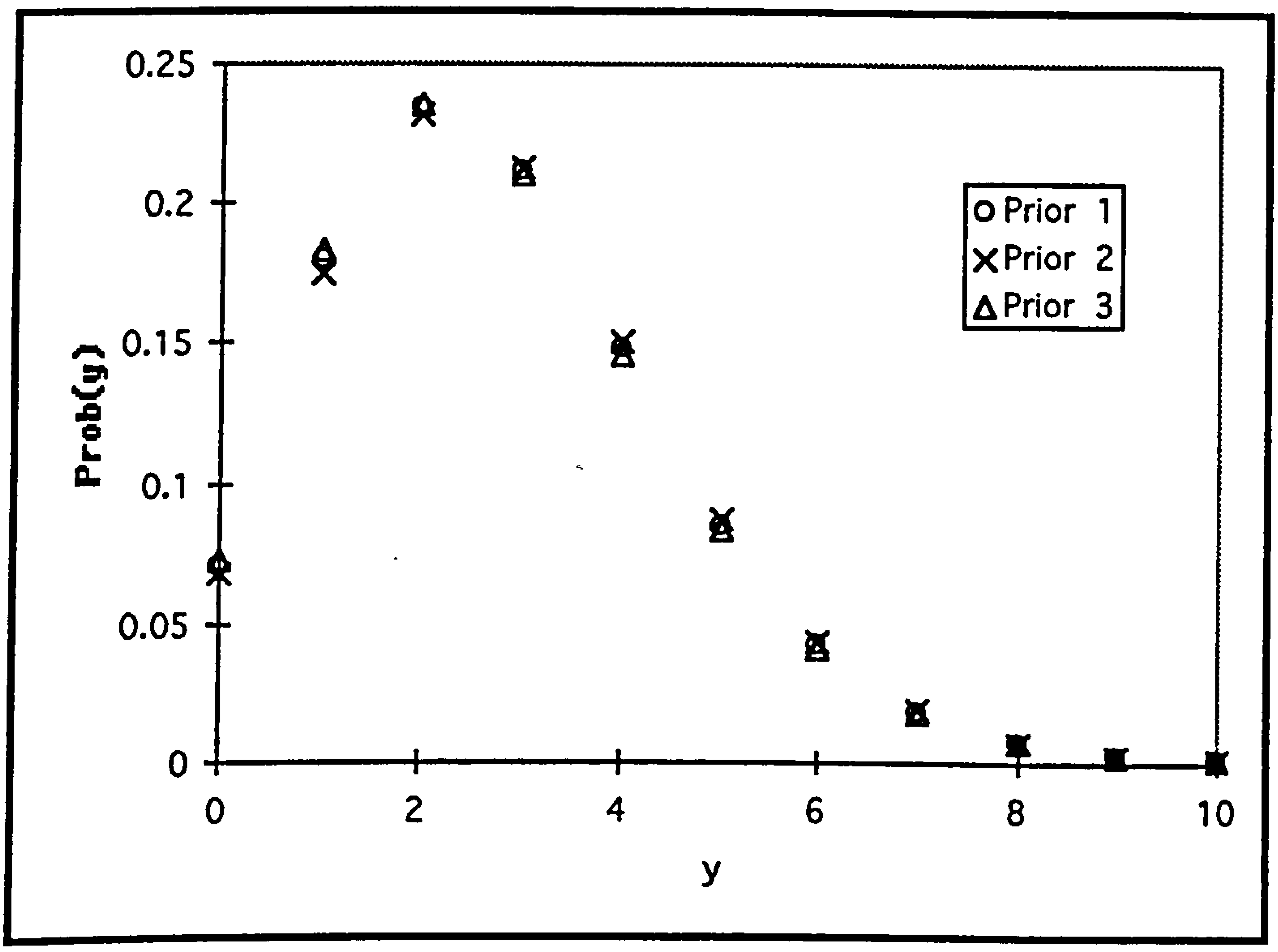

Figure 6.2: Comparison of the predictive distribution when small or moderate changes in the parameters of the prior structure are considered.

For a more drastic change, Priors 4 and 5 illustrate that some care is necessary in the specification of the first stage parameters, since as shown in Figure 6.3, variations in the predictive distribution become apparent.

Prior 4

First stage:

$b_1 = 8.07$	$d_1 = 40.92$	$g_1 = 0.1$
$b_2 = 2.14$	$d_2 = 12.75$	$g_2 = 92.51$
		$g_3 = 94.72$
		$g_4 = 60.91$

Prior 5

First stage:

$b_1 = 0.0000001$	$d_1 = 0.0000001$	$g_1 = 0.0000001$
$b_2 = 0.0000001$	$d_2 = 0.0000001$	$g_2 = 0.0000001$
		$g_3 = 0.0000001$
		$g_4 = 0.0000001$

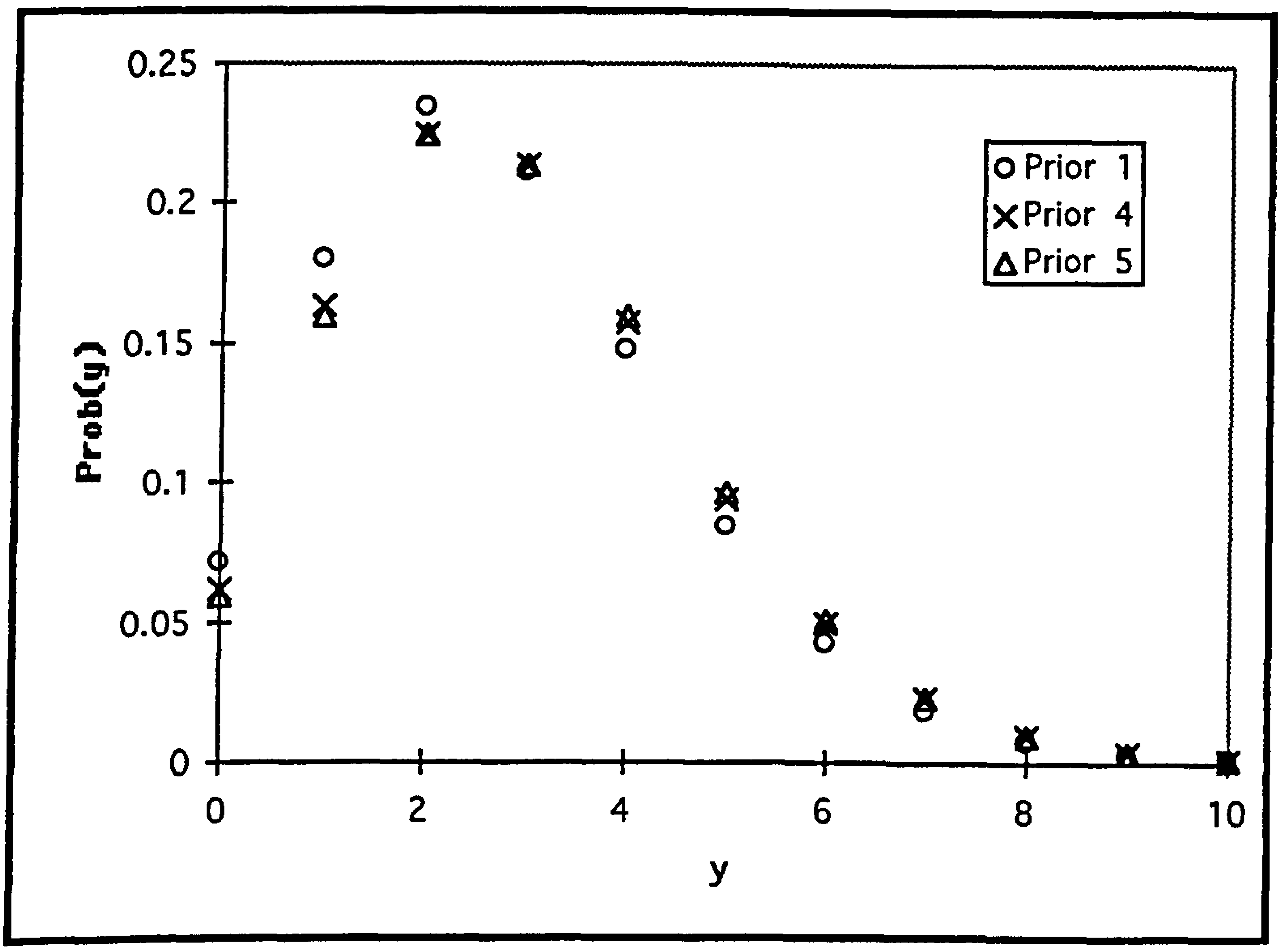

Figure 6.3: Comparison of the predictive distribution when larger changes in the parameters of the prior structure are considered.

General Prior

Moving away from a vague second stage prior assumption, Prior 6 illustrates the drastic changes which can result with differing prior assumptions, as shown in Figure 6.4.

Prior 6

First stage:

$b_1 = 4.039$	$d_1 = 31.131$	$g_1 = 0.007$
$b_2 = 4.563$	$d_2 = 27.983$	$g_2 = 442.178$
		$g_3 = 78.103$
		$g_4 = 130.064$

Second stage:

$u_1 = 1.42$	$v_1 = 2.03$	$r_1 = 0.04$	$s_1 = 1.57$	$t_1 = 4.72$	$w_1 = 2.51$
$u_2 = 0.75$	$v_2 = 1.12$	$r_2 = 3.56$	$s_2 = 2.01$	$t_2 = 0.07$	$w_2 = 4.31$
				$t_3 = 0.14$	$w_3 = 0.03$
				$t_4 = 1.37$	$w_4 = 3.04$

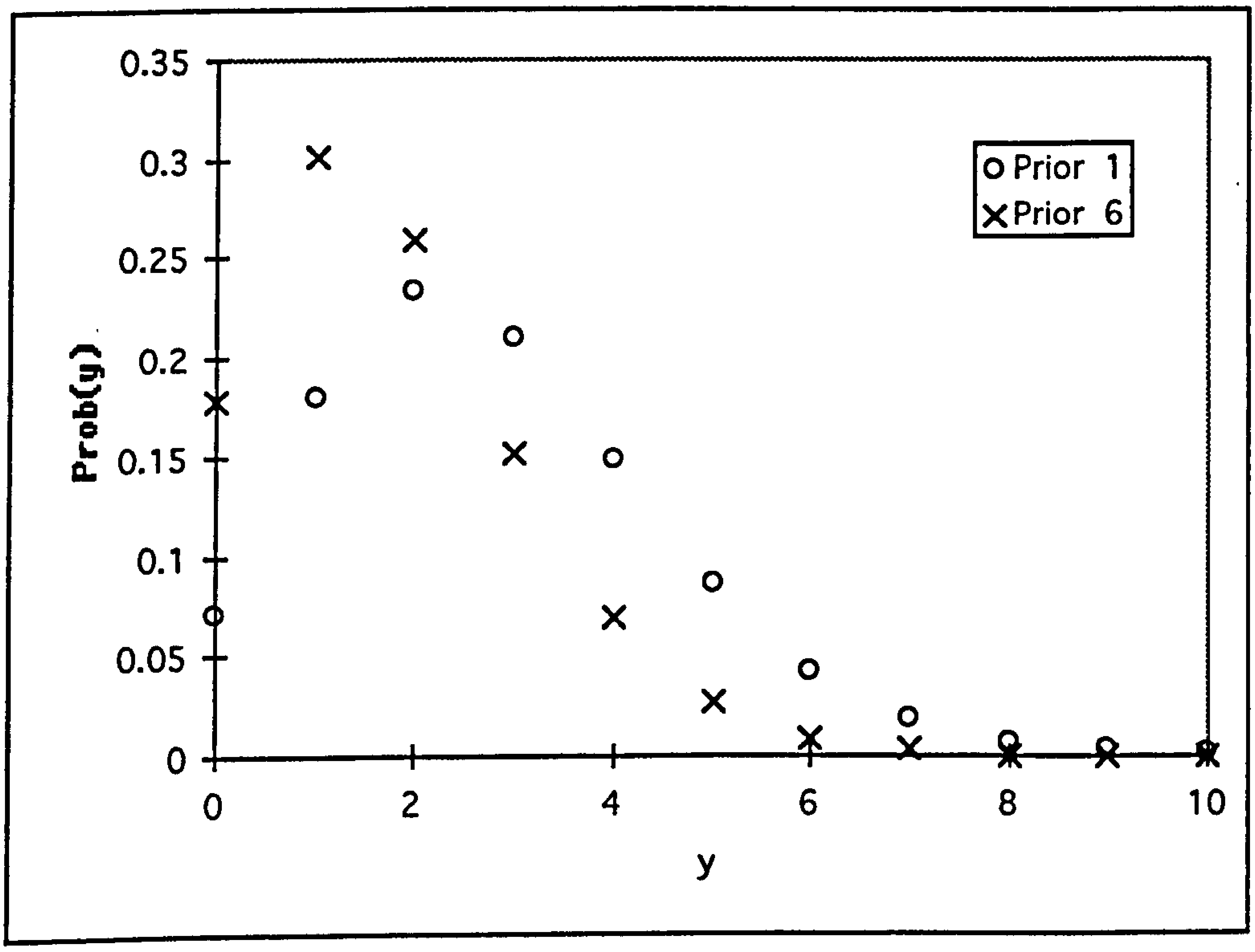

Figure 6.4: Comparison of the predictive distribution when a vague and a non-vague second stage prior structure are considered.

6.2.5.5. Effect of the Covariates

We now investigate the relative importance of the covariates for predicting the number of accidents. Figure 6.5 compares the predictive distribution obtained from using all four covariates, via the Laplace method (and referred to as model M4), and the one which just takes into consideration the two traffic flows ignoring the covariates, referred to as model M0. It is clear that the predictive distributions obtained using M0 and M4 are quite different. This suggests that the covariates, or at least some of them, make important contributions for the prediction of the number of accidents.

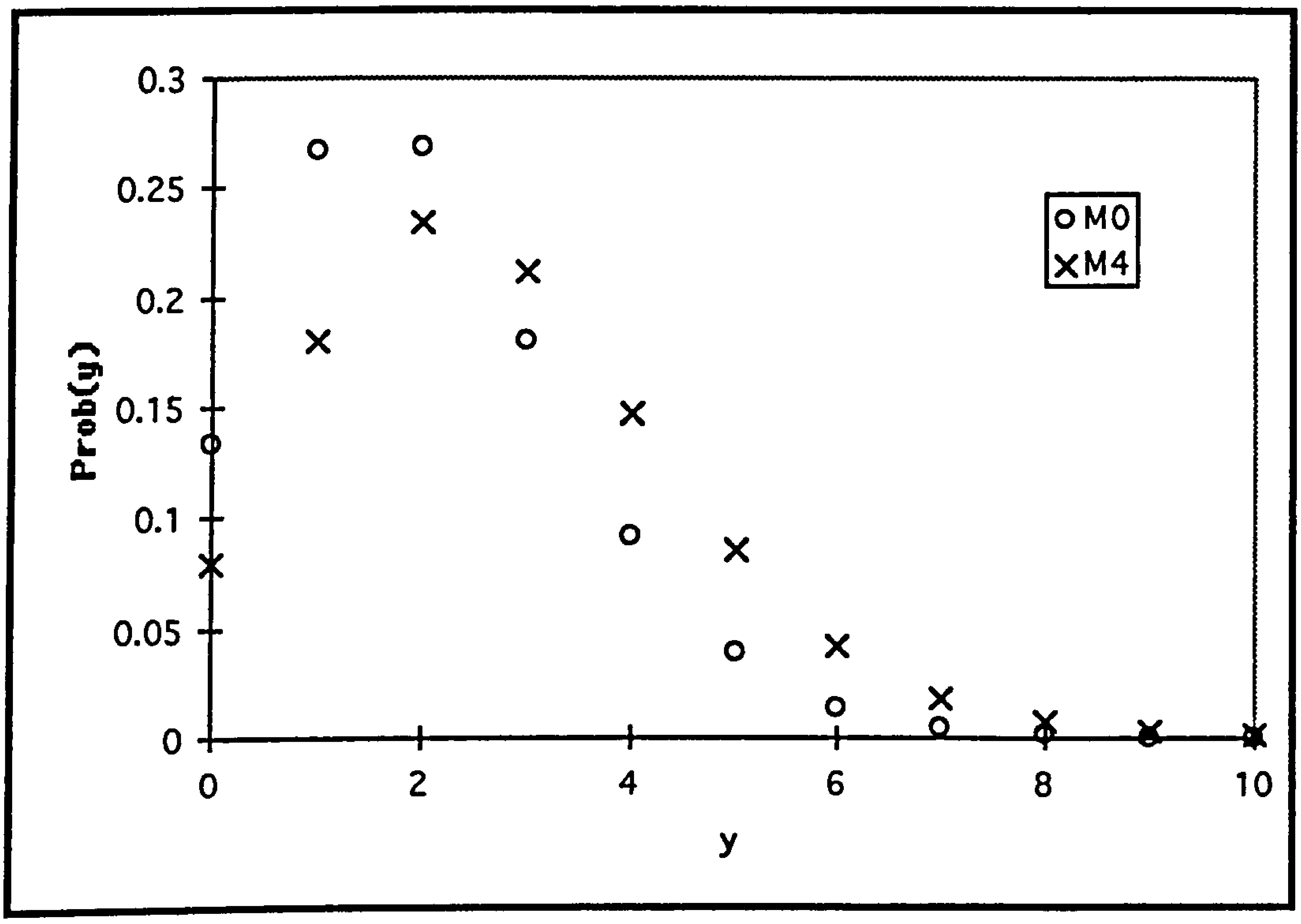

Figure 6.5: Comparison between models M0 and M4.

Stepwise Procedure

We consider a stepwise approach to building up the most suitable model, that is, the one which contains the smallest number of covariates but which still provides an adequate approximation to the full predictive distribution. Taking model M4 as being the initial model, we will remove one covariate at a time, in each case assessing the alteration of the predictive distribution. The covariate to be removed from the model is the one which least alters the predictive probabilities. To assess such alterations, we use the Kullback-Leibler measure of divergence (Kullback & Leibler (1951) and Aitchison (1975)), which is defined by

$$D(p_1(y), p_2(y)) = \sum_{y \in D_y} p_1(y) \ln\left(\frac{p_1(y)}{p_2(y)}\right),$$

where $p_1(y)$ and $p_2(y)$ are the distributions we want to compare and D_y is the domain of y. The reference distribution, that is, the one obtained with the best model, is $p_1(y)$. Here we take this to be the one from M4. The Kullback-Leibler measure of divergence between the two predictive distributions in M0 and M4 is $D=0.1222$.

Now, if we remove from model M4 one covariate at a time, the Kullback-Leibler measure of divergence assumes the values

	Removing z_1	Removing z_2	Removing z_3	Removing z_4
D	*0.1213*	*0.1685*	*0.0101*	*0.0110*

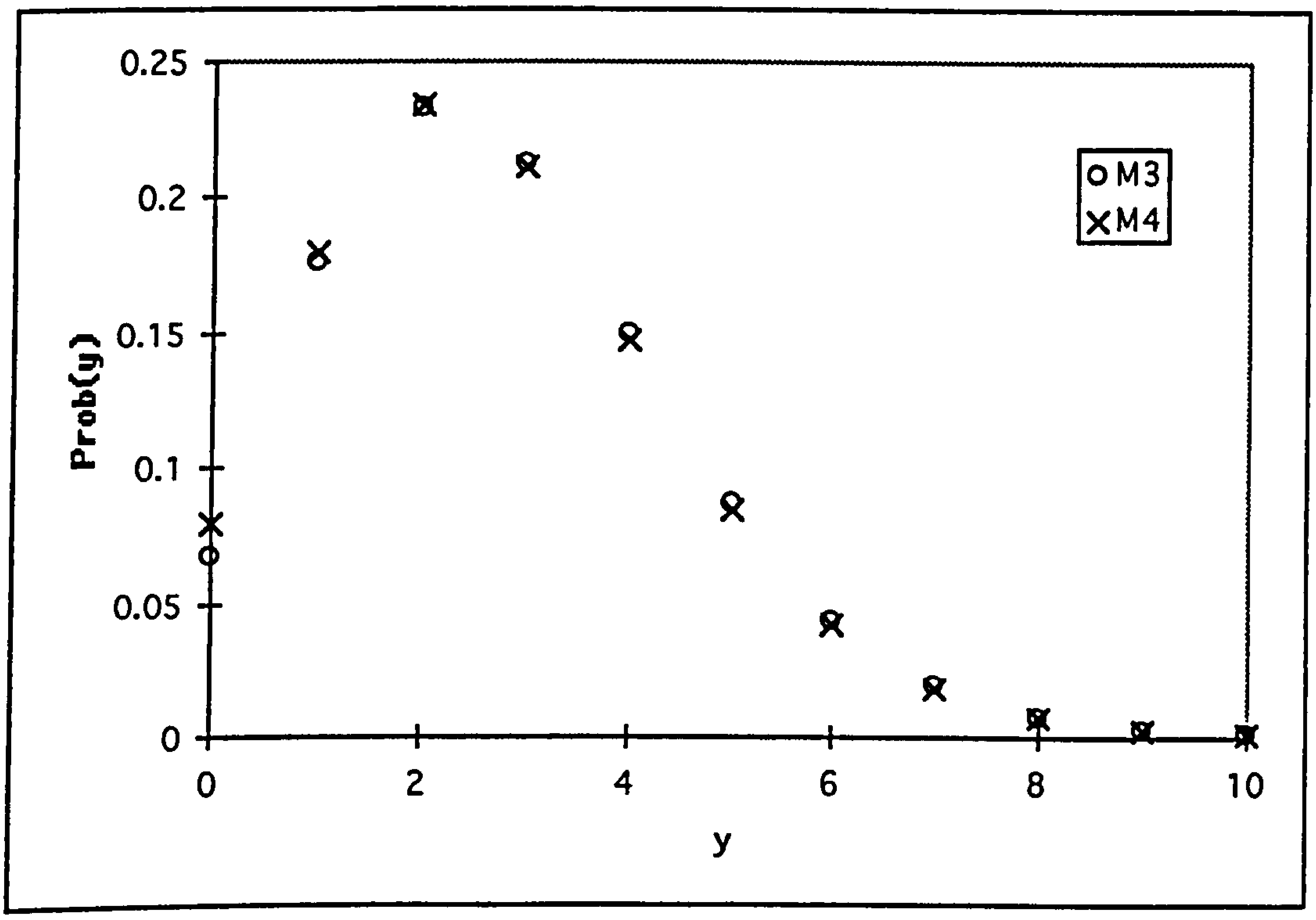

Figure 6.6: Comparison between models M3 and M4

Therefore we choose z_3 to be removed from the model since this appears to have the smallest effect, that is, its omission changes the predictive distribution least. Let us name the resulting model (which includes z_1, z_2 and z_4) by model M3. Figure 6.6 shows the predictive probabilities obtained through models M3 and M4.

It is clear that the two predictive distributions are very similar and therefore we conclude that we can discard z_3 from the model since the predictive probabilities are not too much affected.

We next consider discarding a further covariate. Thus we remove z_1, z_2 and z_4, one at a time, from model M3 and evaluate D, taking the predictive distribution from M4 as reference. The results are

	Removing z_1	Removing z_2	Removing z_4
D	0.1264	*0.1981*	*0.0122*

Therefore, we remove z_4 from the model, thus defining model M2 with covariates z_1 and z_2. The predictive distributions obtained with model M2 and M4 are plotted in Figure 6.7.

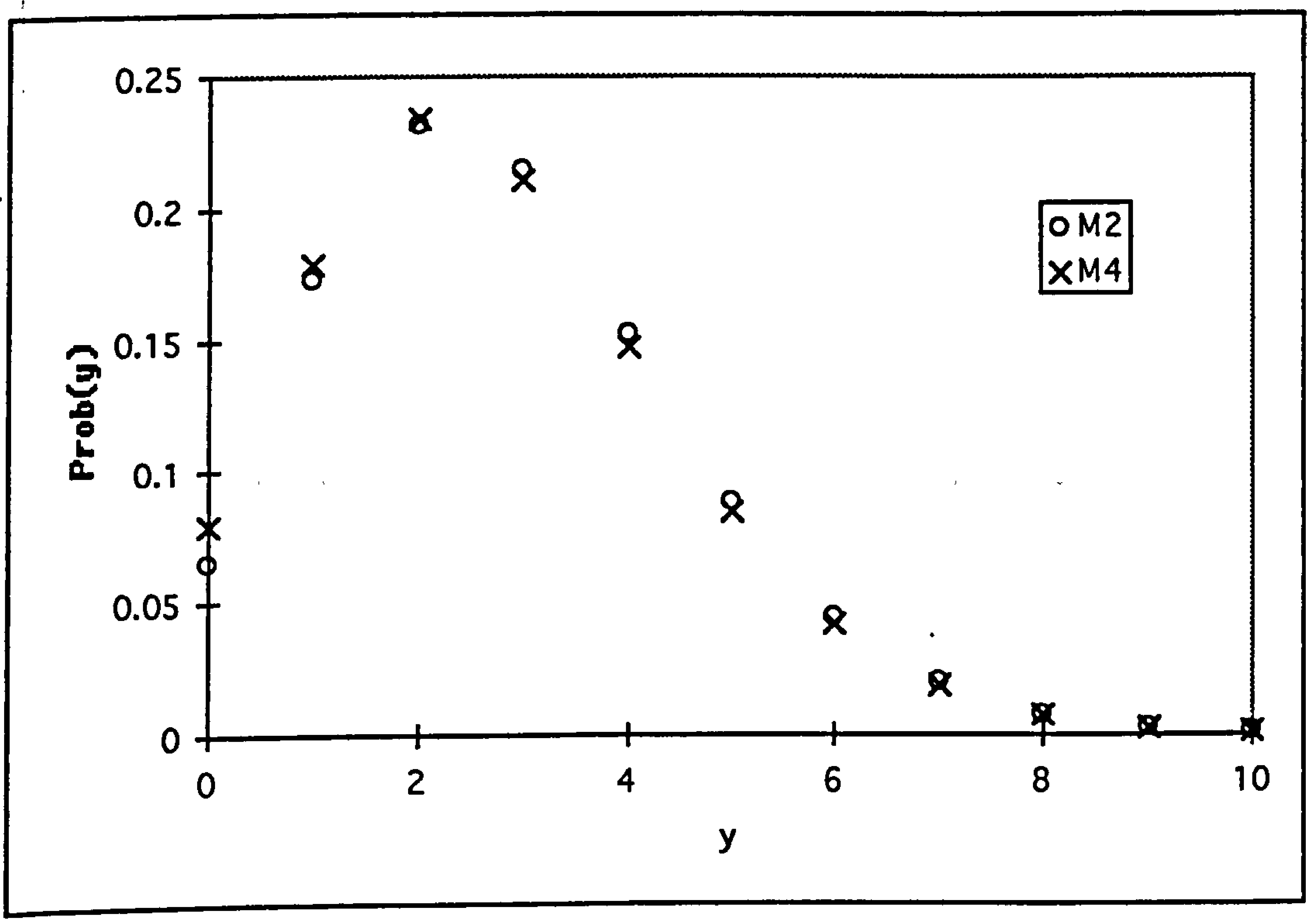

Figure 6.7: Comparison between models M2 and M4.

Thus, removing z_4 from the model does not decrease the accuracy of the approximate predictive distribution to any great extent.

A third covariate, z_1 or z_2, is then removed from the model and the Kullback-Leibler measure is evaluated as follows:

	Removing z_1	Removing z_2
D	*0.1275*	*0.1551*

Therefore, z_1 would be the next candidate for removal, but we note the large D values. Figure 6.8 shows the predictive distribution obtained using either M1 (with covariate z_2 alone) and M4.

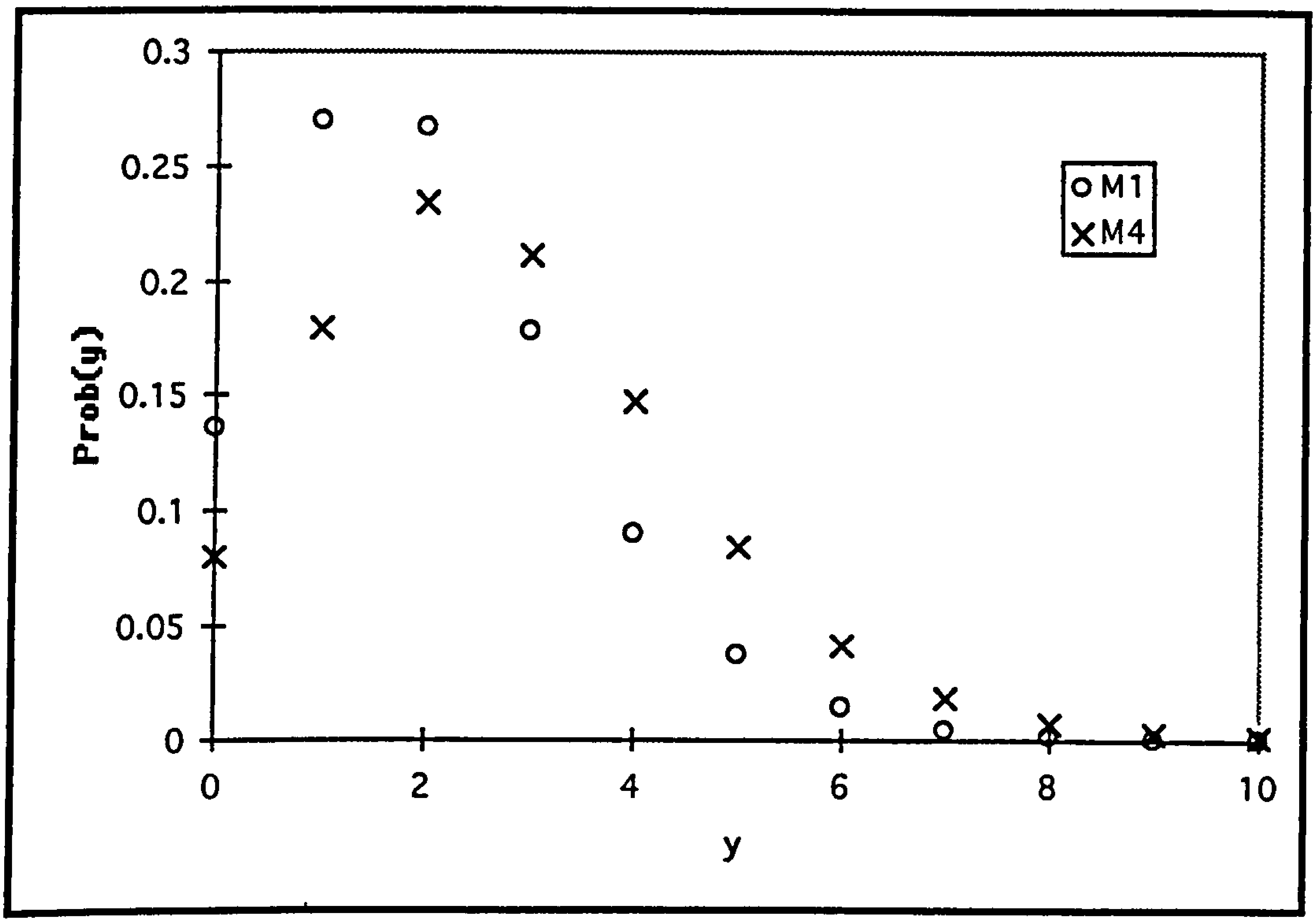

Figure 6.8: Comparison between models M1 and M4.

Observation of this figure shows that both predictive distributions are very different and therefore we cannot remove z_1, since this covariate seems to give an important contribution for predicting the number of accidents.

Therefore, we conclude that the entry path curvature (z_1) and the entry width of the junction (z_2) are important features that should be taken into consideration when we predict the number of accidents which will occur at the junction. On the other hand, the addition of the approach gradient (z_4) and the percentage of motorcycles (z_3) do not seem to alter the predictive distribution to any great extent, and so are felt to be less important.

All possible models procedure

Since we have only four covariates available, the number of possible models here is just *16*. Consequently, it is realistic to evaluate the importance of the covariates by analysing the resulting predictive distributions from each one of the possible models instead of following the stepwise approach considered earlier. Table 6.1 shows the possible models and the Kullback-Leibler divergence measure D, taking the full predictive model (M4) as reference.

Covariates in the model	Model	D
None	M0	0.1222
z_1	M1.1	0.1551
z_2	M1.2	0.1276
z_3	M1.3	0.1514
z_4	M1.4	0.0665
z_1, z_2	M2.12	0.0122
z_1, z_3	M2.13	0.1176
z_1, z_4	M2.14	0.1981
z_2, z_3	M2.23	0.1020
z_2, z_4	M2.24	0.1264
z_3, z_4	M2.34	0.0857
z_1, z_2, z_3	M3.123	0.0110
z_1, z_2, z_4	M3.124	0.0101
z_1, z_3, z_4	M3.134	0.1685
z_2, z_3, z_4	M3.234	0.1213

Table 6.1: Kullback-Leibler divergence measure D between the predictive distributions obtained with various models and the one obtained using model M4.

Note that the best models (M3.124, M3.123 and M2.12) coincide with the ones chosen via the stepwise approach. The next one to be considered would be model M1.4, a model which just takes into consideration the covariate z_4. Figure 6.9 compares the predictive distributions obtained using models M1.4 and M4, and the difference between them is noticeable. Therefore, we would conclude that the model using the covariates z_1

and z_2 is the one that should be considered. Note that this conclusion confirms the one drawn following a stepwise approach.

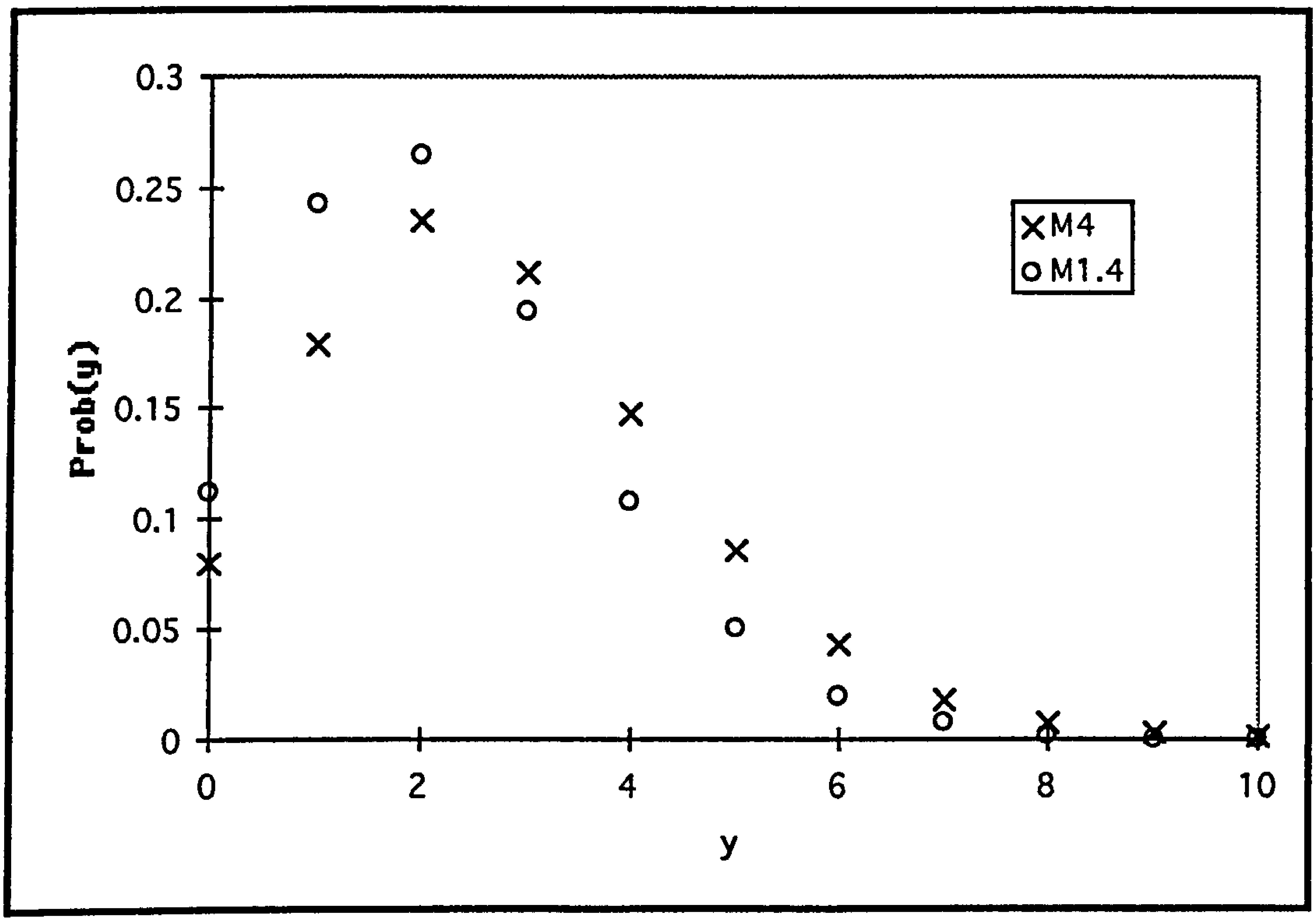

Figure 6.9: Comparison between models M4 and M1.4.

Calibration of D is very difficult, since it depends on both the data set and on the measurements of the traffic flows and covariates. Bhattacharjee and Dunsmore (1991, 1995) consider similar problems of calibration within the logistic and normal frameworks.

CHAPTER 7

PREDICTION IN SOME EXPONENTIAL ERRORS IN VARIABLES MODELS

The work developed in the Chapters 2, 3 and 4 motivated us to consider other elements of the exponential family as underlying distributions in the models. In this chapter we will assume that all observations come from exponential distributions. Because the methodology to be applied is the same, now we will present the results omitting the details of their derivation.

7.1. The Multiplicative Effect of a Treatment

Suppose we have independently distributed random variables (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n), where X_i and Y_i represent the observed measurement on the i-th individual before and after the treatment is applied, respectively. To model this situation, we assume that

- given θ_i, $X_i \sim Ex(\exp(\theta_i))$,
- given α and θ_i, $Y_i \sim Ex(\exp(\alpha + \theta_i))$,

where $Ex(\mu)$ represents the exponential distribution with mean $1/\mu$. While $\theta_1, \theta_2, ..., \theta_n$ are nuisance parameters used to model the particular characteristics of each individual, α is a parameter which models the multiplicative effect of the treatment upon the individuals.

Given a new individual with observed measurement x_{n+1}, we want to predict its outcome Y_{n+1}. We assume that, for the $(n+1)$-th individual,

$$X_{n+1} \sim Ex(\exp(\theta_{n+1})),$$
$$Y_{n+1} \sim Ex(\exp(\alpha + \theta_{n+1})).$$

The predictions for Y_{n+1} will be based on the data set

$$D^n = \{(x_i, y_i),\, i = 1, 2, \ldots, n\}$$

and on an observed x_{n+1}.

7.1.1. A Classical Approach

The likelihood function is

$$L(\theta^n, \theta_{n+1}, \alpha;\, D^n, x_{n+1}) = \exp\left\{-\sum_{i=1}^{n} e^{\theta_i}(x_i + e^{\alpha} y_i) - e^{\theta_{n+1}} x_{n+1} + n\alpha + 2\sum_{i=1}^{n} \theta_i + \theta_{n+1}\right\},$$

so that the maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n, \hat{\theta}_{n+1}$ are

$$\hat{\theta}_{n+1} = \ln\left\{\frac{1}{x_{n+1}}\right\}, \qquad \hat{\theta}_i = \ln\left\{\frac{2}{x_i + e^{\hat{\alpha}} y_i}\right\},\ i = 1, 2, \ldots, n,$$

and $\hat{\alpha}$ is obtained iteratively from

$$e^{\alpha}\sum_{i=1}^{n}\left\{\frac{y_i}{x_i + e^{\alpha} y_i}\right\} = \frac{n}{2}.$$

A simple plug-in estimate for the predictive distribution of Y_{n+1} would then be

$$p(y_{n+1} \mid D^n, x_{n+1}) = Ex\left(\exp\left(\hat{\alpha} + \hat{\theta}_{n+1}\right)\right).$$

Problems arise if $x_{n+1} = 0$.

7.1.2. A Bayesian Approach

Given the posterior distribution $p(\theta_{n+1}, \alpha \mid D^n, x_{n+1})$ for (θ_{n+1}, α), the predictive distribution of Y_{n+1} is given by

$$p(y_{n+1} \mid D^n, x_{n+1}) = \int_{\Re^2} p(y_{n+1} \mid \theta_{n+1}, \alpha)\, p(\theta_{n+1}, \alpha \mid D^n, x_{n+1})\, d\theta_{n+1}\, d\alpha. \qquad (7.1)$$

7.1.2.1. The Exact Predictive Distribution

To derive the exact predictive distribution of Y_{n+1}, given by (7.1), we will consider a hierarchical prior structure taking at the first stage,

$$p(\theta^n, \theta_{n+1}, \alpha \mid \xi, \eta) = \prod_{i=1}^{n+1} \{p(\theta_i \mid \xi)\}\, p(\alpha \mid \eta),$$

and at the second stage,

$$p(\xi, \eta) = p(\xi)\, p(\eta).$$

We will assume that

$$e^{\theta_i} \sim Ga(k, e^{\xi}) \qquad\qquad e^{\alpha} \sim Ga(g, e^{\eta})$$

$$e^{\xi} \sim Ga(u, v) \qquad\qquad e^{\eta} \sim Ga(r, s)$$

with k, g, u, v, r and s assumed to be known. The posterior distribution is given by

$$p(\theta^n, \theta_{n+1}, \alpha, \xi, \eta \mid D^n, x_{n+1}) \propto \exp\left\{-\sum_{i=1}^{n} e^{\theta_i}\left(x_i + e^{\alpha} y_i + e^{\xi}\right) - e^{\theta_{n+1}}\left(x_{n+1} + e^{\xi}\right)\right\} \times$$

$$\times \exp\left\{-e^{\eta+\alpha} - v e^{\xi} - s e^{\eta}\right\} \exp\left\{((n+1)k + u)\xi\right\} \times$$

$$\times \exp\left\{(2+k)\sum_{i=1}^{n} \theta_i + \theta_{n+1}(k+1) + (n+g)\alpha + (g+r)\eta\right\}.$$

After eliminating $\theta_1, \theta_2, ..., \theta_n$ and η, we obtain the marginal posterior distribution of $(\theta_{n+1}, \alpha, \xi)$ which is

$$p(\theta_{n+1}, \alpha, \xi \mid D^n, x_{n+1}) \propto \frac{\exp\left\{-e^{\theta_{n+1}}\left(x_{n+1} + e^{\xi}\right) - ve^{\xi}\right\}}{\left(e^{\alpha} + s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{((n+1)k + u)\xi + \theta_{n+1}(k+1) + (n+g)\alpha\right\}}{\prod_{i=1}^{n}\left\{\left(x_i + e^{\alpha}y_i + e^{\xi}\right)^{k+2}\right\}},$$

but we are unable to eliminate ξ. The exact predictive distribution of Y_{n+1} will then be given by

$$p(y_{n+1} \mid D^n, x_{n+1}) \propto \int_{\Re^2} \frac{\exp\left\{-ve^{\xi}\right\}\exp\left\{((n+1)k+u)\xi + (n+g+1)\alpha\right\} d\alpha\, d\xi}{\left(e^{\alpha} + s\right)^{g+r} \prod_{i=1}^{n+1}\left\{\left(x_i + e^{\alpha}y_i + e^{\xi}\right)^{k+2}\right\}},$$

whose evaluation requires a two-dimensional integration technique, which should be reasonably efficient.

No relevant simplifications occur when we consider vague second stage priors $(u, v, r, s \rightarrow 0)$.

7.1.2.2. Estimates and Approximations

We have developed the Gibbs sampling, normal approximations for the posterior distribution and Laplace methods as in previous chapters, but do not include the details here.

7.2. Treatments Effects in a Biased Allocation Model

Suppose we have $k \geq 2$ possible treatments, $T_1, T_2, ..., T_k$, all with effects that we assume to be of a multiplicative form. Let (X_i, Y_i), $i=1, 2, ..., n$, be independent random variables such that

- given θ_i, $X_i \sim Ex(\exp(\theta_i))$,
- given θ_i, $\alpha_1, \alpha_2, ..., \alpha_k$ and x_i, $Y_i \sim Ex(\exp(\alpha_j + \theta_i))$ if treatment T_j is used,
- given x_i, we assign treatment T_j if and only if $x_i \in C_j$,

where $C_1, C_2, ..., C_k$ are defined as in section 3.4.

Given a new individual with observed measurement x_{n+1}, we want to predict its outcome assuming that he will receive treatment T_j $(j=1, 2, ..., k)$. Such an outcome is represented by the random variable $Y_{n+1,j}$. We assume that

$$X_{n+1} \sim Ex(\exp(\theta_{n+1})),$$
$$Y_{n+1,j} \sim Ex(\exp(\alpha_j + \theta_{n+1})),\ j = 1, 2, ..., k.$$

We define a treatment indicator function

$$\delta_{ij} = \begin{cases} 1, & \text{if treatment } T_j \text{ is given to individual } i \\ 0, & \text{otherwise} \end{cases}$$

and n_j is the number of individuals who received treatment T_j, that is, $n_j = \sum_{i=1}^{n} \delta_{ij}$.

7.2.1. A Classical Approach

The likelihood function is

$$L(\theta^n, \theta_{n+1}, \alpha^k; D^n, x_{n+1}) = \exp\left\{-\sum_{i=1}^{n+1} e^{\theta_i} x_i - \sum_{i=1}^{n}\sum_{j=1}^{k} e^{\alpha_j + \theta_i} y_i \delta_{ij}\right\} \times$$

$$\times \exp\left\{\sum_{i=1}^{n+1}\theta_i + \sum_{i=1}^{n}\sum_{j=1}^{k}(\alpha_j + \theta_i)\delta_{ij}\right\},$$

which leads to the maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n, \hat{\theta}_{n+1}$, given by

$$\hat{\theta}_{n+1} = \ln\left(\frac{1}{x_{n+1}}\right), \qquad \hat{\theta}_i = \ln\left(\frac{2}{x_i + y_i \sum_{j=1}^{k} e^{\hat{\alpha}_j}\delta_{ij}}\right), \quad i = 1, 2, ..., n, \tag{7.2}$$

and $\hat{\alpha}_j$ $(j=1, 2, ..., k)$ as being the solutions of the equations

$$\sum_{i=1}^{n}\left\{\frac{y_i e^{\alpha_j}\delta_{ij}}{x_i + e^{\alpha_j} y_i \delta_{ij}}\right\} = \frac{n_j}{2}, \quad j = 1, 2, ..., k, \tag{7.3}$$

obtained through a numerical method. These maximum likelihood estimates can then, in a classical framework, be used to get plug-in estimates for the predictive distributions:

$$Y_{n+1,j} \sim Ex\left(\exp\left(\hat{\alpha}_j + \hat{\theta}_{n+1}\right)\right), \quad j = 1, 2, ..., k.$$

If $x_{n+1} = 0$ the implementation of this method is clearly impossible.

7.2.2. A Bayesian Approach

To derive the predictive distribution of $Y_{n+1,j}$, $j = 1, 2, ..., k$, in a Bayesian framework, we will consider a hierarchical prior structure for the parameters. Let us define the hyperparameter vector $\eta^k = (\eta_1, \eta_2, ..., \eta_k)$. At the first stage we take

$$p\left(\theta^n, \theta_{n+1}, \alpha^k \mid \xi, \eta^k\right) = \prod_{i=1}^{n+1}\left\{p(\theta_i \mid \xi)\right\}\prod_{j=1}^{k}\left\{p(\alpha_j \mid \eta_j)\right\},$$

and at the second stage we take

$$p\left(\xi, \eta^k\right) = p(\xi)\prod_{j=1}^{k}\left\{p\left(\eta_j\right)\right\}.$$

One possible appropriate choice for the prior distributions is

$$e^{\theta_i} \sim Ga\left(h, e^{\xi}\right), i = 1, 2, \ldots, n+1 \qquad e^{\alpha_j} \sim Ga\left(g_j, e^{\eta_j}\right), j = 1, 2, \ldots, k$$

$$e^{\xi} \sim Ga(u, v) \qquad e^{\eta_j} \sim Ga\left(r_j, s_j\right), j = 1, 2, \ldots, k$$

where h, g_j $(j=1, 2, \ldots, k), u, v, r_j$ $(j=1, 2, \ldots, k)$ and s_j $(j=1, 2, \ldots, k)$ are assumed to be known. The joint posterior distribution is given by

$$\begin{aligned} p\left(\theta^n, \theta_{n+1}, \alpha^k, \xi, \eta^k \mid D^n, x_{n+1}\right) \propto \exp&\left\{-\sum_{i=1}^{n+1} e^{\theta_i}\left(x_i + e^{\xi}\right) - \sum_{j=1}^{k} e^{\eta_j}\left(e^{\alpha_j} + s_j\right)\right\}\times \\ &\times\exp\left\{-ve^{\xi} - \sum_{i=1}^{n}\sum_{j=1}^{k} e^{\alpha_j+\theta_i} y_i \delta_{ij}\right\}\times \\ &\times\exp\left\{\sum_{i=1}^{n+1} h\left(\xi+\theta_i\right) + \sum_{j=1}^{k} g_j\left(\eta_j + \alpha_j\right) + u\,\xi\right\}\times \\ &\times\exp\left\{\sum_{j=1}^{k} r_j \eta_j + \sum_{i=1}^{n+1}\theta_i + \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\alpha_j + \theta_i\right)\delta_{ij}\right\}. \end{aligned} \quad (7.4)$$

7.2.2.1. The Exact Predictive Distributions

The exact predictive distribution of $Y_{n+1,j}$ is given by

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) = \int_{\Re^2} p\left(y_{n+1,j} \mid \theta_{n+1}, \alpha_j\right) p\left(\theta_{n+1}, \alpha_j \mid D^n, x_{n+1}\right) d\theta_{n+1} d\alpha_j.$$

From (7.4) we obtain the marginal posterior distribution for $\left(\theta_{n+1}, \alpha^k, \xi\right)$

$$p\left(\theta_{n+1}, \alpha^k, \xi \mid D^n, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}} + v + x_{n+1}\right)\right\}}{\prod_{i=1}^{n}\left\{\left(x_i + y_i \sum_{j=1}^{k} e^{\alpha_j}\delta_{ij} + e^{\xi}\right)^{h+2}\right\}} \times$$

$$\times \frac{\exp\left\{((n+1)h+u)\xi + \theta_{n+1}(1+h) + \sum_{j=1}^{k}\alpha_j\left(g_j + n_j\right)\right\}}{\prod_{j=1}^{k}\left\{\left(e^{\alpha_j} + s_j\right)^{g_j + r_j}\right\}}. \quad (7.5)$$

The predictive distribution of $Y_{n+1,j}$ will be given by

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) \propto \int_{\Re^{k+2}} \frac{\exp\left\{-e^{\theta_{n+1}}\left(x_{n+1} + y_{n+1,j}e^{\alpha_j} + e^{\xi}\right) - ve^{\xi}\right\}}{\prod_{i=1}^{n}\left\{\left(x_i + y_i\sum_{p=1}^{k} e^{\alpha_p}\delta_{ip} + e^{\xi}\right)^{h+2}\right\}} \times$$

$$\times \frac{\exp\left\{((n+1)h+u)\xi + \theta_{n+1}(2+h) + \alpha_j\left(1 + g_j + n_j\right)\right\}}{\prod_{p=1}^{k}\left\{\left(e^{\alpha_p} + s_p\right)^{g_p + r_p}\right\}} \times$$

$$\times \exp\left\{\sum_{\substack{p=1 \\ p\neq j}}^{k}\alpha_p\left(g_p + n_p\right)\right\} d\theta_{n+1}\, d\alpha^k\, d\xi, \quad (7.6)$$

which requires the evaluation of a $(k+2)$-dimensional integral, that must be solved numerically or approximations must be derived.

Vague second stage priors ($u, v \rightarrow 0$ and $r_j, s_j \rightarrow 0, j=1, 2, ..., k$) provide little relevant simplification.

7.2.2.2. Estimation Via Gibbs Sampling

In order to overcome the numerical problems that would appear evaluating the high-dimensional integral in (7.6), we can instead estimate the predictive distributions of $Y_{n+1,j}$

through the Gibbs sampling algorithm (section 1.2.1). The implementation of the Gibbs routine requires the full conditional distributions which are

$$p\left(\theta_{n+1} \mid \alpha^k, \xi, D^n, x_{n+1}\right) \propto \exp\left\{-e^{\theta_{n+1}}\left(x_{n+1} + e^{\xi}\right)\right\}\exp\left\{\theta_{n+1}(1+h)\right\},$$

$$p\left(\alpha_j \mid \theta_{n+1}, \alpha_{p\neq j}, \xi, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\alpha_j\left(g_j + n_j\right)\right\}}{\left(e^{\alpha_j} + s_j\right)^{g_j + r_j} \prod_{i=1}^{n}\left\{\left(x_i + y_i \sum_{p=1}^{k} e^{\alpha_p}\delta_{ip} + e^{\xi}\right)^{h+2}\right\}},$$

$$j=1, 2, \ldots, k,$$

$$p\left(\xi \mid \theta_{n+1}, \alpha^k, D^n, x_{n+1}\right) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}} + v\right)\right\}\exp\left\{\left((n+1)h + u\right)\xi\right\}}{\prod_{i=1}^{n}\left\{\left(x_i + y_i \sum_{p=1}^{k} e^{\alpha_p}\delta_{ip} + e^{\xi}\right)^{h+2}\right\}}.$$

From Table A1.1, we see that the full conditional distribution of θ_{n+1} is a transformed gamma distribution which makes the random generation of values of θ_{n+1} to be quite simple. The rejection sampling algorithm (section 1.2.3) is required to generate values of α_j ($j=1, 2, \ldots, k$) and ξ.

These distributions do not get much simpler when we consider the vague second stage priors case.

7.2.2.3. Estimation Via Asymptotic Results

Assuming that n is large enough, the posterior distribution can be approximated by a multivariate normal distribution, as shown in section 1.2.2.

7.2.2.3.1. Posterior Normality Based on the Likelihood Function

From the method in section 1.2.2.1 we find that θ_{n+1} and α_j ($j=1, 2, \ldots, k$) are independent a posteriori with distributions

$$p\left(\theta_{n+1} \mid D^n, x_{n+1}\right) = N\left(\hat{\theta}_{n+1}, 1\right),$$

$$p\left(\alpha_j \mid D^n, x_{n+1}\right) = N\left(\hat{\alpha}_j, \frac{1}{f_j}\right), \quad j = 1, 2, \ldots, k,$$

where $\hat{\theta}_{n+1}$ is given in (7.2), $\hat{\alpha}_j$ ($j=1, 2, \ldots, k$) are the solutions of the equations (7.3) and the constants f_j ($j=1, 2, \ldots, k$) are defined by

$$f_j = e^{\hat{\alpha}_j} \sum_{i=1}^{n} \frac{e^{\hat{\theta}_i} x_i y_i \delta_{ij}}{x_i + y_i e^{\hat{\alpha}_j} \delta_{ij}}, \quad j = 1, 2, \ldots, k.$$

Problems arise when $x_{n+1} = 0$ or $x_i = y_i = 0$ for any $i = 1, 2, \ldots, n$.

7.2.2.3.2. Posterior Normality Based on Characteristics of the Posterior Distribution

Given the posterior distribution (7.5) for $\left(\theta_{n+1}, \alpha^k, \xi\right)$ we can use the result suggested by O'Hagan (1994) and approximate it by

$$p\left(\theta_{n+1}, \alpha^k, \xi \mid D^n, x_{n+1}\right) \approx N_{k+2}(m, V), \tag{7.7}$$

according to (1.6). The first parameter of this normal distribution, $m = \left(\tilde{\theta}_{n+1}, \tilde{\alpha}^k, \tilde{\xi}\right)$, is the posterior mode and it is obtained iteratively from the system formed by the $(k+2)$ equations

$$e^{\theta_{n+1}}\left(e^{\xi} + x_{n+1}\right) - h - 1 = 0,$$

$$e^{\xi}\left(e^{\theta_{n+1}} + v\right) - (n+1)h - u + \sum_{i=1}^{n} \left\{ \frac{(h+2)e^{\xi}}{x_i + y_i \sum_{p=1}^{k} e^{\alpha_p} \delta_{ip} + e^{\xi}} \right\} = 0, \tag{7.8}$$

$$g_j + n_j - \sum_{i=1}^{n} \left\{ \frac{(h+2) e^{\alpha_j} y_i \delta_{ij}}{x_i + y_i e^{\alpha_j} \delta_{ij} + e^{\xi}} \right\} - \frac{(g_j + r_j) e^{\alpha_j}}{e^{\alpha_j} + s_j} = 0, \quad j = 1, 2, \ldots, k.$$

The second parameter of the normal distribution above is the modal dispersion matrix, defined as in (1.5).

If we define the constants

$$\tilde{b} = e^{\tilde{\theta}_{n+1} + \tilde{\xi}}, \qquad \tilde{a} = \tilde{b} + e^{\tilde{\theta}_{n+1}} x_{n+1},$$

$$\tilde{c}_j = (h+2) e^{\tilde{\alpha}_j} \sum_{i=1}^{n} \left\{ \frac{\left(x_i + e^{\tilde{\xi}}\right) \delta_{ij} y_i}{\left(x_i + y_i e^{\tilde{\alpha}_j} \delta_{ij} + e^{\tilde{\xi}}\right)^2} \right\} + \frac{(g_j + r_j) e^{\tilde{\alpha}_j} s_j}{\left(e^{\tilde{\alpha}_j} + s_j\right)^2}, \quad j = 1, 2, \ldots, k, \tag{7.9}$$

$$\tilde{d}_j = -(h+2) e^{\tilde{\alpha}_j + \tilde{\xi}} \sum_{i=1}^{n} \left\{ \frac{y_i \delta_{ij_i}}{\left(x_i + y_i e^{\tilde{\alpha}_j} \delta_{ij} + e^{\tilde{\xi}}\right)^2} \right\}, \quad j = 1, 2, \ldots, k,$$

$$\tilde{f} = e^{\tilde{\xi}} \left(e^{\tilde{\theta}_{n+1}} + v\right) + (h+2) e^{\tilde{\xi}} \sum_{i=1}^{n} \left\{ \frac{x_i + y_i \sum_{p=1}^{k} e^{\tilde{\alpha}_p} \delta_{ip}}{\left(x_i + y_i \sum_{p=1}^{k} e^{\tilde{\alpha}_p} \delta_{ip} + e^{\tilde{\xi}}\right)^2} \right\},$$

the approximated distribution (7.7) leads to the full conditional distributions

$$p\left(\theta_{n+1} \mid \alpha^k, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_{n+1} - \frac{\tilde{b}}{\tilde{a}}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{a}}\right),$$

$$p\left(\alpha_j \mid \theta_{n+1}, \alpha_{p \neq j}, \xi, D^n, x_{n+1}\right) = N\left(\tilde{\alpha}_j - \frac{\tilde{d}_j}{\tilde{c}_j}\left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{c}_j}\right), \quad j = 1, 2, \ldots, k,$$

$$p\left(\xi \mid \theta_{n+1}, \alpha^k, D^n, x_{n+1}\right) = N\left(\tilde{\xi} - \frac{\left(\theta_{n+1} - \tilde{\theta}_{n+1}\right)\tilde{b} + \sum_{j=1}^{k}\left(\alpha_j - \tilde{\alpha}_j\right)\tilde{d}_j}{\tilde{f}}, \frac{1}{\tilde{f}}\right).$$

The vague second stage priors case again do not lead to relevant simplifications.

7.2.2.4. Laplace Approximation

To approximate (7.6) via the Laplace approximation (section 1.2.4), we define two functions, $h(\theta_{n+1}, \alpha^k, \xi)$ and $h^*(\theta_{n+1}, \alpha^k, \xi)$ such that

$$-n h(\theta_{n+1}, \alpha^k, \xi) = -e^{\xi+\theta_{n+1}} - v e^{\xi} - x_{n+1} e^{\theta_{n+1}} + ((n+1)h + u)\xi + \theta_{n+1}(1+h) +$$

$$+\sum_{j=1}^{k} \alpha_j (g_j + n_j) - \sum_{i=1}^{n} \left\{ (h+2)\ln\left(x_i + y_i \sum_{j=1}^{k} e^{\alpha_j} \delta_{ij} + e^{\xi} \right) \right\} -$$

$$-\sum_{j=1}^{k} \left\{ (g_j + r_j) \ln\left(e^{\alpha_j} + s_j\right) \right\}$$

and

$$-n h^*(\theta_{n+1}, \alpha^k, \xi) = -e^{\theta_{n+1}+\alpha_j} y_{n+1,j} + \theta_{n+1} + \alpha_j - e^{\xi+\theta_{n+1}} - v e^{\xi} - x_{n+1} e^{\theta_{n+1}} +$$

$$+((n+1)h + u)\xi + \theta_{n+1}(1+h) + \sum_{p=1}^{k} \alpha_p (g_p + n_p) -$$

$$-\sum_{p=1}^{k} \left\{ (g_p + r_p) \ln\left(e^{\alpha_p} + s_p\right) \right\} - \sum_{i=1}^{n} \left\{ (h+2)\ln\left(x_i + y_i \sum_{p=1}^{k} e^{\alpha_p} \delta_{ip} + e^{\xi} \right) \right\}.$$

Defining $(\tilde{\theta}_{n+1}, \tilde{\alpha}^k, \tilde{\xi})$, $\tilde{\sigma}$, $(\theta^*_{n+1}, \alpha^{k*}, \xi^*)$ and σ^* such that (1.16) holds, we conclude that $(\tilde{\theta}_{n+1}, \tilde{\alpha}^k, \tilde{\xi})$=*m*, the posterior mode, the solution of the system of equations (7.8); $\tilde{\sigma}$ is given by

$$\tilde{\sigma} = \left\{ \tilde{a} \left(\frac{\tilde{b}^2}{\tilde{a}} + \sum_{m=1}^{k} \frac{\tilde{d}_m^2}{\tilde{c}_m} \right) \prod_{p=1}^{k} \{\tilde{c}_p\} \right\}^{-1/2},$$

where the constants are defined in (7.9); $(\theta^*_{n+1}, \alpha^{k*}, \xi^*)$ is the solution of the system formed by the equations

$$e^{\theta_{n+1}}\left(x_{n+1} + e^{\alpha_j} y_{n+1,j} + e^{\xi}\right) - h - 2 = 0,$$

$$g_j + n_j + 1 - y_{n+1,j}\, e^{\theta_{n+1}+\alpha_j} - \frac{(g_j + r_j)e^{\alpha_j}}{e^{\alpha_j} + s_j} - (h+2)e^{\alpha_j} \sum_{i=1}^{n} \left\{ \frac{y_i \delta_{ij}}{x_i + y_i e^{\alpha_j}\delta_{ij} + e^{\xi}} \right\} = 0,$$

$$g_p + n_p - \frac{(g_p + r_p)e^{\alpha_p}}{e^{\alpha_p} + s_p} - (h+2)e^{\alpha_p} \sum_{i=1}^{n} \left\{ \frac{y_i \delta_{ip}}{x_i + y_i e^{\alpha_p}\delta_{ip} + e^{\xi}} \right\} = 0\,,\ p=1, 2, \ldots, k,\ p \neq j,$$

$$e^{\xi}\left(e^{\theta_{n+1}} + v\right) - (n+1)h - u + \sum_{i=1}^{n} \left\{ \frac{(h+2)e^{\xi}}{x_i + y_i \sum_{p=1}^{k} e^{\alpha_p}\delta_{ip} + e^{\xi}} \right\} = 0,$$

which must be obtained numerically; σ^* is given by

$$\sigma^* = \left\{ q^* \prod_{\substack{p=1 \\ p \neq j}}^{k} \{c_p^*\} \left[\left(f^* - \frac{d_j^{*2}}{q^*} - \sum_{\substack{m=1 \\ m \neq j}}^{k} \frac{d_m^{*2}}{c_m^*} \right)\left(a^* - \frac{t^{*2}}{q^*} \right) - \left(b^* - \frac{d_j^* t^*}{q^*} \right)^2 \right] \right\}^{-1/2},$$

with

$$b^* = e^{\theta_{n+1}^* + \xi^*}, \qquad t^* = y_{n+1,j}\, e^{\theta_{n+1}^* + \alpha_j^*}, \qquad a^* = e^{\theta_{n+1}^*} x_{n+1} + t^* + b^*,$$

$$c_p^* = (h+2)e^{\alpha_p^*} \sum_{i=1}^{n} \left\{ \frac{\left(x_i + e^{\xi^*}\right) y_i \delta_{ip}}{\left(x_i + y_i e^{\alpha_p^*}\delta_{ip} + e^{\xi^*}\right)^2} \right\} + \frac{(g_p + r_p)s_p e^{\alpha_p^*}}{\left(e^{\alpha_p^*} + s_p\right)^2}, \quad p = 1, 2, \ldots, k,$$

$$d_p^* = -e^{\xi^* + \alpha_p^*}(h+2) \sum_{i=1}^{n} \left\{ \frac{y_i \delta_{ip}}{\left(x_i + y_i e^{\alpha_p^*}\delta_{ip} + e^{\xi^*}\right)^2} \right\}, \quad p = 1, 2, \ldots, k,$$

$$f^* = e^{\xi^*}\left(e^{\theta_{n+1}^*} + v\right) + (h+2)e^{\xi^*} \sum_{i=1}^{n} \left\{ \frac{x_i + y_i \sum_{p=1}^{k} e^{\alpha_p^*}\delta_{ip}}{\left(x_i + y_i \sum_{p=1}^{k} e^{\alpha_p^*}\delta_{ip} + e^{\xi^*}\right)^2} \right\}, \qquad q^* = t^* + c_j^*.$$

Finally, based on these quantities, the Laplace approximation for the predictive distribution of $Y_{n+1,j}$ is given by (1.17) as being

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) \tilde{\propto} \left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\left\{-nh^*\left(\theta^*_{n+1}, \alpha^{k*}, \xi^*\right) + nh\left(\tilde{\theta}_{n+1}, \tilde{\alpha}^k, \tilde{\xi}\right)\right\}.$$

No major simplifications will occur in this Laplace approximation if we assume vague second stage priors.

7.3. A Crossover Design to Compare Two Treatments Effects

The crossover design for the exponential model corresponding to Chapter 4 can be modelled by

$$\text{Period } 1: \quad W_{i1} \sim Ex\left(\exp\left(\theta_i + \delta_{i1}\alpha\right)\right)$$

$$\text{Period } 2: \quad W_{i2} \sim Ex\left(\exp\left(\theta_i + \beta + \delta_{i2}\alpha\right)\right)$$

with treatment indicators δ_{i1} and δ_{i2} as defined in (4.1). Based on

$$D^n = \left\{\left(w_{i1}, w_{i2}, \delta_{i1}, \delta_{i2}\right), i = 1, 2, \dots, n\right\},$$

our aim is to derive the predictive distributions for the outcomes in a new individual who receives T_1 or T_2. These outcomes are represented by the random variables Z_1 and Z_2. Conditional on θ_{n+1}, Z_1 and Z_2 are independent random variables such that

$$Z_1 \sim Ex\left(\exp\left(\theta_{n+1}\right)\right),$$

$$Z_2 \sim Ex\left(\exp\left(\alpha + \theta_{n+1}\right)\right).$$

7.3.1. A Classical Approach

The likelihood function is

$$L\left(\theta^n, \alpha, \beta; D^n\right) = \exp\left\{-\sum_{i=1}^{n} e^{\theta_i + \delta_{i1}\alpha} w_{i1} - \sum_{i=1}^{n} e^{\theta_i + \beta + \delta_{i2}\alpha} w_{i2}\right\} \times$$

$$\times \exp\left\{\sum_{i=1}^{n}\left(\theta_i+\delta_{i1}\alpha\right)+\sum_{i=1}^{n}\left(\theta_i+\beta+\delta_{i2}\alpha\right)\right\}$$

which leads to the normal equations

$$e^{\theta_i+\delta_{i1}\alpha}w_{i1}+e^{\theta_i+\beta+\delta_{i2}\alpha}w_{i2}-2=0, \quad i=1,2,\dots,n,$$

$$\sum_{i=1}^{n}\delta_{i1}-\sum_{i=1}^{n}\delta_{i1}e^{\theta_i+\delta_{i1}\alpha}w_{i1}+\sum_{i=1}^{n}\delta_{i2}-\sum_{i=1}^{n}\delta_{i2}e^{\theta_i+\beta+\delta_{i2}\alpha}w_{i2}=0,$$

$$\sum_{i=1}^{n}e^{\theta_i+\beta+\delta_{i2}\alpha}w_{i2}-n=0.$$

Using a numerical method to solve the system formed by the above $(n+2)$ equations, we obtain the maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n$, $\hat{\alpha}$ and $\hat{\beta}$.

Notice that this system does not provide an estimate for θ_{n+1}, because D^n has no information about the $(n+1)$-th individual. Therefore, it is impossible to derive a plug-in estimate for the predictive distributions we are interested in, since such an approach would require $\hat{\alpha}$ and $\hat{\theta}_{n+1}$.

7.3.2. A Bayesian Approach

Solving the problem in a Bayesian framework, we will derive the predictive distributions using a hierarchical prior structure. At the first stage we take

$$p\left(\theta^n, \theta_{n+1}, \alpha, \beta \mid \xi, \eta, \zeta\right)=\prod_{i=1}^{n+1}\left\{p\left(\theta_i \mid \xi\right)\right\}p\left(\alpha \mid \eta\right)p\left(\beta \mid \zeta\right)$$

and at the second stage

$$p(\xi, \eta, \zeta)=p(\xi)p(\eta)p(\zeta)$$

with

$$e^{\theta_i} \sim Ga\left(k, e^{\xi}\right) \qquad e^{\alpha} \sim Ga\left(h, e^{\eta}\right) \qquad e^{\beta} \sim Ga\left(g, e^{\zeta}\right)$$

$$e^{\xi} \sim Ga(l, m) \qquad e^{\eta} \sim Ga(u, v) \qquad e^{\zeta} \sim Ga(r, s)$$

We assume k, h, g, l, m, u, v, r and s to be specified.

7.3.2.1. The Exact Predictive Distributions

The joint posterior distribution is given by

$$p(\theta^n, \theta_{n+1}, \alpha, \beta, \xi, \eta, \zeta \mid D^n) \propto \exp\left\{-\sum_{i=1}^{n+1} e^{\xi+\theta_i} - e^{\eta+\alpha} - e^{\zeta+\beta} - me^{\xi} - ve^{\eta} - se^{\zeta}\right\} \times$$

$$\times \exp\left\{-\sum_{i=1}^{n} e^{\theta_i+\delta_{i1}\alpha} w_{i1} - \sum_{i=1}^{n} e^{\theta_i+\beta+\delta_{i2}\alpha} w_{i2}\right\} \times$$

$$\times \exp\left\{k\sum_{i=1}^{n+1} \theta_i + h(\eta+\alpha) + g(\zeta+\beta) + u\eta + r\zeta\right\} \times$$

$$\times \exp\left\{((n+1)k + l)\xi + \sum_{i=1}^{n} (\theta_i + \delta_{i1}\alpha)\right\} \times$$

$$\times \exp\left\{\sum_{i=1}^{n} (\theta_i + \beta + \delta_{i2}\alpha)\right\}.$$

As Z_1 and Z_2 only depend on θ_{n+1} and α, we would like to eliminate the remaining parameters. The best we can do is to eliminate θ^n, η and ζ and we obtain the marginal posterior distribution

$$p(\theta_{n+1}, \alpha, \beta, \xi \mid D^n) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}} + m\right)\right\} \exp\left\{((n+1)k + l)\xi + k\theta_{n+1}\right\}}{\left(e^{\alpha} + v\right)^{h+u} \left(e^{\beta} + s\right)^{g+r}} \times$$

$$\times \frac{\exp\left\{\alpha\left(h + \sum_{i=1}^{n} \delta_{i1} + \sum_{i=1}^{n} \delta_{i2}\right) + \beta\,(g+n)\right\}}{\prod_{i=1}^{n} \left\{\left(e^{\delta_{i1}\alpha} w_{i1} + e^{\beta+\delta_{i2}\alpha} w_{i2} + e^{\xi}\right)^{k+2}\right\}}. \qquad (7.10)$$

Then, the predictive distributions of (Z_1, Z_2), Z_1 and Z_2 will be, respectively,

$$p(z_1, z_2 \mid D^n) \propto \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\}\exp\left\{((n+1)k+1)\xi+\alpha\left(1+h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}\right)\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}\prod_{i=1}^{n}\left\{\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)^{k+2}\right\}}\times$$

$$\times\frac{\exp\left\{\beta(g+n)\right\}d\alpha\, d\beta\, d\xi}{\left(e^{\xi}+z_1+e^{\alpha}z_2\right)^{k+2}}, \tag{7.11}$$

$$p(z_1 \mid D^n) \propto \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\}\exp\left\{((n+1)k+1)\xi+\alpha\left(h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}\right)\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}\prod_{i=1}^{n}\left\{\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)^{k+2}\right\}}\times$$

$$\times\frac{\exp\left\{\beta(g+n)\right\}d\alpha\, d\beta\, d\xi}{\left(e^{\xi}+z_1\right)^{k+1}}, \tag{7.12}$$

$$p(z_2 \mid D^n) \propto \int_{\Re^3} \frac{\exp\left\{-me^{\xi}\right\}\exp\left\{((n+1)k+1)\xi+\alpha\left(1+h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}\right)\right\}}{\left(e^{\alpha}+v\right)^{h+u}\left(e^{\beta}+s\right)^{g+r}\prod_{i=1}^{n}\left\{\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)^{k+2}\right\}}\times$$

$$\times\frac{\exp\left\{\beta(g+n)\right\}d\alpha\, d\beta\, d\xi}{\left(e^{\xi}+e^{\alpha}z_2\right)^{k+1}}. \tag{7.13}$$

The evaluation of these exact predictive distributions will then require some numerical integration technique to solve the above three-dimensional integrals, or the use of some approximate methods. The assumption of vague second stage priors does not lead to important simplifications.

7.3.2.2. Estimation Via Gibbs Sampling

From (7.10), the full conditional distributions required for the implementation of the Gibbs routine are easily derived as being

$$p\left(\theta_{n+1} \mid \alpha, \beta, \xi, D^n\right) \propto \exp\left\{-e^{\theta_{n+1}+\xi}\right\}\exp\left\{k\,\theta_{n+1}\right\},$$

$$p\left(\alpha \mid \theta_{n+1}, \beta, \xi, D^n\right) \propto \frac{\exp\left\{\alpha\left(h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}\right)\right\}}{\prod_{i=1}^{n}\left\{\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)^{k+2}\right\}\left(e^{\alpha}+v\right)^{h+u}},$$

$$p\left(\beta \mid \theta_{n+1}, \alpha, \xi, D^n\right) \propto \frac{\exp\left\{\beta\left(g+n\right)\right\}}{\prod_{i=1}^{n}\left\{\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)^{k+2}\right\}\left(e^{\beta}+s\right)^{g+r}},$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, \beta, D^n\right) \propto \frac{\exp\left\{-e^{\xi}\left(e^{\theta_{n+1}}+m\right)\right\}\exp\left\{\left((n+1)k+l\right)\xi\right\}}{\prod_{i=1}^{n}\left\{\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)^{k+2}\right\}},$$

and they do not get significantly simpler when we consider vague second stage priors.

The generation of values of θ_{n+1} is direct noting by Table A1.1 that its full conditional distribution is a transformed gamma distribution. The generation of values of α, β and ξ requires the use of a sampling technique such as the rejection sampling algorithm (section 1.2.3).

7.3.2.3. Estimation Via Asymptotic Results

The first asymptotic normal approximation for the posterior distribution presented in section 1.2.2 (Bernardo & Smith, 1994) is based on the maximum likelihood estimates for the parameters. Since we cannot derive an estimate for θ_{n+1}, as shown in section 7.3.1, such an approach is not suitable to solve our problem.

The second asymptotic result in section 1.2.2, suggested by O'Hagan (1994), is based in the posterior mode and on the modal dispersion matrix, both obtained from (7.10).

Solving numerically the system formed by the equations

$$e^{\xi+\theta_{n+1}} - k = 0,$$

$$h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}-\frac{(h+u)e^{\alpha}}{e^{\alpha}+v}-\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha}w_{i1}+\delta_{i2}e^{\beta+\delta_{i2}\alpha}w_{i2}\right)}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}=0, \quad (7.14)$$

$$g+n-\frac{(g+r)e^{\beta}}{e^{\beta}+s}-\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta+\delta_{i2}\alpha}w_{i2}}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}=0,$$

$$e^{\xi}\left(e^{\theta_{n+1}}+m\right)+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\xi}}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}-(n+1)k-l=0,$$

we obtain the posterior mode $m=\left(\tilde{\theta}_{n+1},\tilde{\alpha},\tilde{\beta},\tilde{\xi}\right)$. Then, defining the constants

$$\tilde{a}=e^{\tilde{\xi}+\tilde{\theta}_{n+1}},$$

$$\tilde{b}=\frac{(h+u)ve^{\tilde{\alpha}}}{\left(e^{\tilde{\alpha}}+v\right)^{2}}+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\tilde{\alpha}}w_{i1}\left((1-\delta_{i2})e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}+e^{\tilde{\xi}}\right)\right)}{\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}+e^{\tilde{\xi}}\right)^{2}}\right\}+$$

$$+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i2}e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}\left((1-\delta_{i1})e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\xi}}\right)\right)}{\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}+e^{\tilde{\xi}}\right)^{2}}\right\},$$

$$\tilde{c}=\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}\left((\delta_{i2}-\delta_{i1})e^{\delta_{i1}\tilde{\alpha}}w_{i1}+\delta_{i2}e^{\tilde{\xi}}\right)}{\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}+e^{\tilde{\xi}}\right)^{2}}\right\},$$

$$\tilde{d}=-\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\tilde{\alpha}}w_{i1}+\delta_{i2}e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}\right)e^{\tilde{\xi}}}{\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}+e^{\tilde{\xi}}\right)^{2}}\right\},$$

$$\tilde{f}=\frac{(g+r)se^{\tilde{\beta}}}{\left(e^{\tilde{\beta}}+s\right)^{2}}+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\xi}}\right)}{\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}+e^{\tilde{\xi}}\right)^{2}}\right\},$$

$$\tilde{q}=-\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}+\tilde{\xi}}w_{i2}}{\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1}+e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}+e^{\tilde{\xi}}\right)^{2}}\right\},$$

$$\tilde{t} = e^{\tilde{\xi}}\left(e^{\tilde{\theta}_{n+1}} + m\right) + \sum_{i=1}^{n}\left\{\frac{(k+2)\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1} + e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2}\right)e^{\tilde{\xi}}}{\left(e^{\delta_{i1}\tilde{\alpha}}w_{i1} + e^{\tilde{\beta}+\delta_{i2}\tilde{\alpha}}w_{i2} + e^{\tilde{\xi}}\right)^2}\right\},$$

the full conditional distributions will become

$$p\left(\theta_{n+1} \mid \alpha, \beta, \xi, D^n\right) = N\left(\tilde{\theta}_{n+1} - \left(\xi - \tilde{\xi}\right), \frac{1}{\tilde{a}}\right),$$

$$p\left(\alpha \mid \theta_{n+1}, \beta, \xi, D^n\right) = N\left(\tilde{\alpha} - \frac{\left(\beta - \tilde{\beta}\right)\tilde{c} + \left(\xi - \tilde{\xi}\right)\tilde{d}}{\tilde{b}}, \frac{1}{\tilde{b}}\right),$$

$$p\left(\beta \mid \theta_{n+1}, \alpha, \xi, D^n\right) = N\left(\tilde{\beta} - \frac{\left(\alpha - \tilde{\alpha}\right)\tilde{c} + \left(\xi - \tilde{\xi}\right)\tilde{q}}{\tilde{f}}, \frac{1}{\tilde{f}}\right),$$

$$p\left(\xi \mid \theta_{n+1}, \alpha, \beta, D^n\right) = N\left(\tilde{\xi} - \frac{\left(\theta_{n+1} - \tilde{\theta}_{n+1}\right)\tilde{a} + \left(\alpha - \tilde{\alpha}\right)\tilde{d} + \left(\beta - \tilde{\beta}\right)\tilde{q}}{\tilde{t}}, \frac{1}{\tilde{t}}\right).$$

The predictive distributions we are interested in can then be estimated using the above full conditional distributions in the Gibbs routine. If we consider the special vague second stage priors case, the implementation of this approximation does not become much simpler.

7.3.2.4. Laplace Approximation

Let us define the function $h\left(\theta_{n+1}, \alpha, \beta, \xi\right)$ as in (1.13), that is, such that

$$-n\,h\left(\theta_{n+1}, \alpha, \beta, \xi\right) = -e^{\xi}\left(e^{\theta_{n+1}} + m\right) + \left((n+1)k + l\right)\xi + k\theta_{n+1} +$$

$$+\alpha\left(h + \sum_{i=1}^{n}\delta_{i1} + \sum_{i=1}^{n}\delta_{i2}\right) + \beta(g+n) -$$

$$-(h+u)\ln\left(e^{\alpha} + v\right) - (g+r)\ln\left(e^{\beta} + s\right) -$$

$$-\sum_{i=1}^{n}\left\{(k+2)\ln\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)\right\},$$

and let us also define $\left(\bar{\theta}_{n+1}, \bar{\alpha}, \bar{\beta}, \bar{\xi}\right)$ and $\bar{\sigma}$ as in (1.16). We then conclude that $\left(\bar{\theta}_{n+1}, \bar{\alpha}, \bar{\beta}, \bar{\xi}\right)=m$, solution of the system of equations (7.14) and that

$$\bar{\sigma}=\left\{\bar{a}\bar{b}\bar{f}\bar{t}-\bar{a}\bar{b}\bar{q}^{2}-\bar{a}\bar{t}\bar{c}^{2}+2\bar{a}\bar{c}\bar{d}\bar{q}-\bar{a}\bar{f}\bar{d}^{2}-\bar{b}\bar{f}\bar{a}^{2}+\bar{a}^{2}\bar{c}^{2}\right\}^{-1/2}.$$

7.3.2.4.1. Joint Predictive Distribution of (Z_1, Z_2)

To approximate (7.11) we define, according to (1.14), a function $h_c^*(\theta_{n+1}, \alpha, \beta, \xi)$ such that

$$-nh_c^*(\theta_{n+1}, \alpha, \beta, \xi)=\alpha+2\theta_{n+1}-e^{\theta_{n+1}}\left(z_1+e^{\alpha}z_2\right)-e^{\xi}\left(e^{\theta_{n+1}}+m\right)+\left((n+1)k+l\right)\xi+$$

$$+k\theta_{n+1}+\alpha\left(h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}\right)+\beta(g+n)-$$

$$-(h+u)\ln\left(e^{\alpha}+v\right)-(g+r)\ln\left(e^{\beta}+s\right)-$$

$$-\sum_{i=1}^{n}\left\{(k+2)\ln\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)\right\}.$$

Then, we define $\left(\theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c, \xi^*_c\right)$ and σ^*_c such that the definitions in (1.16) hold. Solving numerically the system formed by the equations

$$e^{\theta_{n+1}}\left(z_1+e^{\alpha}z_2+e^{\xi}\right)-k-2=0,$$

$$e^{\theta_{n+1}+\alpha}z_2+\frac{(h+u)e^{\alpha}}{e^{\alpha}+v}+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha}w_{i1}+\delta_{i2}e^{\beta+\delta_{i2}\alpha}w_{i2}\right)}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}-$$

$$-\sum_{i=1}^{n}\delta_{i1}-\sum_{i=1}^{n}\delta_{i2}-h-1=0,$$

$$g+n-\frac{(g+r)e^{\beta}}{e^{\beta}+s}-\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta+\delta_{i2}\alpha}w_{i2}}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}=0,$$

$$e^{\xi}\left(e^{\theta_{n+1}}+m\right)+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\xi}}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}-(n+1)k-l=0,$$

we obtain $\left(\theta^{*}_{n+1(c)}, \alpha^{*}_{c}, \beta^{*}_{c}, \xi^{*}_{c}\right)$, and σ^{*}_{c} will be given by

$$\sigma^{*}_{c}=\left\{a^{*}_{c}d^{*}_{c}t^{*}_{c}p^{*}_{c}-a^{*}_{c}d^{*}_{c}o^{*2}_{c}-a^{*}_{c}p^{*}_{c}f^{*2}_{c}+2a^{*}_{c}f^{*}_{c}q^{*}_{c}o^{*}_{c}-a^{*}_{c}t^{*}_{c}q^{*2}_{c}-t^{*}_{c}p^{*}_{c}b^{*2}_{c}+\right.$$

$$\left.+b^{*2}_{c}o^{*2}_{c}-2b^{*}_{c}f^{*}_{c}c^{*}_{c}o^{*}_{c}+2b^{*}_{c}q^{*}_{c}c^{*}_{c}t^{*}_{c}-d^{*}_{c}t^{*}_{c}c^{*2}_{c}+f^{*2}_{c}c^{*2}_{c}\right\}^{-1/2},$$

where

$$b^{*}_{c}=e^{\theta^{*}_{n+1(c)}+\alpha^{*}_{c}}z_{2},\qquad c^{*}_{c}=e^{\theta^{*}_{n+1(c)}+\xi^{*}_{c}},\qquad a^{*}_{c}=e^{\theta^{*}_{n+1(c)}}z_{1}+b^{*}_{c}+c^{*}_{c},$$

$$d^{*}_{c}=e^{\theta^{*}_{n+1(c)}+\alpha^{*}_{c}}z_{2}+\frac{(h+u)ve^{\alpha^{*}_{c}}}{\left(e^{\alpha^{*}_{c}}+v\right)^{2}}+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}\left((1-\delta_{i2})e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}+e^{\xi^{*}_{c}}\right)\right)}{\left(e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}+e^{\xi^{*}_{c}}\right)^{2}}\right\}+$$

$$+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i2}e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}\left((1-\delta_{i1})e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+e^{\xi^{*}_{c}}\right)\right)}{\left(e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}+e^{\xi^{*}_{c}}\right)^{2}}\right\},$$

$$f^{*}_{c}=\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}\left((\delta_{i2}-\delta_{i1})e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+\delta_{i2}e^{\xi^{*}_{c}}\right)}{\left(e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}+e^{\xi^{*}_{c}}\right)^{2}}\right\},$$

$$q^{*}_{c}=-\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+\delta_{i2}e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}\right)e^{\xi^{*}_{c}}}{\left(e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}+e^{\xi^{*}_{c}}\right)^{2}}\right\},$$

$$t^{*}_{c}=\frac{(g+r)se^{\beta^{*}_{c}}}{\left(e^{\beta^{*}_{c}}+s\right)^{2}}+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}\left(e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+e^{\xi^{*}_{c}}\right)}{\left(e^{\delta_{i1}\alpha^{*}_{c}}w_{i1}+e^{\beta^{*}_{c}+\delta_{i2}\alpha^{*}_{c}}w_{i2}+e^{\xi^{*}_{c}}\right)^{2}}\right\},$$

$$o_c^* = -\sum_{i=1}^{n}\left\{\frac{(k+2)\,e^{\beta_c^*+\delta_{i2}\alpha_c^*+\xi_c^*}w_{i2}}{\left(e^{\delta_{i1}\alpha_c^*}w_{i1}+e^{\beta_c^*+\delta_{i2}\alpha_c^*}w_{i2}+e^{\xi_c^*}\right)^2}\right\},$$

$$p_c^* = e^{\xi_c^*}\left(e^{\theta_{n+1(c)}^*}+m\right)+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(e^{\delta_{i1}\alpha_c^*}w_{i1}+e^{\beta_c^*+\delta_{i2}\alpha_c^*}w_{i2}\right)e^{\xi_c^*}}{\left(e^{\delta_{i1}\alpha_c^*}w_{i1}+e^{\beta_c^*+\delta_{i2}\alpha_c^*}w_{i2}+e^{\xi_c^*}\right)^2}\right\}.$$

Finally, the joint predictive distribution (7.11) is approximately given by

$$p(z_1, z_2 \mid D^n) \cong \left(\frac{\sigma_c^*}{\bar{\sigma}}\right)\exp\left\{-nh_c^*\left(\theta_{n+1(c)}^*, \alpha_c^*, \beta_c^*, \xi_c^*\right)+nh\left(\bar{\theta}_{n+1}, \bar{\alpha}, \bar{\beta}, \bar{\xi}\right)\right\}.$$

7.3.2.4.2. Marginal Predictive Distribution of Z_1

The Laplace approximation for (7.12) is similarly derived. We consider a function $h_1^*(\theta_{n+1}, \alpha, \beta, \xi)$ such that

$$-nh_1^*(\theta_{n+1}, \alpha, \beta, \xi) = \theta_{n+1} - e^{\theta_{n+1}}z_1 - e^{\xi}\left(e^{\theta_{n+1}}+m\right)+((n+1)k+l)\xi +$$

$$+k\theta_{n+1}+\alpha\left(h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}\right)+\beta(g+n)-$$

$$-(h+u)\ln\left(e^{\alpha}+v\right)-(g+r)\ln\left(e^{\beta}+s\right)-$$

$$-\sum_{i=1}^{n}\left\{(k+2)\ln\left(e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}\right)\right\}$$

and, from (1.16), we derive $\left(\theta_{n+1(1)}^*, \alpha_1^*, \beta_1^*, \xi_1^*\right)$ solving numerically the system formed by the equations

$$e^{\theta_{n+1}}\left(z_1+e^{\xi}\right)-k-1=0,$$

$$h+\sum_{i=1}^{n}\delta_{i1}+\sum_{i=1}^{n}\delta_{i2}-\frac{(h+u)e^{\alpha}}{e^{\alpha}+v}-\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha}w_{i1}+\delta_{i2}e^{\beta+\delta_{i2}\alpha}w_{i2}\right)}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}=0,$$

$$g+n-\frac{(g+r)e^{\beta}}{e^{\beta}+s}-\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta+\delta_{i2}\alpha}w_{i2}}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}=0,$$

$$e^{\xi}\left(e^{\theta_{n+1}}+m\right)+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\xi}}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}-(n+1)k-l=0.$$

Then, defining the constants

$$a_1^{*}=e^{\theta_{n+1(1)}^{*}}\left(z_1+e^{\xi_1^{*}}\right), \qquad b_1^{*}=e^{\theta_{n+1(1)}^{*}+\xi_1^{*}},$$

$$c_1^{*}=\frac{(h+u)ve^{\alpha_1^{*}}}{\left(e^{\alpha_1^{*}}+v\right)^2}+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha_1^{*}}w_{i1}\left((1-\delta_{i2})e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)\right)}{\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)^2}\right\}+$$

$$+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i2}e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}\left((1-\delta_{i1})e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\xi_1^{*}}\right)\right)}{\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)^2}\right\},$$

$$d_1^{*}=\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}\left((\delta_{i2}-\delta_{i1})e^{\delta_{i1}\alpha_1^{*}}w_{i1}+\delta_{i2}e^{\xi_1^{*}}\right)}{\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)^2}\right\},$$

$$f_1^{*}=-\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha_1^{*}}w_{i1}+\delta_{i2}e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}\right)e^{\xi_1^{*}}}{\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)^2}\right\},$$

$$q_1^{*}=\frac{(g+r)se^{\beta_1^{*}}}{\left(e^{\beta_1^{*}}+s\right)^2}+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\xi_1^{*}}\right)}{\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)^2}\right\},$$

$$t_1^{*}=-\sum_{i=1}^{n}\left\{\frac{(k+2)\,e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}+\xi_1^{*}}w_{i2}}{\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)^2}\right\},$$

$$p_1^{*}=e^{\xi_1^{*}}\left(e^{\theta_{n+1(1)}^{*}}+m\right)+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}\right)e^{\xi_1^{*}}}{\left(e^{\delta_{i1}\alpha_1^{*}}w_{i1}+e^{\beta_1^{*}+\delta_{i2}\alpha_1^{*}}w_{i2}+e^{\xi_1^{*}}\right)^2}\right\},$$

we derive σ_1^* in definition (1.16) as being

$$\sigma_1^* = \left\{a_1^* c_1^* q_1^* p_1^* - a_1^* c_1^* t_1^{*2} - a_1^* p_1^* d_1^{*2} + 2a_1^* d_1^* f_1^* t_1^* - a_1^* q_1^* f_1^{*2} - c_1^* q_1^* b_1^{*2} + d_1^{*2} b_1^{*2}\right\}^{-1/2}$$

and the predictive distribution of Z_1 will be approximated by

$$p(z_1 \mid D^n) \propto \left(\frac{\sigma_1^*}{\tilde{\sigma}}\right) \exp\left\{-nh_1^*\left(\theta_{n+1(1)}^*, \alpha_1^*, \beta_1^*, \xi_1^*\right) + nh\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\right\}.$$

7.3.2.4.3. Marginal Predictive Distribution of Z_2

The approximation of (7.13) requires a function $h_2^*(\theta_{n+1}, \alpha, \beta, \xi)$ defined such that

$$-nh_2^*(\theta_{n+1}, \alpha, \beta, \xi) = \alpha + \theta_{n+1} - e^{\alpha+\theta_{n+1}} z_2 - e^{\xi}\left(e^{\theta_{n+1}} + m\right) + ((n+1)k + l)\xi +$$

$$+ k\theta_{n+1} + \alpha\left(h + \sum_{i=1}^{n} \delta_{i1} + \sum_{i=1}^{n} \delta_{i2}\right) + \beta(g+n) -$$

$$-(h+u)\ln\left(e^{\alpha} + v\right) - (g+r)\ln\left(e^{\beta} + s\right) -$$

$$-\sum_{i=1}^{n}\left\{(k+2)\ln\left(e^{\delta_{i1}\alpha} w_{i1} + e^{\beta+\delta_{i2}\alpha} w_{i2} + e^{\xi}\right)\right\}.$$

Following definition (1.16), we solve numerically the system formed by the equations

$$e^{\theta_{n+1}}\left(e^{\alpha} z_2 + e^{\xi}\right) - k - 1 = 0,$$

$$e^{\theta_{n+1}+\alpha} z_2 + \frac{(h+u)e^{\alpha}}{e^{\alpha}+v} + \sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1} e^{\delta_{i1}\alpha} w_{i1} + \delta_{i2} e^{\beta+\delta_{i2}\alpha} w_{i2}\right)}{e^{\delta_{i1}\alpha} w_{i1} + e^{\beta+\delta_{i2}\alpha} w_{i2} + e^{\xi}}\right\} -$$

$$-\sum_{i=1}^{n} \delta_{i1} - \sum_{i=1}^{n} \delta_{i2} - h - 1 = 0,$$

$$g + n - \frac{(g+r)e^{\beta}}{e^{\beta}+s} - \sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta+\delta_{i2}\alpha} w_{i2}}{e^{\delta_{i1}\alpha} w_{i1} + e^{\beta+\delta_{i2}\alpha} w_{i2} + e^{\xi}}\right\} = 0,$$

$$e^{\xi}\left(e^{\theta_{n+1}}+m\right)+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\xi}}{e^{\delta_{i1}\alpha}w_{i1}+e^{\beta+\delta_{i2}\alpha}w_{i2}+e^{\xi}}\right\}-(n+1)k-l=0,$$

obtaining $\left(\theta^{*}_{n+1(2)}, \alpha^{*}_{2}, \beta^{*}_{2}, \xi^{*}_{2}\right)$, and we define σ^{*}_{2} as

$$\sigma^{*}_{2}=\left\{a^{*}_{2}d^{*}_{2}t^{*}_{2}p^{*}_{2}-a^{*}_{2}d^{*}_{2}o^{*2}_{2}-a^{*}_{2}p^{*}_{2}f^{*2}_{2}+2a^{*}_{2}f^{*}_{2}q^{*}_{2}o^{*}_{2}-a^{*}_{2}t^{*}_{2}q^{*2}_{2}-t^{*}_{2}p^{*}_{2}b^{*2}_{2}+\right.$$

$$\left.+b^{*2}_{2}o^{*2}_{2}-2b^{*}_{2}f^{*}_{2}c^{*}_{2}o^{*}_{2}+2b^{*}_{2}q^{*}_{2}c^{*}_{2}t^{*}_{2}-d^{*}_{2}t^{*}_{2}c^{*2}_{2}+f^{*2}_{2}c^{*2}_{2}\right\}^{-1/2},$$

with

$$a^{*}_{2}=e^{\theta^{*}_{n+1(2)}}\left(e^{\alpha^{*}_{2}}z_{2}+e^{\xi^{*}_{2}}\right),\qquad b^{*}_{2}=e^{\theta^{*}_{n+1(2)}+\alpha^{*}_{2}}z_{2},\qquad c^{*}_{2}=e^{\theta^{*}_{n+1(2)}+\xi^{*}_{2}},$$

$$d^{*}_{2}=e^{\theta^{*}_{n+1(2)}+\alpha^{*}_{2}}z_{2}+\frac{(h+u)ve^{\alpha^{*}_{2}}}{\left(e^{\alpha^{*}_{2}}+v\right)^{2}}+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}\left((1-\delta_{i2})e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}+e^{\xi^{*}_{2}}\right)\right)}{\left(e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}+e^{\xi^{*}_{2}}\right)^{2}}\right\}+$$

$$+\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i2}e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}\left((1-\delta_{i1})e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\xi^{*}_{2}}\right)\right)}{\left(e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}+e^{\xi^{*}_{2}}\right)^{2}}\right\},$$

$$f^{*}_{2}=\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}\left((\delta_{i2}-\delta_{i1})e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+\delta_{i2}e^{\xi^{*}_{2}}\right)}{\left(e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}+e^{\xi^{*}_{2}}\right)^{2}}\right\},$$

$$q^{*}_{2}=-\sum_{i=1}^{n}\left\{\frac{(k+2)\left(\delta_{i1}e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+\delta_{i2}e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}\right)e^{\xi^{*}_{2}}}{\left(e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}+e^{\xi^{*}_{2}}\right)^{2}}\right\},$$

$$t^{*}_{2}=\frac{(g+r)se^{\beta^{*}_{2}}}{\left(e^{\beta^{*}_{2}}+s\right)^{2}}+\sum_{i=1}^{n}\left\{\frac{(k+2)e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}\left(e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\xi^{*}_{2}}\right)}{\left(e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}+e^{\xi^{*}_{2}}\right)^{2}}\right\},$$

$$o^{*}_{2}=-\sum_{i=1}^{n}\left\{\frac{(k+2)\,e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}+\xi^{*}_{2}}w_{i2}}{\left(\left(e^{\delta_{i1}\alpha^{*}_{2}}w_{i1}+e^{\beta^{*}_{2}+\delta_{i2}\alpha^{*}_{2}}w_{i2}+e^{\xi^{*}_{2}}\right)^{2}\right.}\right\},$$

$$p_2^* = e^{\xi_2^*}\left(e^{\theta^*_{n+1(2)}} + m\right) + \sum_{i=1}^{n}\left\{\frac{(k+2)\left(e^{\delta_{i1}\alpha_2^*}w_{i1} + e^{\beta_2^*+\delta_{i2}\alpha_2^*}w_{i2}\right)e^{\xi_2^*}}{\left(e^{\delta_{i1}\alpha_2^*}w_{i1} + e^{\beta_2^*+\delta_{i2}\alpha_2^*}w_{i2} + e^{\xi_2^*}\right)^2}\right\}.$$

Hence, the approximation for the predictive distribution of Z_2 is

$$p\left(z_2 \mid D^n\right) \mathrel{\tilde{\propto}} \left(\frac{\sigma_2^*}{\tilde{\sigma}}\right) \exp\left\{-nh_2^*\left(\theta^*_{n+1(2)}, \alpha_2^*, \beta_2^*, \xi_2^*\right) + nh\left(\tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}, \tilde{\xi}\right)\right\}.$$

These Laplace approximations do not get easier to implement when vague second stage priors are considered.

CHAPTER 8

PREDICTION IN SOME BINOMIAL ERRORS IN VARIABLES MODELS

In this chapter we consider binomial models for the situations discussed in Chapters 2, 3 and 4. Once again we will omit the details, presenting only the final results. Such details will only be discussed when they are significantly different from the ones in the other chapters.

8.1. The Multiplicative Effect of a Treatment

Suppose we have independent random variables (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) where X_i and Y_i $(i=1, 2, ..., n)$ represent, respectively, the observed number of successes on the i-th individual, before and after a treatment with multiplicative effect is applied. We assume that

- given r and θ_i, $X_i \sim Bi\left(r, \dfrac{e^{\theta_i}}{1+e^{\theta_i}}\right)$,

- given s, θ_i and α, $Y_i \sim Bi\left(s, \dfrac{e^{\alpha+\theta_i}}{1+e^{\alpha+\theta_i}}\right)$.

Here, $\theta_1, \theta_2, ..., \theta_n$ are parameters used to model the characteristics of the individuals and α is a parameter which models the effect of the treatment upon the individuals. The parameterisation chosen was taken such that the parameters involved would be unrestricted and also in a way that we can be sure that the probability parameter would lie in $]0, 1[$. Note that the log odds are linear; namely, they are θ_i for X_i and $\alpha + \theta_i$ for Y_i.

After observing n individuals, we get a data set

$$D^n = \{(x_i, y_i), i = 1, 2, ..., n\}.$$

Our aim is to make predictions about Y_{n+1}, the outcome on a new individual after the treatment is used, based on D^n and on an observed x_{n+1}, the number of successes on the $(n+1)$-th individual before the treatment is applied. We assume

$$X_{n+1} \sim Bi\left(r, \frac{e^{\theta_{n+1}}}{1+e^{\theta_{n+1}}}\right),$$

$$Y_{n+1} \sim Bi\left(s, \frac{e^{\alpha+\theta_{n+1}}}{1+e^{\alpha+\theta_{n+1}}}\right).$$

8.1.1. A Classical Approach

The likelihood function is

$$L(\theta^n, \theta_{n+1}, \alpha; D^n, x_{n+1}) \propto \frac{\exp\left\{\sum_{i=1}^{n+1} \theta_i x_i + \sum_{i=1}^{n} (\alpha + \theta_i) y_i\right\}}{\prod_{i=1}^{n} \left\{(1+e^{\theta_i})^r (1+e^{\alpha+\theta_i})^s\right\} (1+e^{\theta_{n+1}})^r}$$

and we easily conclude that the maximum likelihood estimate for θ_{n+1} is

$$\hat{\theta}_{n+1} = \ln\left(\frac{x_{n+1}}{r - x_{n+1}}\right),$$

whilst the maximum likelihood estimates $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n$ and $\hat{\alpha}$ are obtained solving numerically the system formed by the $(n+1)$ equations

$$x_i + y_i - \frac{re^{\theta_i}}{1+e^{\theta_i}} - \frac{se^{\alpha+\theta_i}}{1+e^{\alpha+\theta_i}} = 0 \quad , \quad i = 1, 2, ..., n,$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \frac{se^{\alpha+\theta_i}}{1+e^{\alpha+\theta_i}} = 0.$$

Then, a simple estimative approximation for the predictive distribution of Y_{n+1} would be

$$Y_{n+1} \sim Bi\left(s, \frac{e^{\hat{\alpha}+\hat{\theta}_{n+1}}}{1+e^{\hat{\alpha}+\hat{\theta}_{n+1}}}\right).$$

Problems arise if $x_{n+1} = 0$ or if $x_{n+1} = r$.

8.1.2. A Bayesian Approach

No simple conjugate priors present themselves, and so, as before, we consider a hierarchical prior structure: at the first stage we take

$$p(\theta^n, \theta_{n+1}, \alpha \mid \xi, \eta) = \prod_{i=1}^{n+1} \{p(\theta_i \mid \xi)\} p(\alpha \mid \eta)$$

and at the second stage we take

$$p(\xi, \eta) = p(\xi)\, p(\eta)$$

with

$$e^{\theta_i} \sim Ga(a, e^{\xi}) \qquad e^{\alpha} \sim Ga(b, e^{\eta})$$

$$e^{\xi} \sim Ga(c, d) \qquad e^{\eta} \sim Ga(g, h)$$

where a, b, c, d, g and h are assumed to be known.

8.1.2.1. The Exact Predictive Distribution

The joint posterior distribution is given by

$$p\left(\theta^n, \theta_{n+1}, \alpha, \xi, \eta \mid D^n, x_{n+1}\right) \propto \frac{\exp\left\{-\sum_{i=1}^{n+1} e^{\xi+\theta_i} - e^{\eta+\alpha} - d e^{\xi} - h e^{\eta}\right\}}{\prod_{i=1}^{n}\left\{\left(1+e^{\theta_i}\right)^r\left(1+e^{\alpha+\theta_i}\right)^s\right\}} \times$$

$$\times \frac{\exp\left\{\left((n+1)a+c\right)\xi + \sum_{i=1}^{n}\theta_i\left(a+x_i+y_i\right) + \theta_{n+1}\left(a+x_{n+1}\right)\right\}}{\left(1+e^{\theta_{n+1}}\right)^r} \times$$

$$\times \exp\left\{(b+g)\eta + \alpha\left(b+\sum_{i=1}^{n} y_i\right)\right\}.$$

Eliminating η and ξ and defining

$$S_x = \sum_{i=1}^{n} x_i \qquad\qquad S_y = \sum_{i=1}^{n} y_i$$

the marginal posterior distribution for $\left(\theta^n, \theta_{n+1}, \alpha\right)$ is

$$p\left(\theta^n, \theta_{n+1}, \alpha \mid D^n, x_{n+1}\right) \propto \frac{\exp\left\{\sum_{i=1}^{n}\theta_i\left(a+x_i+y_i\right) + \theta_{n+1}\left(a+x_{n+1}\right)\right\}}{\prod_{i=1}^{n}\left\{\left(1+e^{\theta_i}\right)^r\left(1+e^{\alpha+\theta_i}\right)^s\right\}\left(1+e^{\theta_{n+1}}\right)^r\left(e^{\alpha}+h\right)^{b+g}} \times$$

$$\times \frac{\exp\left\{\alpha\left(b+S_y\right)\right\}}{\left(d+\sum_{i=1}^{n+1} e^{\theta_i}\right)^{(n+1)a+c}}, \tag{8.1}$$

and the predictive distribution of Y_{n+1} will be

$$p\left(y_{n+1} \mid D^n, x_{n+1}\right) \propto \binom{s}{y_{n+1}} \int_{\Re^{n+2}} \frac{\exp\left\{\sum_{i=1}^{n+1}\theta_i\left(a+x_i+y_i\right) + \alpha\left(b+S_y+y_{n+1}\right)\right\} d\theta^n\, d\theta_{n+1}\, d\alpha}{\prod_{i=1}^{n+1}\left\{\left(1+e^{\theta_i}\right)^r\left(1+e^{\alpha+\theta_i}\right)^s\right\}\left(e^{\alpha}+h\right)^{b+g}\left(d+\sum_{i=1}^{n+1} e^{\theta_i}\right)^{(n+1)a+c}}$$

(8.2)

Note that the evaluation of this exact predictive distribution requires the use of a numerical technique to solve the $(n+2)$-dimensional integral involved. Due to its high dimensionality, we should expect numerical problems.

With second stage vague priors $(c, d, g, h \to 0)$ the predictive distribution (8.2) does not simplify much.

8.1.2.2. Estimation Via Gibbs Sampling

One of the possible ways of avoiding the integral in (8.2) is to estimate such a predictive distribution through the Gibbs routine (section 1.2.1). To do that we need to know the full conditional distributions. From (8.1), they are

$$p\left(\theta_i \mid \theta_{j \neq i}, \theta_{n+1}, \alpha, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\theta_i\left(a + x_i + y_i\right)\right\}}{\left(1 + e^{\theta_i}\right)^r \left(1 + e^{\alpha + \theta_i}\right)^s \left(d + \sum_{i=1}^{n+1} e^{\theta_i}\right)^{(n+1)a + c}} ,$$

$$i = 1, 2, \ldots, n,$$

$$p\left(\theta_{n+1} \mid \theta^n, \alpha, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\theta_{n+1}\left(a + x_{n+1}\right)\right\}}{\left(1 + e^{\theta_{n+1}}\right)^r \left(d + \sum_{i=1}^{n+1} e^{\theta_i}\right)^{(n+1)a + c}} ,$$

$$p\left(\alpha \mid \theta^n, \theta_{n+1}, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\alpha\left(b + S_y\right)\right\}}{\prod_{i=1}^{n}\left\{\left(1 + e^{\alpha + \theta_i}\right)^s\right\}\left(h + e^{\alpha}\right)^{b + g}} .$$

Even when we consider vague second stage priors, these full conditional distributions will not become significantly simpler.

This approach should be seen as possibly problematic because there will be a large number of values to generate in each cycle of the Gibbs routine, which can make the algorithm very time consuming.

8.1.2.3. Estimation Via Asymptotic Results

The normal approximations for the posterior distribution summarised in section 1.2.2 will now be considered in order to estimate the predictive distribution of Y_{n+1}.

8.1.2.3.1. Posterior Normality Based on the Likelihood Function

The estimation of the predictive distribution of Y_{n+1} only requires the posterior distribution for (θ_{n+1}, α). Following the asymptotic result presented by Bernardo & Smith (1994) and summarised in section 1.2.2.1, we find that, if n is large enough, θ_{n+1} and α are independent a posteriori, with

$$p\left(\theta_{n+1} \mid D^n, x_{n+1}\right) = N\left(\hat{\theta}_{n+1}, \frac{1}{m}\right),$$

$$p\left(\alpha \mid D^n, x_{n+1}\right) = N\left(\hat{\alpha}, \frac{1}{t}\right),$$

where

$$m = \frac{r\, e^{\hat{\theta}_{n+1}}}{\left(1 + e^{\hat{\theta}_{n+1}}\right)^2} = \frac{\left(r - x_{n+1}\right)x_{n+1}}{r}, \qquad t = \sum_{i=1}^{n}\left\{\frac{rs\, e^{\hat{\alpha}+\hat{\theta}_i}}{r\left(1 + e^{\hat{\alpha}+\hat{\theta}_i}\right)^2 + s\, e^{\hat{\alpha}}\left(1 + e^{\hat{\theta}_i}\right)^2}\right\}.$$

This method cannot be implemented if $x_{n+1} = 0$ or if $x_{n+1} = r$.

8.1.2.3.2. Posterior Normality Based on Characteristics of the Posterior Distribution

The normal approximation for the posterior distribution suggested by O'Hagan (1994) and summarised in section 1.2.2.2, is based on the posterior mode and on the modal dispersion matrix. Considering (8.1), the posterior mode $m = \left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}\right)$ is obtained, according to (1.4), solving through a numerical method the system formed by the equations

$$a+x_i+y_i-\frac{re^{\theta_i}}{1+e^{\theta_i}}-\frac{se^{\alpha+\theta_i}}{1+e^{\alpha+\theta_i}}-\frac{((n+1)a+c)e^{\theta_i}}{d+\sum_{i=1}^{n+1}e^{\theta_i}}=0 \quad, i=1,2,\dots,n,$$

$$a+x_{n+1}-\frac{re^{\theta_{n+1}}}{1+e^{\theta_{n+1}}}-\frac{((n+1)a+c)e^{\theta_{n+1}}}{d+\sum_{i=1}^{n+1}e^{\theta_i}}=0, \tag{8.3}$$

$$b+S_y-\frac{(b+g)e^{\alpha}}{e^{\alpha}+h}-\sum_{i=1}^{n}\frac{se^{\alpha+\theta_i}}{1+e^{\alpha+\theta_i}}=0.$$

Then, defining the constants

$$\tilde{m}_i=\frac{se^{\tilde{\alpha}+\tilde{\theta}_i}}{\left(1+e^{\tilde{\alpha}+\tilde{\theta}_i}\right)^2} \quad, i=1,2,\dots,n, \qquad \tilde{t}=\frac{(b+g)he^{\tilde{\alpha}}}{\left(e^{\tilde{\alpha}}+h\right)^2}+\sum_{i=1}^{n}\tilde{m}_i,$$

$$\tilde{f}_i=\frac{re^{\tilde{\theta}_i}}{\left(1+e^{\tilde{\theta}_i}\right)^2}+\tilde{m}_i+\frac{((n+1)a+c)e^{\tilde{\theta}_i}\left(d+\sum_{j=1,\,j\neq i}^{n+1}e^{\tilde{\theta}_j}\right)}{\left(d+\sum_{i=1}^{n+1}e^{\tilde{\theta}_i}\right)^2} \quad, i=1,2,\dots,n, \tag{8.4}$$

$$\tilde{k}_{ij}=\tilde{k}_{ji}=-\frac{((n+1)a+c)e^{\tilde{\theta}_i+\tilde{\theta}_j}}{\left(d+\sum_{i=1}^{n+1}e^{\tilde{\theta}_i}\right)^2} \quad, i\neq j \quad, i,j=1,2,\dots,n+1,$$

$$\tilde{q}=\frac{re^{\tilde{\theta}_{n+1}}}{\left(1+e^{\tilde{\theta}_{n+1}}\right)^2}+\frac{((n+1)a+c)e^{\tilde{\theta}_{n+1}}\left(d+\sum_{i=1}^{n}e^{\tilde{\theta}_i}\right)}{\left(d+\sum_{i=1}^{n+1}e^{\tilde{\theta}_i}\right)^2},$$

and following (1.5) and (1.6), we derive the approximate posterior distribution for $\left(\theta^n, \theta_{n+1}, \alpha\right)$ and obtain the approximate full conditional distributions to be used in the Gibbs sampling algorithm as being

$$p\left(\theta_i \mid \theta_{j\neq i}, \theta_{n+1}, \alpha, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_i - \frac{(\alpha-\tilde{\alpha})\tilde{m}_i + \sum_{\substack{j=1\\ j\neq i}}^{n+1}\left(\theta_j - \tilde{\theta}_j\right)\tilde{k}_{ij}}{\tilde{f}_i}, \frac{1}{\tilde{f}_i}\right),$$

$$i = 1, 2, \ldots, n,$$

$$p\left(\theta_{n+1} \mid \theta^n, \alpha, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_{n+1} - \frac{\sum_{j=1}^{n}\left(\theta_j - \tilde{\theta}_j\right)\tilde{k}_{n+1,j}}{\tilde{q}}, \frac{1}{\tilde{q}}\right),$$

$$p\left(\alpha \mid \theta^n, \theta_{n+1}, D^n, x_{n+1}\right) = N\left(\tilde{\alpha} - \frac{\sum_{j=1}^{n}\left(\theta_j - \tilde{\theta}_j\right)\tilde{m}_j}{\tilde{t}}, \frac{1}{\tilde{t}}\right).$$

These do not simplify significantly when we consider vague second stage priors.

8.1.2.4. Laplace Approximation

From (1.13) we define

$$-nh\left(\theta^n, \theta_{n+1}, \alpha\right) = \sum_{i=1}^{n}\theta_i\left(a + x_i + y_i\right) + \theta_{n+1}\left(a + x_{n+1}\right) + \alpha\left(b + S_y\right) - \sum_{i=1}^{n+1} r\ln\left(1 + e^{\theta_i}\right) -$$

$$-\sum_{i=1}^{n} s\ln\left(1 + e^{\alpha+\theta_i}\right) - (b+g)\ln\left(e^{\alpha} + h\right) - ((n+1)a + c)\ln\left(d + \sum_{i=1}^{n+1} e^{\theta_i}\right).$$

According to definition (1.16), $\left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}\right) = \boldsymbol{m}$, the solution of the system of equations (8.3), and

$$\tilde{\sigma} = \left\{\tilde{q}\tilde{t}\left|\tilde{A}\right|\right\}^{-1/2},$$

where $\tilde{A}$ is a $(n \times n)$ symmetric matrix whose elements are

$$\tilde{A}_{ii} = \tilde{f}_i - \frac{\tilde{k}^2_{n+1,i}}{\tilde{q}} - \frac{\tilde{m}_i^2}{\tilde{t}} \quad , i = 1, 2, \ldots, n,$$

$$\tilde{A}_{ij} = \tilde{k}_{ij} - \frac{\tilde{k}_{i,n+1}\tilde{k}_{j,n+1}}{\tilde{q}} - \frac{\tilde{m}_i\tilde{m}_j}{\tilde{t}} \quad , i \neq j, \quad i, j = 1, 2, \ldots, n,$$

and all constants are as defined in (8.4). The evaluation of the determinant of $\tilde{A}$ must be done numerically.

Then, as in (1.14), we define a function $h^*(\theta^n, \theta_{n+1}, \alpha)$ such that

$$-n h^*(\theta^n, \theta_{n+1}, \alpha) = (\alpha + \theta_{n+1})y_{n+1} - s\ln(1 + e^{\alpha+\theta_{n+1}}) + \sum_{i=1}^{n}\theta_i(a + x_i + y_i) +$$

$$+\theta_{n+1}(a + x_{n+1}) + \alpha(b + S_y) - \sum_{i=1}^{n+1} r\ln(1 + e^{\theta_i}) - \sum_{i=1}^{n} s\ln(1 + e^{\alpha+\theta_i}) -$$

$$-(b+g)\ln(e^\alpha + h) - ((n+1)a + c)\ln\left(d + \sum_{i=1}^{n+1} e^{\theta_i}\right),$$

and, following the definitions in (1.16), $(\theta^{n*}, \theta^*_{n+1}, \alpha^*)$ will be the solution of the system formed by the equations

$$a + x_i + y_i - \frac{re^{\theta_i}}{1+e^{\theta_i}} - \frac{se^{\alpha+\theta_i}}{1+e^{\alpha+\theta_i}} - \frac{((n+1)a+c)e^{\theta_i}}{d + \sum_{i=1}^{n+1} e^{\theta_i}} = 0 \quad , i = 1, 2, \ldots, n+1,$$

$$b + S_y + y_{n+1} - \frac{(b+g)e^\alpha}{e^\alpha + h} - \sum_{i=1}^{n+1}\frac{se^{\alpha+\theta_i}}{1+e^{\alpha+\theta_i}} = 0.$$

We also have that σ^* will be given by

$$\sigma^* = |A^*|^{-1/2},$$

where A^* is a symmetric matrix with order $(n+2)$ whose elements are

$$A^*_{ii} = f^*_i \quad , i = 1, 2, \ldots, n+1, \qquad A^*_{n+2,n+2} = t^*,$$

$$A^*_{ij} = k^*_{ij} \quad , i \neq j \quad , i, j = 1, 2, \ldots, n+1, \qquad A^*_{i,n+2} = m^*_i \quad , i = 1, 2, \ldots, n+1,$$

with

$$m^*_i = \frac{s e^{\alpha^* + \theta^*_i}}{\left(1 + e^{\alpha^* + \theta^*_i}\right)^2} \quad , i = 1, 2, \ldots, n+1, \qquad t^* = \frac{(b+g) h e^{\alpha^*}}{\left(e^{\alpha^*} + h\right)^2} + \sum_{i=1}^{n+1} m^*_i$$

$$f^*_i = \frac{r e^{\theta^*_i}}{\left(1 + e^{\theta^*_i}\right)^2} + m^*_i + \frac{((n+1)a + c) e^{\theta^*_i} \left(d + \sum_{\substack{j=1 \\ j \neq i}}^{n+1} e^{\theta^*_j}\right)}{\left(d + \sum_{i=1}^{n+1} e^{\theta^*_i}\right)^2} \quad , i = 1, 2, \ldots, n+1,$$

$$k^*_{ij} = -\frac{((n+1)a + c) e^{\theta^*_i + \theta^*_j}}{\left(d + \sum_{i=1}^{n+1} e^{\theta^*_i}\right)^2} \quad , i \neq j \quad , i, j = 1, 2, \ldots, n+1.$$

The determinant of A^*, involved in the definition of σ^*, must be evaluated using a numerical method.

Finally, by (1.17), the Laplace approximation for the predictive distribution of Y_{n+1} will be

$$p\left(y_{n+1} \mid D^n, x_{n+1}\right) \propto \binom{s}{y_{n+1}} \left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\left\{-nh^*\left(\theta^{n*}, \theta^*_{n+1}, \alpha^*\right) + nh\left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}\right)\right\}.$$

No important simplifications will occur taking vague second stage priors.

8.2. Treatments Effects in a Biased Allocation Model

In this section we will consider the 'treatments effects in a biased allocation model' problem assuming that the underlying distributions are binomial. The available treatments are $T_1, T_2, \ldots, T_k$.

Suppose that X_i and Y_i $(i = 1, 2, \ldots, n)$ are random variables representing the countings on the i-th individual before and after a treatment is applied. We assume that

- given r and θ_i, $X_i \sim Bi\left(r, \dfrac{e^{\theta_i}}{1+e^{\theta_i}}\right)$,

- given s, θ_i, $\alpha_1, \alpha_2, \ldots, \alpha_k$ and x_i, $Y_i \sim Bi\left(s, \dfrac{e^{\theta_i+\alpha_j}}{1+e^{\theta_i+\alpha_j}}\right)$, if treatment T_j is used,

- given x_i, treatment T_j is used if and only if $x_i \in C_j$,

where $C_1, C_2, \ldots, C_k$ are defined as in section 3.4.

In the model above $\theta_1, \theta_2, \ldots, \theta_n$ are parameters used to model the individuals' particular characteristics and $\alpha_1, \alpha_2, \ldots, \alpha_k$ are parameters which model the treatments effects upon the individuals.

Let $Y_{n+1,j}$ $(j = 1, 2, \ldots, k)$ be random variables representing the outcome on the $(n+1)$-th individual after T_j is applied. We assume that

$$X_{n+1} \sim Bi\left(r, \frac{e^{\theta_{n+1}}}{1+e^{\theta_{n+1}}}\right),$$

$$Y_{n+1,j} \sim Bi\left(s, \frac{e^{\theta_{n+1}+\alpha_j}}{1+e^{\theta_{n+1}+\alpha_j}}\right) \quad , j = 1, 2, \ldots, k.$$

Our aim is to derive the predictive distribution of $Y_{n+1,j}$. Let us define two parameter vectors $\theta^n = (\theta_1, \theta_2, \ldots, \theta_n)$ and $\alpha^k = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, an indicator function

$$\delta_{ij} = \begin{cases} 1, & \text{if treatment } T_j \text{ is applied to individual } i \\ 0, & \text{otherwise} \end{cases}$$

and n_j is the number of individuals who received treatment T_j, that is,

$$n_j = \sum_{i=1}^{n} \delta_{ij} .$$

The predictions will be based on a data set

$$D^n = \{(x_i, y_i), i = 1, 2, ..., n\}$$

and on an observed x_{n+1}.

8.2.1. A Classical Approach

Given the likelihood function

$$L(\theta^n, \theta_{n+1}, \alpha^k; D^n, x_{n+1}) \propto \frac{\exp\left\{\sum_{i=1}^{n+1} \theta_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{k} (\theta_i + \alpha_j) y_i \delta_{ij}\right\}}{\prod_{i=1}^{n+1}\left\{(1+e^{\theta_i})^r\right\}\prod_{i=1}^{n}\prod_{j=1}^{k}\left\{(1+e^{\theta_i+\alpha_j})^{s\delta_{ij}}\right\}},$$

the maximum likelihood estimates $\hat{\theta}^n$ and $\hat{\alpha}^k$ are obtained numerically solving the system formed by the $(n+k)$ equations

$$x_i + y_i - \frac{re^{\theta_i}}{1+e^{\theta_i}} - s\sum_{j=1}^{k}\left\{\frac{\delta_{ij} e^{\theta_i+\alpha_j}}{1+e^{\theta_i+\alpha_j}}\right\} = 0 \quad , i = 1, 2, ..., n,$$

$$\sum_{i=1}^{n} y_i \delta_{ij} - s\sum_{i=1}^{n}\left\{\frac{\delta_{ij} e^{\theta_i+\alpha_j}}{1+e^{\theta_i+\alpha_j}}\right\} = 0 \quad , j = 1, 2, ..., k, \tag{8.5}$$

whilst

$$\hat{\theta}_{n+1} = \ln\left(\frac{x_{n+1}}{r - x_{n+1}}\right). \tag{8.6}$$

Then, a simple estimate for the predictive distribution of the random variable $Y_{n+1,j}$ would be

$$Y_{n+1,j} \sim Bi\left(s, \frac{e^{\hat{\theta}_{n+1}+\hat{\alpha}_j}}{1+e^{\hat{\theta}_{n+1}+\hat{\alpha}_j}}\right).$$

Note that this approach is not possible if $x_{n+1} = 0$ or if $x_{n+1} = r$.

8.2.2. A Bayesian Approach

Let us consider a hierarchical prior structure in which, at the first stage we take

$$p\left(\theta^n, \theta_{n+1}, \alpha^k \mid \xi, \eta^k\right) = \prod_{i=1}^{n+1}\left\{p(\theta_i \mid \xi)\right\}\prod_{j=1}^{k}\left\{p(\alpha_j \mid \eta_j)\right\}$$

and at the second stage we take

$$p(\xi, \eta^k) = p(\xi)\prod_{j=1}^{k}\left\{p(\eta_j)\right\},$$

where $\eta^k = (\eta_1, \eta_2, \ldots, \eta_k)$. We will assume that

$$e^{\theta_i} \sim Ga\left(a, e^{\xi}\right) \quad, i = 1, 2, \ldots, n+1 \qquad e^{\alpha_j} \sim Ga\left(b_j, e^{\eta_j}\right) \quad, j = 1, 2, \ldots, k$$

$$e^{\xi} \sim Ga(c, d) \qquad e^{\eta_j} \sim Ga\left(g_j, h_j\right) \quad, j = 1, 2, \ldots, k$$

The joint posterior distribution is given by

$$p\left(\theta^n, \theta_{n+1}, \alpha^k, \xi, \eta^k \mid D^n, x_{n+1}\right) \propto \frac{\exp\left\{\sum_{i=1}^{n+1}\theta_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\theta_i + \alpha_j\right)y_i\delta_{ij}\right\}}{\prod_{i=1}^{n+1}\left\{\left(1+e^{\theta_i}\right)^r\right\}\prod_{i=1}^{n}\prod_{j=1}^{k}\left\{\left(1+e^{\theta_i+\alpha_j}\right)^{s\delta_{ij}}\right\}} \times$$

$$\times \exp\left\{-\sum_{i=1}^{n+1}e^{\xi+\theta_i} - \sum_{j=1}^{k}e^{\eta_j+\alpha_j} - \sum_{j=1}^{k}h_j e^{\eta_j} - d e^{\xi}\right\} \times$$

$$\times \exp\left\{\sum_{i=1}^{n+1}a\left(\xi+\theta_i\right) + \sum_{j=1}^{k}b_j\left(\eta_j + \alpha_j\right) + \sum_{j=1}^{k}g_j\,\eta_j + c\,\xi\right\}. \tag{8.7}$$

8.2.2.1. The Exact Predictive Distribution

The exact predictive distribution of $Y_{n+1,j}$ is given by

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) = \int_{\Re^2} p\left(y_{n+1,j} \mid \theta_{n+1}, \alpha_j\right) p\left(\theta_{n+1}, \alpha_j \mid D^n, x_{n+1}\right) d\theta_{n+1} d\alpha_j .$$

Hence, the problem would become much easier if we could derive from (8.7) the marginal posterior distribution of $\left(\theta_{n+1}, \alpha_j\right)$; however, the best we can do is to eliminate ξ and η^k to obtain

$$p\left(\theta^n, \theta_{n+1}, \alpha^k \mid D^n, x_{n+1}\right) \propto \frac{\exp\left\{\sum_{i=1}^{n+1} \theta_i (x_i + a) + \sum_{j=1}^{k} b_j \alpha_j\right\}}{\left(d + \sum_{i=1}^{n+1} e^{\theta_i}\right)^{(n+1)a+c} \prod_{i=1}^{n+1} \left\{\left(1 + e^{\theta_i}\right)^r\right\} \prod_{j=1}^{k} \left\{\left(e^{\alpha_j} + h_j\right)^{b_j + g_j}\right\}} \times$$

$$\times \frac{\exp\left\{\sum_{i=1}^{n} \sum_{j=1}^{k} \left(\theta_i + \alpha_j\right) y_i \delta_{ij}\right\}}{\prod_{i=1}^{n} \prod_{j=1}^{k} \left\{\left(1 + e^{\theta_i + \alpha_j}\right)^{s\delta_{ij}}\right\}}, \qquad (8.8)$$

and the exact predictive distribution of $Y_{n+1,j}$ will be

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) \propto \binom{s}{y_{n+1,j}} \int_{\Re^{n+k+1}} \frac{\exp\left\{\left(\theta_{n+1} + \alpha_j\right) y_{n+1,j}\right\}}{\left(1 + e^{\theta_{n+1} + \alpha_j}\right)^s} \times$$

$$\times\, p\left(\theta^n, \theta_{n+1}, \alpha^k \mid D^n, x_{n+1}\right) d\theta^n d\theta_{n+1} d\alpha^k, \qquad (8.9)$$

which involves the evaluation of a $(n+k+1)$-dimensional integral requiring the use of a numerical integration technique. Because of the high dimensionality we might expect numerical problems. Such problems would not disappear if we would consider vague second stage priors in the hierarchical prior structure.

8.2.2.2. Estimation Via Gibbs Sampling

The Gibbs routine (section 1.2.1) can be used to estimate the predictive probabilities of $Y_{n+1,j}$ overcoming the numerical problems that may arise in (8.9). The full conditional distributions to be used are

$$p\left(\theta_i \mid \theta_{j\neq i}, \theta_{n+1}, \alpha^k, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\theta_i\left(x_i + y_i + a\right)\right\}}{\left(d + \sum_{i=1}^{n+1} e^{\theta_i}\right)^{(n+1)a+c} \left(1 + e^{\theta_i}\right)^r \prod_{j=1}^{k}\left\{\left(1 + e^{\theta_i + \alpha_j}\right)^{s\delta_{ij}}\right\}},$$

$$i = 1, 2, \ldots, n,$$

$$p\left(\theta_{n+1} \mid \theta^n, \alpha^k, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\theta_{n+1}\left(x_{n+1} + a\right)\right\}}{\left(d + \sum_{i=1}^{n+1} e^{\theta_i}\right)^{(n+1)a+c} \left(1 + e^{\theta_{n+1}}\right)^r},$$

$$p\left(\alpha_j \mid \theta^n, \theta_{n+1}, \alpha_{i\neq j}, D^n, x_{n+1}\right) \propto \frac{\exp\left\{\alpha_j\left(b_j + \sum_{i=1}^{n} y_i \delta_{ij}\right)\right\}}{\left(e^{\alpha_j} + h_j\right)^{b_j + g_j} \prod_{i=1}^{n}\left\{\left(1 + e^{\theta_i + \alpha_j}\right)^{s\delta_{ij}}\right\}}, \quad j = 1, 2, \ldots, k,$$

which do not simplify greatly assuming vague second stage priors. The random generation of all necessary values for the Gibbs routine requires a sampling technique such as the rejection sampling algorithm (section 1.2.3).

Again problems are likely because of the high dimensionality.

8.2.2.3. Estimation Via Asymptotic Results

In this section we will use the asymptotic results summarised in section 1.2.2 in order to approximate the posterior distribution by a multivariate normal distribution. Then, we easily obtain the approximate full conditional distributions to be used when estimating the predictive probabilities through the Gibbs routine (section 1.2.1).

8.2.2.3.1. Posterior Normality Based on the Likelihood Function

Following Bernardo & Smith (1994), we find that θ_{n+1}, α_1, α_2, ..., α_k are independent a posteriori with

$$p\left(\theta_{n+1} \mid D^n, x_{n+1}\right) = N\left(\hat{\theta}_{n+1}, \frac{r}{x_{n+1}\left(r - x_{n+1}\right)}\right),$$

$$p\left(\alpha_j \mid D^n, x_{n+1}\right) = N\left(\hat{\alpha}_j, \frac{1}{v_j}\right) \quad , j = 1, 2, \ldots, k,$$

where $\hat{\theta}_i$ $(i = 1, 2, \ldots, n)$, $\hat{\theta}_{n+1}$ and $\hat{\alpha}_j$ $(j = 1, 2, \ldots, k)$ are the maximum likelihood estimates, given by (8.5) and (8.6), and

$$v_j = u_j - \sum_{i=1}^{n} \frac{m_{ij}^2}{f_j} \quad , j = 1, 2, \ldots, k,$$

with

$$m_{ij} = \frac{s\,\delta_{ij}\, e^{\hat{\theta}_i + \hat{\alpha}_j}}{\left(1 + e^{\hat{\theta}_i + \hat{\alpha}_j}\right)^2} \quad , i = 1, 2, \ldots, n \quad ; j = 1, 2, \ldots, k,$$

$$f_i = \frac{r e^{\hat{\theta}_i}}{\left(1 + e^{\hat{\theta}_i}\right)^2} + \sum_{j=1}^{k} m_{ij} \quad , i = 1, 2, \ldots, n, \qquad u_j = \sum_{i=1}^{n} m_{ij} \quad , j = 1, 2, \ldots, k.$$

Problems arise when $x_{n+1} = 0$ or $x_{n+1} = r$.

8.2.2.3.2. Posterior Normality Based on Characteristics of the Posterior Distribution

O'Hagan (1994) suggested another asymptotic approximation for the posterior distribution (8.8) based on the posterior mode and on the modal dispersion matrix (section 1.2.2.2). The posterior mode $m = \left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}^k\right)$ is the solution of the system formed by the $(n+k+1)$ equations

$$x_i + y_i + a - \frac{((n+1)a+c)e^{\theta_i}}{d+\sum_{i=1}^{n+1} e^{\theta_i}} - \frac{re^{\theta_i}}{1+e^{\theta_i}} - s\sum_{j=1}^{k}\left\{\frac{\delta_{ij}e^{\theta_i+\alpha_j}}{1+e^{\theta_i+\alpha_j}}\right\} = 0 \quad , i = 1, 2, \ldots, n,$$

$$x_{n+1} + a - \frac{((n+1)a+c)e^{\theta_{n+1}}}{d+\sum_{i=1}^{n+1} e^{\theta_i}} - \frac{re^{\theta_{n+1}}}{1+e^{\theta_{n+1}}} = 0, \tag{8.10}$$

$$b_j + \sum_{i=1}^{n} y_i\delta_{ij} - \frac{(b_j+g_j)e^{\alpha_j}}{e^{\alpha_j}+b_j} - s\sum_{i=1}^{n}\left\{\frac{\delta_{ij}e^{\theta_i+\alpha_j}}{1+e^{\theta_i+\alpha_j}}\right\} = 0 \quad , j = 1, 2, \ldots, k,$$

which must be obtained using a numerical technique. Then, defining the constants

$$\tilde{t}_{ij} = \frac{s\delta_{ij}e^{\tilde{\theta}_i+\tilde{\alpha}_j}}{\left(1+e^{\tilde{\theta}_i+\tilde{\alpha}_j}\right)^2} \quad , i = 1, 2, \ldots, n \quad ; j = 1, 2, \ldots, k,$$

$$\tilde{f}_i = \frac{((n+1)a+c)e^{\tilde{\theta}_i}\left(d+\sum_{\substack{p=1\\p\neq i}}^{n+1} e^{\tilde{\theta}_p}\right)}{\left(d+\sum_{p=1}^{n+1} e^{\tilde{\theta}_p}\right)^2} + \frac{re^{\tilde{\theta}_i}}{\left(1+e^{\tilde{\theta}_i}\right)^2} + \sum_{j=1}^{k}\tilde{t}_{ij} \quad , i = 1, 2, \ldots, n,$$

$$\tilde{q}_{il} = -\frac{((n+1)a+c)e^{\tilde{\theta}_i+\tilde{\theta}_l}}{\left(d+\sum_{p=1}^{n+1} e^{\tilde{\theta}_p}\right)^2} \quad , i \neq l \quad , i, l = 1, 2, \ldots, n+1, \tag{8.11}$$

$$\tilde{u} = \frac{((n+1)a+c)e^{\tilde{\theta}_{n+1}}\left(d+\sum_{i=1}^{n} e^{\tilde{\theta}_i}\right)}{\left(d+\sum_{i=1}^{n+1} e^{\tilde{\theta}_i}\right)^2} + \frac{re^{\tilde{\theta}_{n+1}}}{\left(1+e^{\tilde{\theta}_{n+1}}\right)^2},$$

$$\tilde{v}_j = \frac{(b_j+g_j)h_je^{\tilde{\alpha}_j}}{\left(e^{\tilde{\alpha}_j}+h_j\right)^2} + \sum_{i=1}^{n}\tilde{t}_{ij} \quad , j = 1, 2, \ldots, k,$$

the predictive distribution of $Y_{n+1,j}$ is estimated through the Gibbs routine using the following full conditional distributions

$$p\left(\theta_i \mid \theta_{j\neq i}, \theta_{n+1}, \alpha^k, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_i - \frac{\sum_{p=1}^{k}\left(\alpha_p - \tilde{\alpha}_p\right)\tilde{t}_{ip} + \sum_{\substack{m=1\\m\neq i}}^{n+1}\left(\theta_m - \tilde{\theta}_m\right)\tilde{q}_{im}}{\tilde{f}_i}, \frac{1}{\tilde{f}_i}\right),$$

$$i = 1, 2, \ldots, n,$$

$$p\left(\theta_{n+1} \mid \theta^n, \alpha^k, D^n, x_{n+1}\right) = N\left(\tilde{\theta}_{n+1} - \frac{\sum_{p=1}^{n}\left(\theta_p - \tilde{\theta}_p\right)\tilde{q}_{n+1,p}}{\tilde{u}}, \frac{1}{\tilde{u}}\right),$$

$$p\left(\alpha_j \mid \theta^n, \theta_{n+1}, \alpha_{i\neq j}, D^n, x_{n+1}\right) = N\left(\tilde{\alpha}_j - \frac{\sum_{p=1}^{n}\left(\theta_p - \tilde{\theta}_p\right)\tilde{t}_{pj}}{\tilde{v}_j}, \frac{1}{\tilde{v}_j}\right), \quad j = 1, 2, \ldots, k.$$

8.2.2.4. Laplace Approximation

The Laplace approximation for the predictive distribution of $Y_{n+1,j}$ is

$$p\left(y_{n+1,j} \mid D^n, x_{n+1}\right) \propto \binom{s}{y_{n+1,j}}\left(\frac{\sigma^*}{\tilde{\sigma}}\right) \exp\left\{-nh^*\left(\theta^{n*}, \theta^*_{n+1}, \alpha^{k*}\right) + nh\left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}^k\right)\right\}.$$

The functions $h\left(\theta^n, \theta_{n+1}, \alpha^k\right)$ and $h^*\left(\theta^n, \theta_{n+1}, \alpha^k\right)$ are defined as in (1.13) and (1.14) by

$$-nh\left(\theta^n, \theta_{n+1}, \alpha^k\right) = \sum_{i=1}^{n+1}\theta_i\left(x_i + a\right) + \sum_{j=1}^{k} b_j\alpha_j + \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\theta_i + \alpha_j\right)y_i\delta_{ij} -$$

$$-\left(\left(n+1\right)a + c\right)\ln\left(d + \sum_{i=1}^{n+1}e^{\theta_i}\right) - \sum_{i=1}^{n+1}\left\{r\ln\left(1 + e^{\theta_i}\right)\right\} -$$

$$-\sum_{j=1}^{k}\left\{\left(b_j+g_j\right)\ln\left(e^{\alpha_j}+h_j\right)\right\}-\sum_{i=1}^{n}\sum_{j=1}^{k}\left\{s\delta_{ij}\ln\left(1+e^{\theta_i+\alpha_j}\right)\right\}$$

and

$$-nh^{*}\left(\theta^{n},\theta_{n+1},\alpha^{k}\right)=\left(\theta_{n+1}+\alpha_j\right)y_{n+1,j}-s\ln\left(1+e^{\theta_{n+1}+\alpha_j}\right)+$$

$$+\sum_{i=1}^{n+1}\theta_i\left(x_i+a\right)+\sum_{j=1}^{k}b_j\alpha_j+\sum_{i=1}^{n}\sum_{j=1}^{k}\left(\theta_i+\alpha_j\right)y_i\delta_{ij}-$$

$$-\left((n+1)a+c\right)\ln\left(d+\sum_{i=1}^{n+1}e^{\theta_i}\right)-\sum_{i=1}^{n+1}\left\{r\ln\left(1+e^{\theta_i}\right)\right\}-$$

$$-\sum_{j=1}^{k}\left\{\left(b_j+g_j\right)\ln\left(e^{\alpha_j}+h_j\right)\right\}-\sum_{i=1}^{n}\sum_{j=1}^{k}\left\{s\delta_{ij}\ln\left(1+e^{\theta_i+\alpha_j}\right)\right\}.$$

Solving the system of equations (8.10), through a numerical method, we obtain $\left(\tilde{\theta}^{n}, \tilde{\theta}_{n+1}, \tilde{\alpha}^{k}\right)$ and $\left(\theta^{n*}, \theta^{*}_{n+1}, \alpha^{k*}\right)$ is derived solving numerically the system formed by the equations

$$x_i+y_i+a-\frac{\left((n+1)a+c\right)e^{\theta_i}}{d+\sum_{p=1}^{n+1}e^{\theta_p}}-\frac{re^{\theta_i}}{1+e^{\theta_i}}-s\sum_{p=1}^{k}\left\{\frac{\delta_{ip}e^{\theta_i+\alpha_p}}{1+e^{\theta_i+\alpha_p}}\right\}=0 \quad ,i=1,2,\ldots,n,$$

$$x_{n+1}+y_{n+1,j}+a-\frac{\left((n+1)a+c\right)e^{\theta_{n+1}}}{d+\sum_{p=1}^{n+1}e^{\theta_p}}-\frac{re^{\theta_{n+1}}}{1+e^{\theta_{n+1}}}-\frac{se^{\theta_{n+1}+\alpha_j}}{1+e^{\theta_{n+1}+\alpha_j}}=0,$$

$$b_j+\sum_{i=1}^{n}y_i\delta_{ij}+y_{n+1,j}-\frac{\left(b_j+g_j\right)e^{\alpha_j}}{e^{\alpha_j}+h_j}-s\sum_{i=1}^{n}\left\{\frac{\delta_{ij}e^{\theta_i+\alpha_j}}{1+e^{\theta_i+\alpha_j}}\right\}-\frac{se^{\theta_{n+1}+\alpha_j}}{1+e^{\theta_{n+1}+\alpha_j}}=0,$$

$$b_m+\sum_{i=1}^{n}y_i\delta_{im}-\frac{\left(b_m+g_m\right)e^{\alpha_m}}{e^{\alpha_m}+h_m}-s\sum_{i=1}^{n}\left\{\frac{\delta_{im}e^{\theta_i+\alpha_m}}{1+e^{\theta_i+\alpha_m}}\right\}=0 \quad ,m\neq j \quad ,\ m=1,2,\ldots,k.$$

Following (1.16), $\tilde{\sigma}$ is defined by

$$\tilde{\sigma} = \left\{ \tilde{u} \prod_{p=1}^{k} \left\{ \tilde{v}_p \right\} \left| \tilde{A} \right| \right\}^{-1/2},$$

where $\tilde{A}$ is a symmetric matrix whose elements are

$$\tilde{A}_{ii} = \tilde{f}_i - \frac{\tilde{q}^2_{i,n+1}}{\tilde{u}} - \sum_{p=1}^{k} \frac{\tilde{t}^2_{ip}}{\tilde{v}_p} \quad , i = 1, 2, \ldots, n,$$

$$\tilde{A}_{ij} = \tilde{q}_{ij} - \frac{\tilde{q}_{i,n+1}\tilde{q}_{j,n+1}}{\tilde{u}} - \sum_{p=1}^{k} \frac{\tilde{t}_{ip}\tilde{t}_{jp}}{\tilde{v}_p} \quad , i \neq j, \quad i, j = 1, 2, \ldots, n,$$

with $\tilde{u}$, $\tilde{v}_p$ $(p = 1, 2, \ldots, k)$, $\tilde{f}_i$ $(i = 1, 2, \ldots, n)$, $\tilde{q}_{ij}$ $(i = 1, 2, \ldots, n+1; j = 1, 2, \ldots, n+1)$ and $\tilde{t}_{ij}$ $(i = 1, 2, \ldots, n; j = 1, 2, \ldots k)$ being the constants defined in (8.11). Still from (1.16), σ^* is given by

$$\sigma^* = \left\{ \left(u^* w^* - z^{*2}\right) \prod_{\substack{p=1 \\ p \neq j}}^{k} \left\{ v^*_p \right\} \left| A^* \right| \right\}^{-1/2},$$

where A^* is a symmetric matrix whose elements are

$$A^*_{ii} = \frac{q^{*2}_{i,n+1} w^* - 2t^*_{ij} z^* q^*_{i,n+1} + t^{*2}_{ij} u^*}{u^* w^* - z^{*2}} + \sum_{\substack{p=1 \\ p \neq j}}^{k} \frac{t^{*2}_{ip}}{v^*_p} \quad , i = 1, 2, \ldots, n,$$

$$A^*_{il} = \frac{q^*_{i,n+1} q^*_{l,n+1} w^* + t^*_{ij} t^*_{lj} u^* - z^* \left(t^*_{ij} q^*_{l,n+1} + q^*_{i,n+1} t^*_{lj} \right)}{u^* w^* - z^{*2}} + \sum_{\substack{p=1 \\ p \neq j}}^{k} \frac{t^*_{ip} t^*_{lp}}{v^*_p} \quad , i \neq l \quad , i, l = 1, 2, \ldots, n,$$

with

$$q^*_{il} = -\frac{\left((n+1)a + c\right) e^{\theta^*_i + \theta^*_l}}{\left(d + \sum_{p=1}^{n+1} e^{\theta^*_p} \right)^2} \quad , i \neq l, \quad i, l = 1, 2, \ldots, n+1,$$

$$t^*_{im} = \frac{s\, \delta_{im}\, e^{\theta^*_i + \alpha^*_m}}{\left(1 + e^{\theta^*_i + \alpha^*_m} \right)^2} \quad , i = 1, 2, \ldots, n; m = 1, 2, \ldots, k,$$

$$u^* = \frac{((n+1)a+c)e^{\theta^*_{n+1}}\left(d+\sum_{i=1}^{n}e^{\theta^*_i}\right)}{\left(d+\sum_{i=1}^{n+1}e^{\theta^*_i}\right)^2} + \frac{re^{\theta^*_{n+1}}}{\left(1+e^{\theta^*_{n+1}}\right)^2} + \frac{se^{\theta^*_{n+1}+\alpha^*_j}}{\left(1+e^{\theta^*_{n+1}+\alpha^*_j}\right)^2},$$

$$v^*_m = \frac{(b_m+g_m)h_m e^{\alpha^*_m}}{\left(e^{\alpha^*_m}+h_m\right)^2} + \sum_{i=1}^{n} t^*_{im} \quad , m \neq j \quad , m = 1, 2, \ldots, k,$$

$$w^* = \frac{(b_j+g_j)h_j e^{\alpha^*_j}}{\left(e^{\alpha^*_j}+h_j\right)^2} + \sum_{i=1}^{n} t^*_{ij} + \frac{se^{\theta^*_{n+1}+\alpha^*_j}}{\left(1+e^{\theta^*_{n+1}+\alpha^*_j}\right)^2},$$

$$z^* = \frac{se^{\theta^*_{n+1}+\alpha^*_j}}{\left(1+e^{\theta^*_{n+1}+\alpha^*_j}\right)^2}.$$

The implementation of this approximation may be somehow difficult because of the high-dimensional determinants which must be evaluated numerically. The assumption of vague second stage priors does not simplify much.

8.3. A Crossover Design to Compare Two Treatments

The crossover design for the binomial model corresponding to Chapter 4 can be modelled by

$$\text{Period } 1: \quad W_{i1} \sim Bi\left(r(1-\delta_{i1}) + s\delta_{i1}, \frac{e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}\right)$$

$$\text{Period } 2: \quad W_{i2} \sim Bi\left(r(1-\delta_{i2}) + s\delta_{i2}, \frac{e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right),$$

where δ_{i1} and δ_{i2} are the treatment indicators defined in (4.1).

Our final goal is to derive the predictive probabilities for the outcomes on a new individual when treatment T_1 or T_2 is applied. Those outcomes are represented by the random variables Z_1 and Z_2. We assume that, conditional on θ_{n+1}, Z_1 and Z_2 are independent random variables such that

$$Z_1 \sim Bi\left(r, \frac{e^{\theta_{n+1}}}{1+e^{\theta_{n+1}}}\right),$$

$$Z_2 \sim Bi\left(s, \frac{e^{\alpha+\theta_{n+1}}}{1+e^{\alpha+\theta_{n+1}}}\right).$$

The predictions will be made based on a data set D^n formed by the information about the outcomes for each individual in both periods as well as by information about the used treatment, that is

$$D^n = \left\{(w_{i1}, w_{i2}, \delta_{i1}, \delta_{i2}),\ i = 1, 2, \ldots, n\right\}.$$

8.3.1. A Classical Approach

The likelihood function is

$$L(\theta^n, \theta_{n+1}, \alpha, \beta; D^n) \propto \frac{\exp\left\{\sum_{i=1}^{n}(\theta_i + \delta_{i1}\alpha)w_{i1} + \sum_{i=1}^{n}(\theta_i + \beta + \delta_{i2}\alpha)w_{i2}\right\}}{\prod_{i=1}^{n}\left\{\left(1+e^{\theta_i+\delta_{i1}\alpha}\right)^{r(1-\delta_{i1})+s\delta_{i1}}\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)^{r(1-\delta_{i2})+s\delta_{i2}}\right\}}$$

and the maximum likelihood estimates $\hat{\theta}^n$, $\hat{\alpha}$ and $\hat{\beta}$ will be obtained solving numerically the system formed by the $(n+2)$ equations

$$w_{i1} + w_{i2} - \frac{\left(r(1-\delta_{i1}) + s\delta_{i1}\right)e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}} - \frac{\left(r(1-\delta_{i2}) + s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}} = 0\ ,\ i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n}\delta_{i1}w_{i1} + \sum_{i=1}^{n}\delta_{i2}w_{i2} - s\sum_{i=1}^{n}\left\{\frac{\delta_{i1}e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}\right\} - s\sum_{i=1}^{n}\left\{\frac{\delta_{i2}e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\} = 0$$

$$\sum_{i=1}^{n}w_{i2} - \sum_{i=1}^{n}\left\{\frac{\left(r(1-\delta_{i2}) + s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\} = 0.$$

Since D^n contains no information about the $(n+1)$-th individual, we are unable of obtaining the maximum likelihood estimate for θ_{n+1}, so that plug-in estimates of the predictive distributions are not obtainable.

8.3.2. A Bayesian Approach

Let us consider a hierarchical prior structure, taking at the first stage

$$p(\theta^n, \theta_{n+1}, \alpha, \beta \mid \xi, \eta, \zeta) = \prod_{i=1}^{n+1} \{p(\theta_i \mid \xi)\} p(\alpha \mid \eta) p(\beta \mid \zeta)$$

and taking at the second stage

$$p(\xi, \eta, \zeta) = p(\xi)p(\eta)p(\zeta).$$

We assume that

$$e^{\theta_i} \sim Ga(k, e^{\xi}) \qquad e^{\alpha} \sim Ga(h, e^{\eta}) \qquad e^{\beta} \sim Ga(g, e^{\zeta})$$

$$e^{\xi} \sim Ga(l, m) \qquad e^{\eta} \sim Ga(q, t) \qquad e^{\zeta} \sim Ga(u, v)$$

with k, h, g, l, m, q, t, u and v specified.

The joint posterior distribution is given by

$$p(\theta^n, \theta_{n+1}, \alpha, \beta, \xi, \eta, \zeta \mid D^n) \propto \exp\left\{-e^{\xi}\left(\sum_{i=1}^{n+1} e^{\theta_i} + m\right) - e^{\eta}(e^{\alpha} + t) - e^{\zeta}(e^{\beta} + v)\right\} \times$$

$$\times \exp\left\{((n+1)k + l)\xi + k\theta_{n+1} + (k+2)\sum_{i=1}^{n} \theta_i + (q+h)\eta + (g+u)\zeta\right\} \times$$

$$\times \frac{\exp\left\{\alpha\left(h + \sum_{i=1}^{n} \delta_{i1} w_{i1} + \sum_{i=1}^{n} \delta_{i2} w_{i2}\right) + \beta\left(g + \sum_{i=1}^{n} w_{i2}\right)\right\}}{\prod_{i=1}^{n}\left\{\left(1 + e^{\theta_i + \delta_{i1}\alpha}\right)^{r(1-\delta_{i1}) + s\delta_{i1}} \left(1 + e^{\theta_i + \beta + \delta_{i2}\alpha}\right)^{r(1-\delta_{i2}) + s\delta_{i2}}\right\}}.$$

After eliminating the hyperparameters ξ, η and ζ, we obtain the marginal posterior distribution

$$p\left(\theta^n, \theta_{n+1}, \alpha, \beta \mid D^n\right) \propto \frac{\exp\left\{(k+2)\sum_{i=1}^{n}\theta_i + k\theta_{n+1}\right\}}{\prod_{i=1}^{n}\left\{\left(1+e^{\theta_i+\delta_{i1}\alpha}\right)^{r(1-\delta_{i1})+s\delta_{i1}}\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)^{r(1-\delta_{i2})+s\delta_{i2}}\right\}} \times$$

$$\times \frac{\exp\left\{\alpha\left(h+\sum_{i=1}^{n}\delta_{i1}w_{i1}+\sum_{i=1}^{n}\delta_{i2}w_{i2}\right)+\beta\left(g+\sum_{i=1}^{n}w_{i2}\right)\right\}}{\left(\sum_{i=1}^{n+1}e^{\theta_i}+m\right)^{(n+1)k+1}\left(e^{\alpha}+t\right)^{q+h}\left(e^{\beta}+v\right)^{g+u}} . \quad (8.12)$$

8.3.2.1. The Exact Predictive Distributions

Given the posterior distribution (8.12) for $\left(\theta^n, \theta_{n+1}, \alpha, \beta\right)$, the exact predictive distributions of $\left(Z_1, Z_2\right)$, Z_1 and Z_2 will be given, respectively, by

$$p\left(z_1, z_2 \mid D^n\right) \propto \binom{r}{z_1}\binom{s}{z_2}\int_{\Re^{n+3}} \frac{\exp\left\{\theta_{n+1}\left(z_1+z_2\right)+\alpha z_2\right\}}{\left(1+e^{\theta_{n+1}}\right)^r\left(1+e^{\alpha+\theta_{n+1}}\right)^s} \times$$

$$\times p\left(\theta^n, \theta_{n+1}, \alpha, \beta \mid D^n\right) d\theta^n d\theta_{n+1} d\alpha\, d\beta, \quad (8.13)$$

$$p\left(z_1 \mid D^n\right) \propto \binom{r}{z_1}\int_{\Re^{n+3}} \frac{\exp\left\{\theta_{n+1} z_1\right\}}{\left(1+e^{\theta_{n+1}}\right)^r}\, p\left(\theta^n, \theta_{n+1}, \alpha, \beta \mid D^n\right) d\theta^n d\theta_{n+1} d\alpha\, d\beta, \quad (8.14)$$

$$p\left(z_2 \mid D^n\right) \propto \binom{s}{z_2}\int_{\Re^{n+3}} \frac{\exp\left\{\left(\alpha+\theta_{n+1}\right) z_2\right\}}{\left(1+e^{\alpha+\theta_{n+1}}\right)^s}\, p\left(\theta^n, \theta_{n+1}, \alpha, \beta \mid D^n\right) d\theta^n d\theta_{n+1} d\alpha\, d\beta \quad (8.15)$$

which evaluation requires the use of a numerical integration technique to solve the $(n+3)$-dimensional integrals involved.

Assuming vague second stage priors, we will not get significant simplifications.

8.3.2.2. Estimation Via Gibbs Sampling

Because of the high dimensionality of the integrals in (8.13), (8.14) and (8.15) we would like to avoid their evaluation. The Gibbs sampling algorithm (section 1.2.1) is one of the possible ways of doing that. From (8.12) we derive the full conditional distributions to be used. They are

$$p\left(\theta_i \mid \theta_{j\neq i}, \theta_{n+1}, \alpha, \beta, D^n\right) \propto \frac{\exp\{(k+2)\theta_i\}}{\left(1+e^{\theta_i+\delta_{i1}\alpha}\right)^{r(1-\delta_{i1})+s\delta_{i1}}\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)^{r(1-\delta_{i2})+s\delta_{i2}}} \times$$

$$\times \frac{1}{\left(\sum_{p=1}^{n+1} e^{\theta_p} + m\right)^{(n+1)k+1}} \quad , i = 1, 2, \ldots, n,$$

$$p\left(\theta_{n+1} \mid \theta^n, \alpha, \beta, D^n\right) \propto \frac{\exp\{k\theta_{n+1}\}}{\left(\sum_{i=1}^{n+1} e^{\theta_i} + m\right)^{(n+1)k+1}} ,$$

$$p\left(\alpha \mid \theta^n, \theta_{n+1}, \beta, D^n\right) \propto \frac{\exp\left\{\alpha\left(h + \sum_{i=1}^{n}\delta_{i1}w_{i1} + \sum_{i=1}^{n}\delta_{i2}w_{i2}\right)\right\}}{\prod_{i=1}^{n}\left\{\left(1+e^{\theta_i+\delta_{i1}\alpha}\right)^{r(1-\delta_{i1})+s\delta_{i1}}\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)^{r(1-\delta_{i2})+s\delta_{i2}}\right\}} \times$$

$$\times \frac{1}{\left(e^{\alpha}+t\right)^{q+h}} ,$$

$$p\left(\beta \mid \theta^n, \theta_{n+1}, \alpha, D^n\right) \propto \frac{\exp\left\{\beta\left(g + \sum_{i=1}^{n} w_{i2}\right)\right\}}{\prod_{i=1}^{n}\left\{\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)^{r(1-\delta_{i2})+s\delta_{i2}}\right\}\left(e^{\beta}+v\right)^{g+u}} ,$$

and they do not simplify much when we consider the vague second stage priors case.

Note that the random generation of values of θ_{n+1} is very easy since its full conditional distribution is a transformed beta distribution (see Table A1.2). The rejection sampling algorithm (section 1.2.3) will be required to generate the values for all the other parameters.

8.3.2.3. Estimation Via Asymptotic Results

The asymptotic result we summarised in section 1.2.2.1, presented by Bernardo & Smith (1994), cannot be applied to the present problem because it is based on the maximum likelihood estimates and, as we saw in section 6.3.1, we are unable of obtaining $\hat{\theta}_{n+1}$ which is a crucial estimate for the estimation of the predictive distributions we are seeking.

O'Hagan's (1994) suggestion can be applied because it uses only characteristics of the posterior distribution (8.12), namely its mode $m = \left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}\right)$ and its modal dispersion matrix.

The posterior mode $m = \left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}\right)$ is the solution of the system formed by the equations

$$\frac{((n+1)k+l)e^{\theta_i}}{\sum_{p=1}^{n+1} e^{\theta_p} + m} + \frac{\left(r(1-\delta_{i1}) + s\delta_{i1}\right)e^{\theta_i + \delta_{i1}\alpha}}{1 + e^{\theta_i + \delta_{i1}\alpha}} +$$

$$+ \frac{\left(r(1-\delta_{i2}) + s\delta_{i2}\right)e^{\theta_i + \beta + \delta_{i2}\alpha}}{1 + e^{\theta_i + \beta + \delta_{i2}\alpha}} - k - 2 = 0\,,\ i = 1, 2, \ldots, n,$$

$$\frac{((n+1)k+l)e^{\theta_{n+1}}}{\sum_{i=1}^{n+1} e^{\theta_i} + m} - k = 0, \qquad (8.16)$$

$$h + \sum_{i=1}^{n} \delta_{i1} w_{i1} + \sum_{i=1}^{n} \delta_{i2} w_{i2} - \frac{(q+h)e^{\alpha}}{e^{\alpha} + t} - \sum_{i=1}^{n} \left\{ \frac{s\delta_{i1} e^{\theta_i + \delta_{i1}\alpha}}{1 + e^{\theta_i + \delta_{i1}\alpha}} \right\} -$$

$$- \sum_{i=1}^{n} \left\{ \frac{s\delta_{i2} e^{\theta_i + \beta + \delta_{i2}\alpha}}{1 + e^{\theta_i + \beta + \delta_{i2}\alpha}} \right\} = 0,$$

$$g + \sum_{i=1}^{n} w_{i2} - \frac{(g+u)e^{\beta}}{e^{\beta} + v} - \sum_{i=1}^{n} \left\{ \frac{\left(r(1-\delta_{i2}) + s\delta_{i2}\right)e^{\theta_i + \beta + \delta_{i2}\alpha}}{1 + e^{\theta_i + \beta + \delta_{i2}\alpha}} \right\} = 0,$$

which must be obtained using a numerical technique. Then, defining the constants

$$\tilde{a}_i = \frac{((n+1)k+l)e^{\tilde{\theta}_i}\left(\sum_{\substack{p=1\\p\neq i}}^{n+1} e^{\tilde{\theta}_p}+m\right)}{\left(\sum_{p=1}^{n+1} e^{\tilde{\theta}_p}+m\right)^2} + \frac{\left(r\left(1-\delta_{i1}\right)+s\delta_{i1}\right)e^{\tilde{\theta}_i+\delta_{i1}\tilde{\alpha}}}{\left(1+e^{\tilde{\theta}_i+\delta_{i1}\tilde{\alpha}}\right)^2} +$$

$$+\frac{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)e^{\tilde{\theta}_i+\tilde{\beta}+\delta_{i2}\tilde{\alpha}}}{\left(1+e^{\tilde{\theta}_i+\tilde{\beta}+\delta_{i2}\tilde{\alpha}}\right)^2} \quad, i=1,2,\ldots,n,$$

$$\tilde{b}_{ij} = \frac{((n+1)k+l)e^{\tilde{\theta}_i+\tilde{\theta}_j}}{\left(\sum_{p=1}^{n+1} e^{\tilde{\theta}_p}+m\right)^2} \quad, i\neq j,\; i,j=1,2,\ldots,n+1, \tag{8.17}$$

$$\tilde{c}_{i1} = \frac{s\delta_{i1}e^{\tilde{\theta}_i+\delta_{i1}\tilde{\alpha}}}{\left(1+e^{\tilde{\theta}_i+\delta_{i1}\tilde{\alpha}}\right)^2}, \quad \tilde{c}_{i2} = \frac{s\delta_{i2}e^{\tilde{\theta}_i+\tilde{\beta}+\delta_{i2}\tilde{\alpha}}}{\left(1+e^{\tilde{\theta}_i+\tilde{\beta}+\delta_{i2}\tilde{\alpha}}\right)^2}, \quad \tilde{c}_i = \tilde{c}_{i1}+\tilde{c}_{i2}, \quad i=1,2,\ldots,n$$

$$\tilde{d}_i = \frac{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)e^{\tilde{\theta}_i+\tilde{\beta}+\delta_{i2}\tilde{\alpha}}}{\left(1+e^{\tilde{\theta}_i+\tilde{\beta}+\delta_{i2}\tilde{\alpha}}\right)^2} \quad, i=1,2,\ldots,n,$$

$$\tilde{f} = \frac{((n+1)k+l)e^{\tilde{\theta}_{n+1}}\left(\sum_{p=1}^{n} e^{\tilde{\theta}_p}+m\right)}{\left(\sum_{p=1}^{n+1} e^{\tilde{\theta}_p}+m\right)^2}, \qquad \tilde{o} = \sum_{i=1}^{n}\tilde{c}_{i2},$$

$$\tilde{p}_1 = \frac{(q+h)te^{\tilde{\alpha}}}{\left(e^{\tilde{\alpha}}+t\right)^2} + \sum_{i=1}^{n}\tilde{c}_{i1} + \sum_{i=1}^{n}\tilde{c}_{i2}, \qquad \tilde{p}_2 = \frac{(g+u)ve^{\tilde{\beta}}}{\left(e^{\tilde{\beta}}+v\right)^2} + \sum_{i=1}^{n}\tilde{d}_i,$$

the full conditional distributions are

$$p\left(\theta_i \mid \theta_{j\neq i}, \theta_{n+1}, \alpha, \beta, D^n\right) = N\left(\tilde{\theta}_i - \frac{\sum_{\substack{j=1\\j\neq i}}^{n+1}\left(\theta_j-\tilde{\theta}_j\right)\tilde{b}_{ij} + \left(\alpha-\tilde{\alpha}\right)\tilde{c}_i + \left(\beta-\tilde{\beta}\right)\tilde{d}_i}{\tilde{a}_i}, \frac{1}{\tilde{a}_i}\right),$$

$$i=1,2,\ldots,n,$$

$$p\left(\theta_{n+1} \mid \theta^n, \alpha, \beta, D^n\right) = N\left(\tilde{\theta}_{n+1} - \frac{\sum_{i=1}^{n}\left(\theta_i - \tilde{\theta}_i\right)\tilde{b}_{n+1,i}}{\tilde{f}}, \frac{1}{\tilde{f}}\right),$$

$$p\left(\alpha \mid \theta^n, \theta_{n+1}, \beta, D^n\right) = N\left(\tilde{\alpha} - \frac{\sum_{i=1}^{n}\left(\theta_i - \tilde{\theta}_i\right)\tilde{c}_i + \left(\beta - \tilde{\beta}\right)\tilde{\sigma}}{\tilde{p}_1}, \frac{1}{\tilde{p}_1}\right),$$

$$p\left(\beta \mid \theta^n, \theta_{n+1}, \alpha, D^n\right) = N\left(\tilde{\beta} - \frac{\sum_{i=1}^{n}\left(\theta_i - \tilde{\theta}_i\right)\tilde{d}_i + \left(\alpha - \tilde{\alpha}\right)\tilde{\sigma}}{\tilde{p}_2}, \frac{1}{\tilde{p}_2}\right).$$

The solution would not become simpler assuming vague second stage priors.

8.3.2.4. Laplace Approximation

We begin by defining a function $h\left(\theta^n, \theta_{n+1}, \alpha, \beta\right)$ as in (1.13), that is, such that

$$-n\,h\left(\theta^n, \theta_{n+1}, \alpha, \beta\right) = (k+2)\sum_{i=1}^{n}\theta_i + k\theta_{n+1} + \alpha\left(h + \sum_{i=1}^{n}\delta_{i1}w_{i1} + \sum_{i=1}^{n}\delta_{i2}w_{i2}\right) +$$

$$+\beta\left(g + \sum_{i=1}^{n}w_{i2}\right) - ((n+1)k+1)\ln\left(\sum_{i=1}^{n+1}e^{\theta_i} + m\right) -$$

$$-(q+h)\ln\left(e^{\alpha} + t\right) - (g+u)\ln\left(e^{\beta} + v\right) -$$

$$-\sum_{i=1}^{n}\left\{\left(r\left(1-\delta_{i1}\right) + s\delta_{i1}\right)\ln\left(1 + e^{\theta_i + \delta_{i1}\alpha}\right)\right\} -$$

$$-\sum_{i=1}^{n}\left\{\left(r\left(1-\delta_{i2}\right) + s\delta_{i2}\right)\ln\left(1 + e^{\theta_i + \beta + \delta_{i2}\alpha}\right)\right\}.$$

Based on this function we define $\left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}\right)$ by (1.16). Such a vector is the solution of the system of equations (8.16). Still using definition (1.16), we obtain

$$\tilde{\sigma} = \left\{\tilde{f}\,\tilde{p}_1 \left|\tilde{A}\right|\right\}^{-1/2},$$

where $\tilde{A}$ is a symmetric matrix of order $(n+1)$ whose elements are

$$\tilde{A}_{ii} = \tilde{a}_i - \frac{\tilde{b}^2_{i,n+1}}{\tilde{f}} - \frac{\tilde{c}^2_i}{\tilde{p}_1} \quad, i = 1, 2, \ldots, n,$$

$$\tilde{A}_{ij} = \tilde{b}_{ij} - \frac{\tilde{b}_{i,n+1}\tilde{b}_{j,n+1}}{\tilde{f}} - \frac{\tilde{c}_i\tilde{c}_j}{\tilde{p}_1} \quad, i \neq j, \quad i, j = 1, 2, \ldots, n,$$

$$\tilde{A}_{n+1,n+1} = \tilde{p}_2 - \frac{\tilde{\sigma}^2}{\tilde{p}_1}, \qquad \tilde{A}_{n+1,i} = \tilde{d}_i - \frac{\tilde{c}_i\tilde{\sigma}}{\tilde{p}_1} \quad, i = 1, 2, \ldots, n,$$

and all constants involved in the definition of these elements are given in (8.17). A numerical algorithm is required to evaluate the determinant of $\tilde{A}$.

8.3.2.4.1. Joint Predictive Distribution of (Z_1, Z_2)

We consider a function $h_c^*\left(\theta^n, \theta_{n+1}, \alpha, \beta\right)$ such that

$$-n\,h_c^*\left(\theta^n, \theta_{n+1}, \alpha, \beta\right) = \theta_{n+1}\left(z_1 + z_2\right) + \alpha z_2 - r\ln\left(1 + e^{\theta_{n+1}}\right) - s\ln\left(1 + e^{\alpha+\theta_{n+1}}\right) +$$

$$+(k+2)\sum_{i=1}^{n}\theta_i + k\theta_{n+1} + \alpha\left(h + \sum_{i=1}^{n}\delta_{i1}w_{i1} + \sum_{i=1}^{n}\delta_{i2}w_{i2}\right) +$$

$$+\beta\left(g + \sum_{i=1}^{n}w_{i2}\right) - \left((n+1)k + l\right)\ln\left(\sum_{i=1}^{n+1}e^{\theta_i} + m\right) -$$

$$-(q+h)\ln\left(e^{\alpha} + t\right) - (g+u)\ln\left(e^{\beta} + v\right) -$$

$$-\sum_{i=1}^{n}\left\{\left(r\left(1-\delta_{i1}\right)+s\delta_{i1}\right)\ln\left(1+e^{\theta_i+\delta_{i1}\alpha}\right)\right\}-$$

$$-\sum_{i=1}^{n}\left\{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)\ln\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)\right\}$$

and, solving numerically the system formed by the equations

$$\frac{((n+1)k+l)e^{\theta_i}}{\sum_{p=1}^{n+1}e^{\theta_p}+m}+\frac{\left(r\left(1-\delta_{i1}\right)+s\delta_{i1}\right)e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}+$$

$$+\frac{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}-k-2=0\,,\; i=1,2,\ldots,n,$$

$$\frac{((n+1)k+l)e^{\theta_{n+1}}}{\sum_{i=1}^{n+1}e^{\theta_i}+m}+\frac{re^{\theta_{n+1}}}{1+e^{\theta_{n+1}}}+\frac{se^{\alpha+\theta_{n+1}}}{1+e^{\alpha+\theta_{n+1}}}-z_1-z_2-k=0,$$

$$h+\sum_{i=1}^{n}\delta_{i1}w_{i1}+\sum_{i=1}^{n}\delta_{i2}w_{i2}+z_2-\frac{se^{\alpha+\theta_{n+1}}}{1+e^{\alpha+\theta_{n+1}}}-\frac{(q+h)e^{\alpha}}{e^{\alpha}+t}-$$

$$-\sum_{i=1}^{n}\left\{\frac{s\delta_{i1}e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}\right\}-\sum_{i=1}^{n}\left\{\frac{s\delta_{i2}e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\}=0,$$

$$g+\sum_{i=1}^{n}w_{i2}-\frac{(g+u)e^{\beta}}{e^{\beta}+v}-\sum_{i=1}^{n}\left\{\frac{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\}=0,$$

we derive $\left(\theta_{(c)}^{n*},\, \theta_{n+1(c)}^{*},\, \alpha_c^{*},\, \beta_c^{*}\right)$.Then, we define the constants

$$a_{i(c)}^{*}=\frac{((n+1)k+l)e^{\theta_{i(c)}^{*}}\left(\sum_{\substack{p=1\\p\neq i}}^{n+1}e^{\theta_{p(c)}^{*}}+m\right)}{\left(\sum_{p=1}^{n+1}e^{\theta_{p(c)}^{*}}+m\right)^2}+\frac{\left(r\left(1-\delta_{i1}\right)+s\delta_{i1}\right)e^{\theta_{i(c)}^{*}+\delta_{i1}\alpha_c^{*}}}{\left(1+e^{\theta_{i(c)}^{*}+\delta_{i1}\alpha_c^{*}}\right)^2}+$$

$$+\frac{\left(r(1-\delta_{i2})+s\delta_{i2}\right)e^{\theta^*_{i(c)}+\beta^*_c+\delta_{i2}\alpha^*_c}}{\left(1+e^{\theta^*_{i(c)}+\beta^*_c+\delta_{i2}\alpha^*_c}\right)^2} \quad, i=1,2,\ldots,n,$$

$$b^*_{ij(c)}=-\frac{((n+1)k+l)e^{\theta^*_{i(c)}+\theta^*_{j(c)}}}{\left(\sum_{p=1}^{n+1}e^{\theta^*_{p(c)}}+m\right)^2} \quad, i\neq j, \quad i,j=1,2,\ldots,n+1,$$

$$c^*_{i1(c)}=\frac{s\delta_{i1}e^{\theta^*_{i(c)}+\delta_{i1}\alpha^*_c}}{\left(1+e^{\theta^*_{i(c)}+\delta_{i1}\alpha^*_c}\right)^2}, \qquad c^*_{i2(c)}=\frac{s\delta_{i2}e^{\theta^*_{i(c)}+\beta^*_c+\delta_{i2}\alpha^*_c}}{\left(1+e^{\theta^*_{i(c)}+\beta^*_c+\delta_{i2}\alpha^*_c}\right)^2}, \qquad c^*_{i(c)}=c^*_{i1(c)}+c^*_{i2(c)},$$

$$i=1,2,\ldots,n,$$

$$d^*_{i(c)}=\frac{\left(r(1-\delta_{i2})+s\delta_{i2}\right)e^{\theta^*_{i(c)}+\beta^*_c+\delta_{i2}\alpha^*_c}}{\left(1+e^{\theta^*_{i(c)}+\beta^*_c+\delta_{i2}\alpha^*_c}\right)^2} \quad, i=1,2,\ldots,n,$$

$$f^*_c=\frac{((n+1)k+l)e^{\theta^*_{n+1(c)}}\left(\sum_{i=1}^{n}e^{\theta^*_{i(c)}}+m\right)}{\left(\sum_{i=1}^{n+1}e^{\theta^*_{i(c)}}+m\right)^2}+\frac{re^{\theta^*_{n+1(c)}}}{\left(1+e^{\theta^*_{n+1(c)}}\right)^2}+\frac{se^{\alpha^*_c+\theta^*_{n+1(c)}}}{\left(1+e^{\alpha^*_c+\theta^*_{n+1(c)}}\right)^2},$$

$$e^*_c=\frac{se^{\alpha^*_c+\theta^*_{n+1(c)}}}{\left(1+e^{\alpha^*_c+\theta^*_{n+1(c)}}\right)^2}, \qquad o^*_c=\sum_{i=1}^{n}c^*_{i2(c)},$$

$$p^*_{1(c)}=e^*_c+\frac{(q+h)t\,e^{\alpha^*_c}}{\left(e^{\alpha^*_c}+t\right)^2}+\sum_{i=1}^{n}c^*_{i1(c)}+\sum_{i=1}^{n}c^*_{i2(c)}, \qquad p^*_{2(c)}=\frac{(g+u)ve^{\beta^*_c}}{\left(e^{\beta^*_c}+v\right)^2}+\sum_{i=1}^{n}d^*_{i(c)},$$

and σ^*_c will be given by

$$\sigma^*_c=\left\{f^*_c\,p^*_{2(c)}\left|A^*_{(c)}\right|\right\}^{-1/2},$$

where $A^*_{(c)}$ is a symmetric matrix whose determinant must be evaluated numerically. The elements of such a matrix are

$$A^*_{(c)ii}=a^*_{i(c)}-\frac{b^{*2}_{i,n+1(c)}}{f^*_c}-\frac{d^{*2}_{i(c)}}{p^*_{2(c)}} \quad, i=1,2,\ldots,n,$$

$$A^*_{(c)ij} = b^*_{ij(c)} - \frac{b^*_{i,n+1(c)}b^*_{j,n+1(c)}}{f^*_c} - \frac{d^*_{i(c)}d^*_{j(c)}}{p^*_{2(c)}} \quad , i \neq j, \quad i, j = 1, 2, \ldots, n,$$

$$A^*_{(c)n+1,n+1} = p^*_{1(c)} - \frac{e^{*2}_c}{f^*_c} - \frac{o^{*2}_c}{p^*_{2(c)}},$$

$$A^*_{(c)n+1,i} = c^*_{i(c)} - \frac{b^*_{i,n+1(c)}e^*_c}{f^*_c} - \frac{d^*_{i(c)}o^*_c}{p^*_{2(c)}} \quad , i = 1, 2, \ldots, n.$$

Hence, the joint predictive distribution (8.13) is approximately given by

$$p\left(z_1, z_2 \mid D^n\right) \,\tilde{\propto}\, \binom{r}{z_1}\binom{s}{z_2}\left(\frac{\sigma^*_c}{\tilde{\sigma}}\right)\exp\left\{-nh^*_c\left(\theta^{n*}_{(c)}, \theta^*_{n+1(c)}, \alpha^*_c, \beta^*_c\right) + nh\left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}\right)\right\}.$$

8.3.2.4.2. Marginal Predictive Distribution of Z_1

The predictive distribution of Z_1 is approximated by

$$p\left(z_1 \mid D^n\right) \,\tilde{\propto}\, \binom{r}{z_1}\left(\frac{\sigma^*_1}{\tilde{\sigma}}\right)\exp\left\{-nh^*_1\left(\theta^{n*}_{(1)}, \theta^*_{n+1(1)}, \alpha^*_1, \beta^*_1\right) + nh\left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}\right)\right\},$$

where the function $h^*_1\left(\theta^n, \theta_{n+1}, \alpha, \beta\right)$ is defined such that

$$-nh^*_1\left(\theta^n, \theta_{n+1}, \alpha, \beta\right) = \theta_{n+1}z_1 - r\ln\left(1 + e^{\theta_{n+1}}\right) + (k+2)\sum_{i=1}^{n}\theta_i + k\theta_{n+1} +$$

$$+\alpha\left(h + \sum_{i=1}^{n}\delta_{i1}w_{i1} + \sum_{i=1}^{n}\delta_{i2}w_{i2}\right) + \beta\left(g + \sum_{i=1}^{n}w_{i2}\right) -$$

$$-((n+1)k + l)\ln\left(\sum_{i=1}^{n+1}e^{\theta_i} + m\right) - (q+h)\ln\left(e^{\alpha} + t\right) -$$

$$-(g+u)\ln\left(e^{\beta} + v\right) - \sum_{i=1}^{n}\left\{\left(r\left(1-\delta_{i1}\right) + s\delta_{i1}\right)\ln\left(1 + e^{\theta_i + \delta_{i1}\alpha}\right)\right\} -$$

$$-\sum_{i=1}^{n}\left\{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)\ln\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)\right\},$$

$\left(\theta_{(1)}^{n*}, \theta_{n+1(1)}^{*}, \alpha_1^{*}, \beta_1^{*}\right)$ is the numerical solution of the system formed by the equations

$$\frac{((n+1)k+l)e^{\theta_i}}{\sum_{p=1}^{n+1} e^{\theta_p}+m}+\frac{\left(r\left(1-\delta_{i1}\right)+s\delta_{i1}\right)e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}+$$

$$+\frac{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}-k-2=0,\ i=1,2,\ldots,n,$$

$$\frac{((n+1)k+l)e^{\theta_{n+1}}}{\sum_{i=1}^{n+1} e^{\theta_i}+m}+\frac{re^{\theta_{n+1}}}{1+e^{\theta_{n+1}}}-z_1-k=0,$$

$$h+\sum_{i=1}^{n}\delta_{i1}w_{i1}+\sum_{i=1}^{n}\delta_{i2}w_{i2}-\frac{(q+h)e^{\alpha}}{e^{\alpha}+t}-\sum_{i=1}^{n}\left\{\frac{s\delta_{i1}e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}\right\}-$$

$$-\sum_{i=1}^{n}\left\{\frac{s\delta_{i2}e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\}=0,$$

$$g+\sum_{i=1}^{n}w_{i2}-\frac{(g+u)e^{\beta}}{e^{\beta}+v}-\sum_{i=1}^{n}\left\{\frac{\left(r\left(1-\delta_{i2}\right)+s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\}=0,$$

and

$$\sigma_1^{*}=\left\{f_1^{*}\,p_{2(1)}^{*}\left|A_{(1)}^{*}\right|\right\}^{-1/2};$$

$A_{(1)}^{*}$ is a symmetric matrix whose elements are given by

$$A_{(1)ii}^{*}=a_{i(1)}^{*}-\frac{b_{i,n+1(1)}^{*2}}{f_1^{*}}-\frac{d_{i(1)}^{*2}}{p_{2(1)}^{*}},\ i=1,2,\ldots,n,$$

$$A_{(1)ij}^{*}=b_{ij(1)}^{*}-\frac{b_{i,n+1(1)}^{*}b_{j,n+1(1)}^{*}}{f_1^{*}}-\frac{d_{i(1)}^{*}d_{j(1)}^{*}}{p_{2(1)}^{*}},\ i\neq j,\quad i,j=1,2,\ldots,n,$$

$$A^*_{(1)n+1,n+1} = p^*_{1(1)} - \frac{o_1^{*2}}{p^*_{2(1)}}, \qquad A^*_{(1)n+1,i} = c^*_{i(1)} - \frac{d^*_{i(1)}o^*_1}{p^*_{2(1)}} \quad , i = 1, 2, \ldots, n,$$

with

$$a^*_{i(1)} = \frac{((n+1)k+l)e^{\theta^*_{i(1)}}\left(\sum_{\substack{p=1\\p\neq i}}^{n+1} e^{\theta^*_{p(1)}} + m\right)}{\left(\sum_{p=1}^{n+1} e^{\theta^*_{p(1)}} + m\right)^2} + \frac{\left(r(1-\delta_{i1}) + s\delta_{i1}\right)e^{\theta^*_{i(1)}+\delta_{i1}\alpha^*_i}}{\left(1 + e^{\theta^*_{i(1)}+\delta_{i1}\alpha^*_i}\right)^2} +$$

$$+ \frac{\left(r(1-\delta_{i2}) + s\delta_{i2}\right)e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}}{\left(1 + e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}\right)^2} \quad , i = 1, 2, \ldots, n,$$

$$b^*_{ij(1)} = -\frac{((n+1)k+l)e^{\theta^*_{i(1)}+\theta^*_{j(1)}}}{\left(\sum_{p=1}^{n+1} e^{\theta^*_{p(1)}} + m\right)^2} \quad , i \neq j, \quad i, j = 1, 2, \ldots, n+1,$$

$$c^*_{i1(1)} = \frac{s\delta_{i1}e^{\theta^*_{i(1)}+\delta_{i1}\alpha^*_i}}{\left(1 + e^{\theta^*_{i(1)}+\delta_{i1}\alpha^*_i}\right)^2}, \quad c^*_{i2(1)} = \frac{s\delta_{i2}e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}}{\left(1 + e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}\right)^2}, \quad c^*_{i(1)} = c^*_{i1(1)} + c^*_{i2(1)}$$

$$i = 1, 2, \ldots, n,$$

$$d^*_{i(1)} = \frac{\left(r(1-\delta_{i2}) + s\delta_{i2}\right)e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}}{\left(1 + e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}\right)^2} \quad , i = 1, 2, \ldots, n,$$

$$f^*_1 = \frac{((n+1)k+l)e^{\theta^*_{n+1(1)}}\left(\sum_{i=1}^{n} e^{\theta^*_{i(1)}} + m\right)}{\left(\sum_{i=1}^{n+1} e^{\theta^*_{i(1)}} + m\right)^2} + \frac{re^{\theta^*_{n+1(1)}}}{\left(1 + e^{\theta^*_{n+1(1)}}\right)^2}, \qquad o^*_1 = \sum_{i=1}^{n} c^*_{i2(1)},$$

$$p^*_{1(1)} = \frac{(q+h)te^{\alpha^*_i}}{\left(e^{\alpha^*_i} + t\right)^2} + \sum_{i=1}^{n} c^*_{i1(1)} + \sum_{i=1}^{n} c^*_{i2(1)},$$

$$p^*_{2(1)} = \frac{(g+u)ve^{\beta^*_i}}{\left(e^{\beta^*_i} + v\right)^2} + \sum_{i=1}^{n}\left\{\frac{\left(r(1-\delta_{i2}) + s\delta_{i2}\right)e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}}{\left(1 + e^{\theta^*_{i(1)}+\beta^*_i+\delta_{i2}\alpha^*_i}\right)^2}\right\}.$$

The evaluation of the determinant of $A^{*}_{(1)}$ requires a numerical algorithm.

8.3.2.4.3. Marginal Predictive Distribution of Z_2

In order to derive an approximation for the predictive distribution (8.15), we define a function $h_2^{*}\left(\theta^{n}, \theta_{n+1}, \alpha, \beta\right)$ such that

$$
\begin{aligned}
-n h_2^{*}\left(\theta^{n}, \theta_{n+1}, \alpha, \beta\right) = {} & (\alpha+\theta_{n+1})z_2 - s\ln\left(1+e^{\alpha+\theta_{n+1}}\right) + (k+2)\sum_{i=1}^{n}\theta_i + k\theta_{n+1} + \\
& +\alpha\left(h+\sum_{i=1}^{n}\delta_{i1}w_{i1}+\sum_{i=1}^{n}\delta_{i2}w_{i2}\right)+\beta\left(g+\sum_{i=1}^{n}w_{i2}\right)- \\
& -((n+1)k+l)\ln\left(\sum_{i=1}^{n+1}e^{\theta_i}+m\right)-(q+h)\ln\left(e^{\alpha}+t\right)- \\
& -(g+u)\ln\left(e^{\beta}+v\right)-\sum_{i=1}^{n}\left\{\left(r(1-\delta_{i1})+s\delta_{i1}\right)\ln\left(1+e^{\theta_i+\delta_{i1}\alpha}\right)\right\}- \\
& -\sum_{i=1}^{n}\left\{\left(r(1-\delta_{i2})+s\delta_{i2}\right)\ln\left(1+e^{\theta_i+\beta+\delta_{i2}\alpha}\right)\right\}.
\end{aligned}
$$

Then, we obtain $\left(\theta_{(2)}^{n*}, \theta^{*}_{n+1(2)}, \alpha_2^{*}, \beta_2^{*}\right)$ solving the system formed by the equations

$$
\begin{aligned}
& \frac{((n+1)k+l)e^{\theta_i}}{\sum_{p=1}^{n+1}e^{\theta_p}+m}+\frac{\left(r(1-\delta_{i1})+s\delta_{i1}\right)e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}+ \\
& \qquad +\frac{\left(r(1-\delta_{i2})+s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}-k-2=0,\ i=1,2,\ldots,n,
\end{aligned}
$$

$$
\frac{((n+1)k+l)e^{\theta_{n+1}}}{\sum_{i=1}^{n+1}e^{\theta_i}+m}+\frac{s e^{\alpha+\theta_{n+1}}}{1+e^{\alpha+\theta_{n+1}}}-z_2-k=0,
$$

$$h+\sum_{i=1}^{n}\delta_{i1}w_{i1}+\sum_{i=1}^{n}\delta_{i2}w_{i2}+z_2-\frac{s e^{\alpha+\theta_{n+1}}}{1+e^{\alpha+\theta_{n+1}}}-\frac{(q+h)e^{\alpha}}{e^{\alpha}+t}-$$

$$-\sum_{i=1}^{n}\left\{\frac{s\delta_{i1}e^{\theta_i+\delta_{i1}\alpha}}{1+e^{\theta_i+\delta_{i1}\alpha}}\right\}-\sum_{i=1}^{n}\left\{\frac{s\delta_{i2}e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\}=0,$$

$$g+\sum_{i=1}^{n}w_{i2}-\frac{(g+u)e^{\beta}}{e^{\beta}+v}-\sum_{i=1}^{n}\left\{\frac{\left(r(1-\delta_{i2})+s\delta_{i2}\right)e^{\theta_i+\beta+\delta_{i2}\alpha}}{1+e^{\theta_i+\beta+\delta_{i2}\alpha}}\right\}=0\,.$$

Using the solution of this system, we define the constants

$$a^*_{i(2)}=\frac{((n+1)k+l)e^{\theta^*_{i(2)}}\left(\sum_{\substack{p=1\\p\neq i}}^{n+1}e^{\theta^*_{p(2)}}+m\right)}{\left(\sum_{p=1}^{n+1}e^{\theta^*_{p(2)}}+m\right)^2}+\frac{\left(r(1-\delta_{i1})+s\delta_{i1}\right)e^{\theta^*_{i(2)}+\delta_{i1}\alpha^*_2}}{\left(1+e^{\theta^*_{i(2)}+\delta_{i1}\alpha^*_2}\right)^2}+$$

$$+\frac{\left(r(1-\delta_{i2})+s\delta_{i2}\right)e^{\theta^*_{i(2)}+\beta^*_2+\delta_{i2}\alpha^*_2}}{\left(1+e^{\theta^*_{i(2)}+\beta^*_2+\delta_{i2}\alpha^*_2}\right)^2}\quad,\ i=1,2,\ldots,n,$$

$$b^*_{ij(2)}=-\frac{((n+1)k+l)e^{\theta^*_{i(2)}+\theta^*_{j(2)}}}{\left(\sum_{p=1}^{n+1}e^{\theta^*_{p(2)}}+m\right)^2}\quad,\ i\neq j,\quad i,j=1,2,\ldots,n+1,$$

$$c^*_{i1(2)}=\frac{s\delta_{i1}e^{\theta^*_{i(2)}+\delta_{i1}\alpha^*_2}}{\left(1+e^{\theta^*_{i(2)}+\delta_{i1}\alpha^*_2}\right)^2},\qquad c^*_{i2(2)}=\frac{s\delta_{i2}e^{\theta^*_{i(2)}+\beta^*_2+\delta_{i2}\alpha^*_2}}{\left(1+e^{\theta^*_{i(2)}+\beta^*_2+\delta_{i2}\alpha^*_2}\right)^2},\qquad c^*_{i(2)}=c^*_{i1(2)}+c^*_{i2(2)},$$

$$i=1,2,\ldots,n,$$

$$d^*_{i(2)}=\frac{\left(r(1-\delta_{i2})+s\delta_{i2}\right)e^{\theta^*_{i(2)}+\beta^*_2+\delta_{i2}\alpha^*_2}}{\left(1+e^{\theta^*_{i(2)}+\beta^*_2+\delta_{i2}\alpha^*_2}\right)^2}\quad,\ i=1,2,\ldots,n,$$

$$f^*_2=\frac{((n+1)k+l)e^{\theta^*_{n+1(2)}}\left(\sum_{i=1}^{n}e^{\theta^*_{i(2)}}+m\right)}{\left(\sum_{i=1}^{n+1}e^{\theta^*_{i(2)}}+m\right)^2}+\frac{s e^{\alpha^*_2+\theta^*_{n+1(2)}}}{\left(1+e^{\alpha^*_2+\theta^*_{n+1(2)}}\right)^2},$$

$$e_2^* = \frac{s e^{\alpha_2^* + \theta_{n+1(2)}^*}}{\left(1 + e^{\alpha_2^* + \theta_{n+1(2)}^*}\right)^2}, \qquad o_2^* = \sum_{i=1}^{n} c_{i2(2)}^*,$$

$$p_{1(2)}^* = e_2^* + \frac{(q+h)te^{\alpha_2^*}}{\left(e^{\alpha_2^*} + t\right)^2} + \sum_{i=1}^{n} c_{i1(2)}^* + \sum_{i=1}^{n} c_{i2(2)}^*, \qquad p_{2(2)}^* = \frac{(g+u)ve^{\beta_2^*}}{\left(e^{\beta_2^*} + v\right)^2} + \sum_{i=1}^{n} d_{i(2)}^*,$$

and then,

$$\sigma_2^* = \left\{f_2^* \, p_{2(2)}^* \left|A_{(2)}^*\right|\right\}^{-1/2}.$$

The elements of $A_{(2)}^*$ are

$$A_{(2)ii}^* = a_{i(2)}^* - \frac{b_{i,n+1(2)}^{*2}}{f_2^*} - \frac{d_{i(2)}^{*2}}{p_{2(2)}^*}, \quad i = 1, 2, \ldots, n,$$

$$A_{(2)ij}^* = b_{ij(2)}^* - \frac{b_{i,n+1(2)}^* b_{j,n+1(2)}^*}{f_2^*} - \frac{d_{i(2)}^* d_{j(2)}^*}{p_{2(2)}^*}, \quad i \neq j, \quad i, j = 1, 2, \ldots, n,$$

$$A_{(2)n+1,n+1}^* = p_{1(2)}^* - \frac{e_2^{*2}}{f_2^*} - \frac{o_2^{*2}}{p_{2(2)}^*},$$

$$A_{(2)n+1,i}^* = c_{i(2)}^* - \frac{b_{i,n+1(2)}^* e_2^*}{f_2^*} - \frac{d_{i(2)}^* o_2^*}{p_{2(2)}^*}, \quad i = 1, 2, \ldots, n,$$

and its determinant must be obtained numerically. Finally, the Laplace approximation for (8.15) will be

$$p(z_2 \mid D^n) \propto \binom{s}{z_2} \left(\frac{\sigma_2^*}{\tilde{\sigma}}\right) \exp\left\{-nh_2^*\left(\theta_{(2)}^{n*}, \theta_{n+1(2)}^*, \alpha_2^*, \beta_2^*\right) + nh\left(\tilde{\theta}^n, \tilde{\theta}_{n+1}, \tilde{\alpha}, \tilde{\beta}\right)\right\}.$$

In the above approximations we need to evaluate numerically determinants of large matrices which can cause some problems. When we consider the special vague second stage priors case, the solution for the problem does not simplify significantly.

CHAPTER 9

CONCLUSIONS AND POSSIBLE FUTURE WORK

We have developed, in a Bayesian framework, methods to derive the predictive distributions for three problems, which can be summarised as follows:

The "Multiplicative Effect of a Treatment" problem:

- given θ_i, $X_i \sim Po(\exp(\theta_i))$,
- given θ_i and α, $Y_i \sim Po(\exp(\alpha + \theta_i))$;

The "Treatment Effect Under Biased Allocation" problem

- given θ_i, $X_i \sim Po(\exp(\theta_i))$,
- given α, β, θ_i and x_i,

  $Y_i \sim Po(\exp(\alpha + \theta_i))$, if treatment T_1 is used,

  $Y_i \sim Po(\exp(\beta + \theta_i))$, if treatment T_2 is used,
- given a, T_1 is used if $x_i < a$ and T_2 is used if $x_i \geq a$;

The "Crossover Design to Compare Two Treatments" problem:

Treatment	Period 1	Period 2
T_1	$Po(\exp(\theta_i))$	$Po(\exp(\beta + \theta_i))$
T_2	$Po(\exp(\alpha + \theta_i))$	$Po(\exp(\alpha + \beta + \theta_i))$

We also generalised the second problem to a situation where more than two treatments are available.

Since the evaluation of the exact predictive distributions requires the use of numerical integration techniques which are sometimes not feasible, because of the form of the function to integrate or because of the dimension of the integral, we derived estimates and approximations for the predictive distributions. These were through the Gibbs sampling approach, normal approximations for the posterior distributions and the Laplace approximation. Analyse of the results obtained in the examples presented allowed us to conclude that the Gibbs sampling approach and the Laplace approximation were excellent approximations. Furthermore, because of the high dimensionality of the parameters the normal approximations for the posterior distributions were somewhat poor, especially the one based on the likelihood function (Bernardo & Smith(1994)), which tends to diverge towards the plug-in estimate. The normal approximation suggested by O'Hagan (1994) improves matter somewhat. One point in favour of the Laplace approximation when compared with the Gibbs sampling approach is the superior computational speed.

We also considered the three problems above with the assumption that the underlying distributions were exponential and binomial. The solutions obtained seem to be more difficult to implement due to the numerical mathematical techniques then required, and is expected to be undertaken in the future. It would also be interesting, in the future, to see if it is possible to generalise the solution of the three problems for the case when the underlying distribution is any element of the exponential family of distributions.

We also studied a model that could be used to predict the number of accidents, y_i, at a road junction based on the traffic flows, x_{ij}, at the junction and on a set of covariates, z_{it}, describing features of the junction. Such a model is

$$X_{ij} \sim Po\left(\exp\left(a_{ij} + k_{ij}\right)\right) \quad , \quad i = 1, 2, \ldots, n \quad ; \quad j = 1, 2, \ldots, f$$

$$Y_i \sim Po\left(\exp\left(\sum_{j=1}^{f} \lambda_j a_{ij} + \sum_{t=1}^{c} \beta_t z_{it}\right)\right) \quad , \quad i = 1, 2, \ldots, n.$$

Barnett & Wright (1990) and Dunsmore & Robson (1992) had considered this problem in a classical framework; here we developed a Bayesian predictive approach. The exact predictive distribution involves high dimension integrals, which make numerical integration difficult. Therefore we developed estimates and approximations for the predictive distribution, as in the problems considered earlier.

In an analysis of a data set, in which we concentrated on the Laplace methods, we found that the model above was not too sensitive to small or moderate changes in the parameters of the prior structure. A method for choice of hyperparameters was suggested. We found that the systems of equations involved in the solution of this problem were extremely unstable and that good starting values must be given in order to achieve convergence of the numerical technique. Methods for the choice of covariates through the use of Kullback-Leibler divergence measure were also considered.

APPENDIX 1

DISTRIBUTIONS OF FUNCTIONS OF RANDOM VARIABLES

Throughout the developments of the methods we have considered various functions of random variables. We record here some distributional results.

Table A1.1. X is $Ga(g,h)$.

Random Variable	Probability Density Function	Range
X	$p(x) = \frac{h^g x^{g-1} e^{-hx}}{\Gamma(g)}$	$x>0$
$Y=\ln(X)$	$p(y) = \frac{h^g \exp\{-he^y\}\exp\{gy\}}{\Gamma(g)}$	$y \in \Re$
$T=\ln(X)-k$	$p(t) = \frac{h^g \exp\{-he^{t+k}\}\exp\{(t+k)g\}}{\Gamma(g)}$	$t \in \Re$
$Z = \frac{\ln(X)}{k}$ $(k>0)$	$p(z) = \frac{kh^g \exp\{-he^{kz}\}\exp\{gkz\}}{\Gamma(g)}$	$z \in \Re$

Table A1.2. X is $Be(g,h)$.

Random Variable	Probability Density Function	Range
X	$p(x)=\dfrac{x^{g-1}(1-x)^{h-1}}{B(g,h)}$	$0<x<1$
$Y=\dfrac{1-X}{X}$	$p(y)=\dfrac{1}{B(h,g)}\dfrac{y^{h-1}}{(y+1)^{g+h}}$	$y>0$
$Z=vY$ $(v>0)$	$p(z)=\dfrac{v^{g}}{B(h,g)}\dfrac{z^{h-1}}{(z+v)^{g+h}}$	$z>0$
$T=\ln(Z)$	$p(t)=\dfrac{v^{g}}{B(h,g)}\dfrac{\exp\{th\}}{(\exp\{t\}+v)^{g+h}}$	$t\in\Re$

APPENDIX 2

TRAFFIC ACCIDENTS DATA SET

This appendix contains the data set used in the example presented in section 6.2.5. Values were recorded at *78* road junctions, as shown in table A2.1. For the *i*-th junction it was considered two traffic flows, x_{i1} and x_{i2}, the number of accidents, y_i, and four covariates, z_{i1}, z_{i2}, z_{i3} and z_{i4}, measuring

z_{i1}- the entry path curvature (*1*/ metres)

z_{i2}- the entry width (metres)

z_{i3}- the percentage of motorcycles (%)

z_{i4}- the approach gradient (in categories between *-3* and *+3*: *-g*=downhill).

Table A2.1:

i	x_{i1}	x_{i2}	y_i	z_{i1}	z_{i2}	z_{i3}	z_{i4}
1	69	104	5	0.0040000	11.4	2.96	0
2	76	93	6	0.0028571	6.1	2.30	-1
3	77	84	0	0.0071429	7.6	2.46	-2
4	113	101	4	0.0050000	17.6	1.26	0
5	92	95	1	0.0028571	12.1	1.18	-1
6	128	109	12	0.0000000	17.0	1.32	0
7	59	76	7	0.0076923	12.0	2.18	-2
8	62	81	2	0.0000000	11.8	2.86	0
9	200	96	10	0.0111111	15.3	2.20	0
10	92	77	1	0.0000000	8.1	1.25	-1

Table A2.1 (cont.):

i	x_{i1}	x_{i2}	y_i	z_{i1}	z_{i2}	z_{i3}	z_{i4}
11	74	60	2	0.0028571	7.7	1.65	1
12	55	59	5	0.0000000	11.7	2.34	−2
13	45	47	0	0.0090909	6.0	1.96	−1
14	165	99	2	0.0055556	13.5	1.56	0
15	75	96	1	0.0100000	9.2	2.52	−1
16	81	33	1	0.0000000	6.2	2.75	−2
17	94	90	5	−0.0055556	12.4	3.66	2
18	93	58	4	0.0000000	9.1	1.62	−1
19	100	28	2	0.0000000	6.9	2.59	0
20	67	103	0	0.0028571	8.5	2.24	0
21	88	51	0	0.0028571	7.2	1.84	1
22	89	85	2	−0.0050000	12.4	3.81	2
23	70	85	2	0.0000000	6.1	1.91	−1
24	75	114	10	0.0040000	11.4	3.04	0
25	159	220	3	0.0105263	8.2	2.62	0
26	195	70	7	0.0066667	14.0	1.66	3
27	55	82	2	0.0031250	11.5	1.95	0
28	82	91	4	0.0050000	10.6	1.46	1
29	48	88	5	0.0083333	12.0	1.53	0
30	146	68	1	0.0028571	13.5	2.13	0
31	42	63	0	0.0028571	9.2	2.40	1
32	195	50	9	0.0000000	10.2	2.00	−1
33	23	2	0	0.0040000	6.5	0.99	−1
34	85	33	3	0.0142857	6.5	2.89	0
35	20	17	1	0.0055556	6.6	2.39	0
36	102	100	0	0.0181818	5.8	2.82	0
37	100	113	0	0.0166667	6.3	2.97	−1
38	77	125	0	0.0166667	6.6	2.28	0
39	100	80	0	0.0285714	6.8	1.64	0
40	52	43	1	0.0037037	6.5	3.57	0

Table A2.1 (cont.);

i	x_{i1}	x_{i2}	y_i	z_{i1}	z_{i2}	z_{i3}	z_{i4}
41	42	42	0	0.0188679	8.4	2.32	0
42	97	58	0	0.0204082	7.4	1.88	0
43	68	78	0	0.0250000	5.5	3.12	0
44	107	49	0	0.0133333	7.1	2.36	−1
45	28	79	0	0.0117647	7.3	4.82	0
46	73	162	3	0.0166667	7.0	1.96	1
47	52	69	5	0.0047619	12.5	2.90	1
48	88	39	0	0.0181818	7.3	1.81	0
49	88	61	0	0.0285714	7.4	1.08	1
50	15	44	0	0.0125000	6.9	0.72	1
51	50	62	0	0.0250000	6.7	2.43	0
52	207	127	3	0.0250000	7.1	2.92	0
53	127	65	1	0.0285714	7.7	4.12	0
54	72	57	0	0.0095238	7.4	1.01	−1
55	21	47	0	0.0153846	7.8	1.06	1
56	87	75	0	0.0111111	7.1	1.99	0
57	149	205	2	0.0285714	7.0	2.51	0
58	139	180	0	0.0476190	7.1	4.08	1
59	106	91	0	0.0192308	7.7	0.90	−3
60	170	106	0	0.0102041	11.4	1.48	2
61	125	54	0	0.0166667	7.5	5.73	1
62	103	58	2	0.0119048	7.0	2.34	1
63	115	82	0	0.0153846	8.7	1.64	1
64	55	53	0	0.0090909	7.5	1.36	−3
65	130	43	1	0.0050000	7.5	2.73	0
66	104	87	2	0.0100000	7.9	2.60	0
67	164	79	0	0.0083333	7.3	2.53	0
68	177	85	0	0.0043478	9.6	2.57	0
69	67	37	0	0.0090909	7.5	1.18	0
70	211	63	1	0.0285714	8.0	1.99	−1

Table A2.1 (cont.):

i	x_{i1}	x_{i2}	y_i	z_{i1}	z_{i2}	z_{i3}	z_{i4}
71	267	108	4	0.0142857	11.6	2.28	2
72	97	81	0	0.0200000	7.7	2.09	-2
73	104	37	0	0.0125000	8.7	1.15	-3
74	70	72	0	0.0166667	8.3	0.84	1
75	95	36	0	0.0105263	7.3	1.79	-1
76	84	78	0	0.0158730	7.9	0.79	0
77	81	49	0	0.0285714	7.5	1.84	-2
78	73	55	0	0.0153846	9.0	2.89	2

APPENDIX 3

SOME NUMERICAL RESULTS

This appendix contains the numerical results obtained when the examples presented in Chapters 2 and 3 were solved.

Table A3.1: Predictive probabilities for example 1 in section 2.2.5.

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
0	0.110520	0.143711	0.155460	0.103723	0.110489	0.107442
1	0.229543	0.278792	0.268081	0.221941	0.229507	0.225594
2	0.250998	0.270422	0.249087	0.248978	0.250990	0.249855
3	0.192367	0.174869	0.166453	0.195354	0.192384	0.193986
4	0.116084	0.084809	0.090090	0.120665	0.116107	0.118523
5	0.058753	0.032905	0.042154	0.062612	0.058770	0.060701
6	0.025946	0.010639	0.017768	0.028439	0.025956	0.027115
7	0.010271	0.002948	0.006942	0.011634	0.010276	0.010856
8	0.003716	0.000715	0.002568	0.004376	0.003718	0.003973
9	0.001247	0.000154	0.000915	0.001537	0.001248	0.001348
10	0.000393	0.000030	0.000318	0.000510	0.000393	0.000429
11	0.000117	0.000005	0.000109	0.000162	0.000117	0.000129
12	0.000033	0.000001	0.000037	0.000049	0.000033	0.000037
13	0.000009	0.000000	0.000012	0.000014	0.000009	0.000010
14	0.000002	0.000000	0.000004	0.000004	0.000002	0.000003
15	0.000001	0.000000	0.000001	0.000001	0.000001	0.000001
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table A3.2: Predictive probabilities for example 2 in section 2.2.5.

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
0	0.000004	0.000001	0.000019	0.000003	0.000004	0.000003
1	0.000045	0.000013	0.000160	0.000031	0.000045	0.000034
2	0.000236	0.000092	0.000713	0.000174	0.000236	0.000186
3	0.000852	0.000425	0.002224	0.000660	0.000851	0.000701
4	0.002373	0.001474	0.005448	0.001922	0.002371	0.002026
5	0.005443	0.004084	0.011116	0.004579	0.005441	0.004791
6	0.010704	0.009433	0.019614	0.009296	0.010700	0.009668
7	0.018548	0.018673	0.030707	0.016545	0.018541	0.017123
8	0.028885	0.032344	0.043475	0.026356	0.028877	0.027178
9	0.041048	0.049799	0.056484	0.038173	0.041038	0.039267
10	0.053858	0.069007	0.068133	0.050894	0.053847	0.052278
11	0.065865	0.086931	0.077026	0.063086	0.065854	0.064757
12	0.075659	0.100385	0.082257	0.073304	0.075651	0.075218
13	0.082162	0.107003	0.083529	0.080398	0.082156	0.082452
14	0.084806	0.105911	0.081115	0.083723	0.084804	0.085756
15	0.083587	0.097842	0.075705	0.083201	0.083588	0.085015
16	0.078983	0.084738	0.068203	0.079255	0.078987	0.080651
17	0.071799	0.069072	0.059541	0.072650	0.071805	0.073472
18	0.062980	0.053175	0.050541	0.064309	0.062987	0.064477
19	0.053450	0.038782	0.041839	0.055137	0.053458	0.054662
20	0.043994	0.026870	0.033866	0.045914	0.044002	0.044884
21	0.035192	0.017731	0.026862	0.037223	0.035200	0.035782
22	0.027412	0.011168	0.020919	0.029442	0.027419	0.027757
23	0.020827	0.006729	0.016022	0.022763	0.020832	0.020993
24	0.015458	0.003885	0.012087	0.017233	0.015462	0.015507
25	0.011224	0.002153	0.008993	0.012794	0.011227	0.011206
26	0.007983	0.001148	0.006610	0.009329	0.007986	0.007933
27	0.005568	0.000589	0.004806	0.006690	0.005570	0.005507
28	0.003813	0.000292	0.003463	0.004726	0.003814	0.003752
29	0.002566	0.000139	0.002477	0.003294	0.002567	0.002511

Table A3.2 (cont.):

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
30	*0.001698*	*0.000064*	*0.001763*	*0.002268*	*0.001699*	*0.001651*
31	*0.001106*	*0.000029*	*0.001250*	*0.001545*	*0.001107*	*0.001067*
32	*0.000710*	*0.000012*	*0.000885*	*0.001043*	*0.000711*	*0.000678*
33	*0.000450*	*0.000005*	*0.000627*	*0.000699*	*0.000450*	*0.000423*
34	*0.000281*	*0.000002*	*0.000445*	*0.000465*	*0.000281*	*0.000260*
35	*0.000173*	*0.000001*	*0.000316*	*0.000307*	*0.000173*	*0.000157*
36	*0.000105*	*0.000000*	*0.000225*	*0.000202*	*0.000105*	*0.000093*
37	*0.000063*	*0.000000*	*0.000160*	*0.000132*	*0.000063*	*0.000054*
38	*0.000038*	*0.000000*	*0.000114*	*0.000086*	*0.000038*	*0.000031*
39	*0.000022*	*0.000000*	*0.000081*	*0.000055*	*0.000022*	*0.000018*
40	*0.000013*	*0.000000*	*0.000057*	*0.000035*	*0.000013*	*0.000010*
41	*0.000007*	*0.000000*	*0.000040*	*0.000022*	*0.000007*	*0.000005*
42	*0.000004*	*0.000000*	*0.000028*	*0.000014*	*0.000004*	*0.000003*
43	*0.000002*	*0.000000*	*0.000019*	*0.000009*	*0.000002*	*0.000001*
44	*0.000001*	*0.000000*	*0.000013*	*0.000005*	*0.000001*	*0.000001*
45	*0.000001*	*0.000000*	*0.000009*	*0.000003*	*0.000001*	*0.000000*
46	*0.000000*	*0.000000*	*0.000006*	*0.000002*	*0.000000*	*0.000000*
47	*0.000000*	*0.000000*	*0.000004*	*0.000001*	*0.000000*	*0.000000*
48	*0.000000*	*0.000000*	*0.000002*	*0.000001*	*0.000000*	*0.000000*
49	*0.000000*	*0.000000*	*0.000002*	*0.000000*	*0.000000*	*0.000000*
50	*0.000000*	*0.000000*	*0.000001*	*0.000000*	*0.000000*	*0.000000*

Table A3.3: Predictive distribution of Y_{211} for the example in section 3.3.

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
0	*0.276792*	*0.169000*	*0.216651*	*0.239476*	*0.276118*	*0.278363*
1	*0.307547*	*0.300000*	*0.261883*	*0.295715*	*0.307381*	*0.308546*
2	*0.208542*	*0.267000*	*0.202399*	*0.216794*	*0.208751*	*0.208825*

Table A3.3 (cont.):

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
3	0.112725	0.158000	0.130760	0.125318	0.112982	0.112346
4	0.053732	0.070000	0.077990	0.064210	0.053913	0.053038
5	0.023755	0.025000	0.045194	0.031074	0.023857	0.023076
6	0.010027	0.007000	0.026235	0.014704	0.010077	0.009526
7	0.004113	0.002000	0.015495	0.006899	0.004137	0.003808
8	0.001660	0.000000	0.009328	0.003210	0.001670	0.001498
9	0.000664	0.000000	0.005679	0.001470	0.000668	0.000587
10	0.000265	0.000000	0.003457	0.000656	0.000267	0.000232
11	0.000106	0.000000	0.002085	0.000283	0.000106	0.000093
12	0.000042	0.000000	0.001237	0.000117	0.000043	0.000037
13	0.000017	0.000000	0.000719	0.000046	0.000017	0.000015
14	0.000007	0.000000	0.000409	0.000017	0.000007	0.000006
15	0.000003	0.000000	0.000227	0.000006	0.000003	0.000002
16	0.000001	0.000000	0.000123	0.000002	0.000001	0.000001
17	0.000000	0.000000	0.000065	0.000001	0.000000	0.000000
18	0.000000	0.000000	0.000033	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000016	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000008	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000004	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000002	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000001	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table A3.4: Predictive distribution of Y_{212} for the example in section 3.3.

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
0	0.003796	0.007000	0.024935	0.002771	0.003783	0.004060
1	0.015988	0.033000	0.067571	0.011997	0.015945	0.016636
2	0.037168	0.083000	0.104912	0.029512	0.037093	0.037968

Table A3.4 (cont.):

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
3	0.063061	0.140000	0.124550	0.053242	0.062967	0.063670
4	0.087235	0.175000	0.126666	0.077811	0.087145	0.087461
5	0.104325	0.175000	0.116685	0.097373	0.104259	0.104216
6	0.111776	0.147000	0.100290	0.108204	0.111745	0.111525
7	0.109867	0.105000	0.081818	0.109570	0.109870	0.109661
8	0.100734	0.066000	0.064120	0.103059	0.100765	0.100662
9	0.087218	0.037000	0.048768	0.091354	0.087266	0.087251
10	0.071977	0.018000	0.036341	0.077157	0.072032	0.072026
11	0.057028	0.008000	0.026767	0.062599	0.057083	0.057005
12	0.043631	0.004000	0.019633	0.049072	0.043681	0.043491
13	0.032384	0.00100	0.014423	0.037319	0.032427	0.032132
14	0.023409	0.000000	0.010652	0.027605	0.023444	0.023084
15	0.016532	0.000000	0.007926	0.019897	0.016559	0.016187
16	0.011438	0.000000	0.005946	0.013992	0.011458	0.011119
17	0.007770	0.000000	0.004497	0.009611	0.007785	0.007508
18	0.005193	0.000000	0.003423	0.006455	0.005203	0.005000
19	0.003420	0.000000	0.002616	0.004245	0.003427	0.003294
20	0.002223	0.000000	0.001999	0.002737	0.002228	0.002152
21	0.001428	0.000000	0.001520	0.001732	0.001431	0.001398
22	0.000907	0.000000	0.001145	0.001077	0.000910	0.000904
23	0.000571	0.000000	0.000850	0.000659	0.000572	0.000582
24	0.000356	0.000000	0.000620	0.000397	0.000357	0.000373
25	0.000220	0.000000	0.000443	0.000236	0.000221	0.000238
26	0.000135	0.000000	0.000309	0.000138	0.000136	0.000151
27	0.000083	0.000000	0.000210	0.000079	0.000083	0.000096
28	0.000050	0.000000	0.000139	0.000045	0.000050	0.000060
29	0.000030	0.000000	0.000090	0.000025	0.000030	0.000037
30	0.000018	0.000000	0.000056	0.000014	0.000018	0.000022
31	0.000011	0.000000	0.000034	0.000007	0.000011	0.000014
32	0.000006	0.000000	0.000020	0.000004	0.000006	0.000008

Table A3.4 (cont.):

y	Exact	MLE	N. Approx. 1	N. Approx. 2	Laplace	Gibbs
33	0.000004	0.000000	0.000012	0.000002	0.000004	0.000005
34	0.000002	0.000000	0.000007	0.000001	0.000002	0.000003
35	0.000001	0.000000	0.000004	0.000000	0.000001	0.000001
36	0.000001	0.000000	0.000002	0.000000	0.000001	0.000001
37	0.000000	0.000000	0.000001	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000001	0.000000	0.000000	0.000000
39	0.000000	0.000000	0.000003	0.000000	0.000000	0.000000
40	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Table A3.5: Predictive distribution of $Z = Y_{21.2} - Y_{21.1}$ for the example in section 3.3.

z	MLE	Laplace	z	MLE	Laplace
-10	0.000000	0.000001	4	0.139059	0.115970
-9	0.000002	0.000002	5	0.105818	0.116895
-8	0.000009	0.000006	6	0.070627	0.108278
-7	0.000046	0.000018	7	0.041849	0.093573
-6	0.000205	0.000057	8	0.022247	0.076329
-5	0.000813	0.000180	9	0.010708	0.059309
-4	0.002828	0.000565	10	0.004703	0.044219
-3	0.008515	0.001715	11	0.001898	0.031817
-2	0.021858	0.004936	12	0.000708	0.022201
-1	0.047138	0.012980	13	0.000246	0.015083
0	0.084396	0.029486	14	0.000080	0.010009
1	0.124826	0.054113	15	0.000024	0.006506
2	0.153273	0.081292	16	0.000007	0.004153
3	0.158115	0.103647	17	0.000002	0.002608

Table A3.5 (cont.):

z	MLE	Laplace	*z*	MLE	Laplace
18	*0.000000*	*0.001614*	*26*	*0.000000*	*0.000024*
19	*0.000000*	*0.000986*	*27*	*0.000000*	*0.000014*
20	*0.000000*	*0.000596*	*28*	*0.000000*	*0.000008*
21	*0.000000*	*0.000356*	*29*	*0.000000*	*0.000004*
22	*0.000000*	*0.000211*	*30*	*0.000000*	*0.000002*
23	*0.000000*	*0.000124*	*31*	*0.000000*	*0.000001*
24	*0.000000*	*0.000072*	*32*	*0.000000*	*0.000001*
25	*0.000000*	*0.000042*	*33*	*0.000000*	*0.000000*

BIBLIOGRAPHY

(A) Papers and books referred to in the thesis:

Aitchison, J. (1975), "Goodness of Prediction Fit", *Biometrika*, **62**, 547-554.

Aitchison, J. and Dunsmore, I.R. (1975), "*Statistical Prediction Analysis*", Cambridge University Press.

Barnett, V. and Wright, D.E. (1990), "*Accident Predictive Relations from Short Period Flow Counts - Final Report*", Statistical Services Unit, University of Sheffield.

Barnett, V. and Wright, D.E. (1992), "Biomedical Applications for Generalized Linear Functional Poisson Model", *Journal of Applied Statistics*, **19**, 41-47.

Bernardo, J.M. and Smith, A.F.M. (1994), "*Bayesian Theory*", Wiley: Chichester.

Bhattacharjee, S.K. and Dunsmore, I.R. (1991), "The Influence of Variables in a Logistic Model", *Biometrika*, **78**, 851-856.

Bhattacharjee, S.K. and Dunsmore, I.R. (1995), "The Predictive Influence of Variables in a Normal Regression Model", *Journal of Information & Optimization Sciences*, **16**, No.2, 327-334.

Dellaportas, P. and Smith, A.F.M. (1993), "Bayesian Inference for Generalized Linear and Proportional Hazards Models via Gibbs Sampling", *Appl. Statist.*, **3**, 443-459.

Dunsmore, I.R. and Robson, D.J. (1992), "*Accident Predictive Relations from Short Period Counts: Extensions to Existing Software - Final Report*", Statistical Services Unit, University of Sheffield.

Dunsmore, I.R. and Robson, D.J. (1994a), "*Prediction in some Poisson Errors in Variables Models*", *Research Report 427/94*, Department of Probability and Statistics, University of Sheffield.

Dunsmore, I.R. and Robson, D.J. (1994b), "*Poisson Predictions in a Crossover Model*", *Research Report 428/94*, Department of Probability and Statistics, University of Sheffield.

Dunsmore, I.R. and Robson, D.J. (1997), "Prediction in some Poisson Errors in Variables Models", *Scand. J. Statist.*(to appear).

Dunsmore, I.R., Magalhaes, F.J.M. and Robson, D.J. (1997), "Prediction of Treatment Effects in a Biased Allocation Model", Submitted to *Biometrika.*

Gaver, D.P. and O'Muircheartaigh, I.G. (1987)."Robust Empirical Bayes Analyses of Events Rates". *Technometrics*, **29**, 1-15.

Geisser, S. (1993), "*Predictive Inference: An Introduction*", Chapman & Hall: New York.

Gelfand, A.E. and Smith, A.F.M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities", *J. Amer. Statist. Ass.*, **85**, 398-409.

Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990), "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling", *J. Amer. Statist. Ass.*, **85**, 972-985.

Gilks, W.R. and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling", *Appl. Statist.*, **41**, 337-348.

Kullback, S. and Leibler, R.A. (1951), "On Information and Sufficiency", *Ann. Math. Statist.*, **22**, 79-86.

O'Hagan, A. (1994), "*Bayesian Inference - Kendall's Advanced Theory of Statistics*", Vol. 2B, London: Edward Arnold.

Robbins, H. and Zhang, C.H. (1988), "Estimating a Treatment Effect Under Biased Sampling", *Proc. Natl. Acad. Sci. USA* **85**, 3670-3672.

Robbins, H. and Zhang, C.H. (1989), "Estimating the superiority of a drug to a placebo when all and only those patients at risk are treated with the drug", *Proc. Natl. Acad. Sci. USA* **86**, 3003-3005.

Robbins, H. and Zhang, C.H. (1991), "Estimating a Multiplicative Treatment Effect Under Biased Allocation", *Biometrika*, **78**, 349-354.

Svensson, A. (1981), "On a goodness-of-fit test for multiplicative Poisson models", *Ann. Statist.* **9**, 697-704.

(B) Other references, not used in the thesis, which provide further information on some background details (e.g. aspects of posterior normality, Laplace approximations, etc.), on alternative approaches not considered here (e.g. predictive likelihood), on various aspects of computational Bayesian methodology, and on Bayesian approaches to generalised linear models.

Aitchison, J. and Sculthorpe, D. (1965), "Some problems of statistical prediction", *Biometrika*, **52**, 469-483.

Alanko, T. and Duffy, J.C. (1996), "Compound binomial distributions for modelling consumption data", *Statistician*, **45**, 269-286.

Albert, J.H. (1988), "Computational methods using a Bayesian hierarchical generalized linear model", *J. Amer. Statist. Assoc.*, **83**, 1037-1044.

Albert, J.H. and Pepple, P.A. (1989). "A Bayesian approach to some overdispersion models", *Canadian J. Statist.*, **17**, 333-344.

Albert, J.H. (1992), "A Bayesian analysis of a Poisson random effects model for home run hitters", *The American Statistician*, **46**, 246-253.

Anderson, D.A. and Aitkin, M. (1985), "Variance component models with binary response: interviewer variability", *J. R. Statist. Soc.* B, **47**, 203-210.

Barndorff-Nielsen, O. (1983), "On a formula for the distribution of the maximum likelihood estimator", *Biometrika*, **70**, 343-465.

Breslow, N.E. (1984), "Extra-Poisson variation in log-linear models", *Appl. Statist.*, **33**, 38-44.

Breslow, N. (1990), "Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models", *J. Amer. Statist. Assoc.*, **85**, 565-571.

Breslow, N.E. and Clayton, D.G. (1993), "Approximate inference in generalized linear mixed models", *J. Amer. Statist. Assoc.*, **88**, 9-25.

Casella, G. and George, E.I. (1992), "Explaining the Gibbs sampler", *The American Statistician*, **46**, 167-174.

Clevenson, M.L. and Zidek, J.V. (1975), "Simultaneous estimation of the means of independent Poisson laws", *J. Amer. Statist. Assoc.*, **70**, 698-705.

Cowles, M.K. and Carlin, B.P. (1996), "Markov chain Monte Carlo convergence diagnostics: a comparative review", *J. Amer. Statist. Assoc.*, **91**, 883-904.

Cui, L., Tanner, M.A., Sinha, D. and Hall, W.J. (1992), "Comment: Monitoring convergence of the Gibbs sampler: Further experience with the Gibbs stopper", *Statistical Science*, **7**, 483-486.

Davison, A.C. (1992), "Treatment Effect Heterogeneity in Paired Data", *Biometrika*, **79**, 463-474.

Diaconis, P. and Efron, B. (1985), "Testing for independence in a two-way table: new interpretations of the chi-square statistic", *Ann. Statist.*, **13**, 845-874.

Dunsmore, I.R. (1976), "Asymptotic prediction analysis", *Biometrika*, **63**, 627-630.

Epstein, L.D. and Fienberg, S.E. (1991), "Using Gibbs sampling for a Bayesian inference in multidimensional contingency tables", Computer Science and Statistics: Proceedings of the 23th Symposium on the Interface, pp. 215-223.

Faulkenberry, G.D. (1973), "A method of obtaining prediction intervals", *J. Amer. Statist. Assoc.*, **68**, 433-435.

Geisser, S. (1971), "The inferential use of predictive distributions", in: *Foundations of Statistical Inference*, Eds. V. P. Godambe and D. A. Sprott, pp. 456-469, Toronto: Holt, Rinehart and Winston.

George, E.I., Makov, U.E. and Smith, A.F.M. (1994), "Fully Bayesian hierarchical analysis for exponential families via Monte Carlo computation", in *Aspects of Uncertainty: A Tribute to D. V. Lindley* (eds. P.R.Freeman and A. F. M. Smith), pp. 181-199, Chichester: Wiley.

Goldstein, H. (1995). "*Multilevel Statistical Models*". London: Edward Arnold.

Good, I.J. (1980), "Some history of the hierarchical Bayesian methodology", in: *Bayesian Sattistics*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds, Valencia, Spain, University Press, pp. 489-519.

Hastings, W.K. (1979), "Monte Carlo sampling methods using Markov chains and their applications", *Biometrika*, **57**, 97-109.

Hinkley, D. (1979), "Predictive likelihood", *Ann. Statist.*, **7**, 718-728.

Hsu, J.S.J. (1995), "Generalized Laplacian approximations in Bayesian inference", *Canadian J. Statist.*, **23**, 399-410.

Hsu, J.S.J. and Leonard, T. (1997), "Hierarchical Bayesian semiparametric procedures for logistic regression", *Biometrika*, **84**, 85-93.

Hsu, J.S.J., Leonard, T., and Tsui, K.W. (1991), "Statistical inference for multiple choice tests", *Psychometrika*, **56**, 327-348.

Johnson, R.A. (1967), "An asymptotic expansion for the posterior distributions", *Ann. Math. Statist.*, **38**, 1899-1906.

Johnson, R.A. (1970), "Asymptotic expansions associated with posterior distributions", *Ann. Math. Statist.*, **41**, 851-864.

Kalbfleisch, J.D. and Sprott, D.A. (1970), "Application of likelihood methods to models involving large number of parameters (with Discussion)", *J. R. Statist. Soc.* B, **32**, 175-208.

Kass, R.E., Tierney, L. and Kadane, J.B. (1988), "Asymptotics in Bayesian computation", in: *Bayesian Statistics 3*, Eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, pp. 261-278, Oxford University Press.

Kass, R.E. and Steffey, D. (1989), "Approximate Bayesian-inference in conditionally independent hierarchical-models (parametric empirical Bayes models)", *J. Amer. Statist. Assoc.*, **84**, 717-726.

Laird, N.M. and Louis, T.A. (1991), "Smoothing the non-parametric estimate of a prior distribution by roughening: a computational study", *Comp. Statist. & Data Anal.*, **12**, 27-37.

Lee, Y. and Nelder, J.A. (1996), "Hierarchical generalized linear models (with Discussion)", *J. R. Statist. Soc.* B, **58**, 619-678.

Lejeune, M. and Faulkenberry, G.D. (1982), "A simple predictive density function", *J. Amer. Statist. Assoc.*, **77**, 654-659.

Leonard, T. (1976), "Some alternative approaches to multiparameter estimation", *Biometrika*, **63**, 69-76.

Leonard, T. (1982), "Comments on 'A simple predictive density function'", *J. Amer. Statist. Assoc.*, **77**, 657-658.

Leonard, T. (1996), "On exchangeable sampling distributions for uncontrolled data", *Statist. Prob. Letters*, **26**, 1-6.

Leonard, T. and Hsu, J.S.J. (1994), "The Bayesian analysis of categorical data - a selective review", in: *Aspects of Uncertainty: A Tribute to D. V. Lindley* (eds. P. R. Freeman and A. F. M. Smith), pp. 283-310, Chichester: Wiley.

Leonard, T. and Hsu, J.S.J. (1996), "On small sample Bayesian inference and sequential design for quantal response curves", in: *Modeling Prediction: Honoring Seymour Geisser*, New York: Springer Verlag.

Leonard, T. Hsu, J.S.J. and Tsui, K.W. (1989), "Bayesian marginal inference", *J. Amer. Statist. Assoc.*, **84**, 1051-1058.

Leonard, T., Hsu, J.S.J. and Ritter, C. (1994), "The Laplacian *t*-approximation in Bayesian inference", *Statistica Sinica*, **4**, 127-142.

Leonard, T. and Novick, M.R. (1986), "Bayesian full rank marginalization for two-way contingency tables", *J. Educ. Behav. Statist.*, **11**, 33-56.

Lindley, D. V. (1971), " *Bayesian Statistics, A Review*", Philadelphia: S.I.A.M.

Lindley, D.V. (1980), "Approximate Bayesian methods", Bayesian Statistics: Proceedings of the First International Meeting held in Valencia, pp. 223-237.

Lindstrom, M.J. and Bates, D.M. (1990), "Nonlinear mixed effects models for repeated measures data", *Biometrics*, **46**, 673-687.

Longford, N.T. (1993), " Random Coefficients Models", Oxford: OUP.

Naylor, J.C. and Smith, A.F.M. (1982), "Applications of a method for the efficient computation of posterior distributions", *Appl. Statist.*, **31**, 214-225.

Paul, S.R. and Plackett, R.L. (1978), "Inference sensitivity for Poisson mixtures", *Biometrika*, **65**, 591-602.

Prentice, R.L. and Zhao, L.P. (1991), "Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses", *Biometrics*, **47**, 825-839.

Racine-Poon, A. (1985), "A Bayesian approach to nonlinear random effects models", *Biometrics*, **41**, 1015-1023.

Raftery, A.E. (1996), "Approximate Bayes factors and accounting for model uncertainty in generalised linear models", *Biometrika*, **83**, 251-266.

Reilly, P.M. (1976), "The numerical computation of posterior distributions", *Appl. Statist.*, **25**, 201-209.

Ritter, C. and Tanner, M.A. (1992), "Facilitating the Gibbs sampler: the Gibbs stopper and the Griddy-Gibbs sampler", *J. Amer. Statist. Assoc.*, **87**, 861-868.

Schwarz, G. (1978), "Estimating the dimension of a model", *Ann. Statist.*, **6**, 461-464.

Searle, S.R. (1971), " *Linear Models*", Wiley: Chichester.

Senn, S. (1993), " *Cross-Over Trials in Clinic Research*", Wiley: Chichester.

Shun, Z.M. and McCullagh, P. (1995), "Laplace approximation of high-dimensional integrals", *J. R. Statist. Soc.* B, **57**, 749-760.

Solomon, P.J. and Cox, D.R. (1992), "Nonlinear component of variance models", *Biometrika*, **79**, 1-11.

Sun, L., Hsu, J.S.J., Guttman, I. and Leonard, T. (1996), "Bayesian methods for variance component models", *J. Amer. Statist. Assoc.*, **91**, 743-752.

Tavare, S. and Altham, P.M.E. (1983), "Serial dependence of observations leading to contingency tables, and corrections to chi-squared statistics", *Biometrika*, **70**, 139-144.

Tierney, L. (1994), "Markov chains for exploring posterior distributions (with Discussion)", *Ann. Statist.*, **22**, 1701-1762.

Tierney, L. and Kadane, J.B. (1986), "Accurate approximations for posterior moments and marginal densities", *J. Amer. Statist. Assoc.*, **81**, 82-86.

Tierney, L., Kass, R.E. and Kadane, J.B. (1989), "Approximate marginal densities of nonlinear functions", *Biometrika*, **76**, 425-433.

Tsui, K.W. (1979), "Multiparameter estimation of discrete exponential distributions", *Canadian J. Statist.*, **7**, 193-200.

Tsui, K.W. (1981), "Simultaneous estimation of several Poisson parameters under squared error loss", *Ann. Inst. Statist. Math.*, **33**, 215-223.

Tsui, K.W. (1984), "Robustness of Clevenson-Zidek-type estimators", *J. Amer. Statist. Assoc.*, **79**, 152-157.

Tsui, K.W. and Press, S.J. (1982), "Simultaneous estimation of several Poisson parameters under K-normalized squared error loss", *Ann. Statist.*, **10**, 93-100.

Williams, D.A. (1982), "Extra-binomial variation in logistic linear models", *Appl. Statist.*, **31**, 144-148.

Wolfinger, R. (1993), "Laplace's approximation for nonlinear mixed models", *Biometrika*, **80**, 791-795.

Wong, W.H. and Li, B. (1992), "Laplace expansion for posterior densities of nonlinear functions of parameters", *Biometrika*, **79**, 393-398.

Zeger, S.L. and Karim, M.R. (1991), "Generalized linear models with random effects; a Gibbs sampling approach", *J. Amer. Statist. Assoc.*, **86**, 79-86.

Zellner, A. and Rossi, P.E. (1984), "Bayesian analysis of dichotomous quantal response models", *J. Econometrics*, **25**, 365-393.

Zellner, A. and Min, C.K. (1995), "Gibbs sampler convergence criteria", *J. Amer. Statist. Assoc.*, **90**, 921-927.