UNIVERSITY OF SOUTHAMPTON

SOCIAL STATISTICS AND DEMOGRAPHY

Social Sciences

Applications of the dependence ratio association measure for multivariate categorical data

by

Robert F. North

Thesis for the degree of Doctor of Philosophy

Abstract

Traditional univariate regression techniques assume independence between observations so are therefore not appropriate for the analysis of multivariate responses. A multivariate response refers to cases where there are more than one response for each unit (cluster). In order to not violate the independence assumption, the entire response profile of each unit is treated as the response. The focus of this thesis is on the analysis of multivariate categorical data using likelihood-based approaches. In addition, population-averaged models are of interest, whereby the effects of the explanatory variables are averaged over the population. The nature of likelihood-based approaches means that the full joint distribution of the multivariate categorical responses must be specified. This thesis puts equal emphasis on both the univariate means (first-order moments) and the associations within a cluster. The likelihood-based approach used in this thesis uses the dependence ratio association measure (Ekhlom, Smith and McDonald, 1995) as opposed to the commonly used odds ratio. A key advantage of the dependence ratio is its ability to cope computationally with larger cluster sizes than the odds ratio.

Papers 1 and 2 consider frequentist applications using maximum likelihood estimation. Paper 1 considers the dependence ratio in the context of square tables (two-way contingency tables with the same row and column categories) as well as extensions to larger contingency tables. Existing models, such as symmetry, are discussed and expressed in terms of dependence ratios where possible. Paper 2 of this thesis considers a specific dataset on the hand joints of patients who are affected with psoriatic arthritis. The (binary) response of interest is damage in the hand joints of each patient (at their last clinic visit). The dependence ratio approach gives equal emphasis to the first-order moments and the associations within a patient. In addition to the dependence ratio approach, a generalised estimating equations (GEE) approach is also presented and contrasted with the dependence ratio approach. The GEE approach focuses on the first-order moments and in this analysis misses some notable conclusions by treating the associations as a nuisance.

Paper 3 considers a Bayesian approach with dependence ratios. Although dependence ratios have been used in a Bayesian context before, it is believed that only empirical Bayes approaches have been considered (Good, 1956; Du Mouchel and Pregibon, 2001). Paper 3 considers a Bayesian approach using dependence ratios and traditional Bayesian approaches whereby prior distributions are considered (for the first-order moments and dependence ratios). The simplest bivariate binary response is the main focus of the paper with emphasis on obtaining non-informative priors. Specifying the dependence ratio to have a uniform prior between its upper and lower bounds was found to be the most appropriate non-informative prior for the bivariate binary case.

Contents

List of Tables

List of Figures

Declaration Of Authorship

I, Robert North, declare that the thesis entitled: Applications of the dependence ratio association measure for multivariate categorical data and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Signed: .....

Date

Acknowledgements

I would firstly like to thank Peter Smith and Mac McDonald for the supervision and time they have dedicated to advising me throughout the last three years. Their knowledge has been of great benefit for not only improving my own understanding but also for determining the best ways in which to take the research forward.

I am grateful to the Economic and Social Research Council (ESRC) for providing funding for this work. I would also like to thank Vern Farewell and Dafna Gladman for allowing the use of the psoriatic arthritis data in paper 2 of this thesis. Finally, I also express thanks to Aidan O'Keeffe for his advise when the dataset was first received.

Chapter 1

Introduction

1.1 Background

Longitudinal or otherwise clustered data can be found in a range of disciplines. They both relate to situations where there are more than one observation for each individual unit (cluster). Longitudinal data are typically used to refer to studies where repeated measurements are observed over time. A typical example of this would be patients in a clinical trial who give a number of measurements over a given time period. In contrast, clustered data typically refers to studies where there are repeated measurements with no clear time element. For example, studies often involve subjects answering a number of similar questions in a questionnaire.

Longitudinal and clustered data can also be referred to as multivariate data. Due to the correlated nature of multivariate data, standard univariate techniques cannot be used since they assume independence. One way to overcome this difficulty is to treat each unit's entire response profile as the response and use multivariate techniques. For normally distributed data, substantial work has been undertaken and there are well established methods (such as those described in Diggle et al. 2002). However, for categorical data, which is the focus of this thesis, research is being more actively carried out. Categorical data refers to situations where the response of interest has only a small number of categories. This could be either nominal or ordinal categorical data. Nominal is when there is no order to the categories and this also includes dichotomous (binary) data. An example of a nominal variable would be religion followed (Christianity, Muslim, etc). In contrast, ordinal categorical data are when the order of the categories is meaningful. A common example of an ordinal categorical variable is when subjects are asked to rate each question in a questionnaire on a scale from 1-5 (ranging from strongly agree to strongly disagree, for example). Ordinal categorical data can also be encountered when there is a discrete variable with few categories or when continuous variables are grouped into categories (Agresti, 2002).

There are two main approaches to analysing multivariate categorical data. Firstly, likelihood based approaches such as the population-averaged approach (discussed in Section 1.2), where emphasis is typically placed on modelling the associations between

the repeated responses (using an association model) and the marginal regression (first-order moments). Secondly, the generalised estimating equations (GEE) approach (Liang and Zeger, 1986) which is a multivariate extension of quasi-likelihood. This method treats the association parameters as a nuisance and emphasis is placed on the marginal regression. In contrast to likelihood approaches, the GEE approach does not specify the entire joint distribution. This approach is discussed in greater detail in Section 1.2.1 as well as paper 2 of this thesis. However, the focus of this thesis is on likelihood based methods in both a frequentist context (papers 1 and 2) and Bayesian context (paper 3). A detailed description of the outline to this thesis is given in Section 1.5.

The joint distribution of a multivariate normal response has the useful property that it is completely specified by the first two moments (the univariate means and the variance-covariance matrix respectively). In contrast, the joint distribution of a multivariate categorical response is more complicated as it is not specified by the first two moments. This is illustrated in Table 1.1 which shows the total number of parameters describing the associations within a cluster (for the saturated case) increases dramatically as the number of subunits within a cluster increases, for a binary response. If the number of subunits within a cluster is of size q, the total number of association parameters is given by $2^q - q - 1$.

Table 1.1: Number of association parameters by cluster size, for a binary response

Cluster size	Number of association parameters
2	1
3	4
4	11
5	26
6	57

Table 1.1 shows that when the response is binary and the number of subunits within a cluster is five or more, that having one parameter for each degree of freedom (for the saturated case) is often not feasible. This is due to potential computational difficulties as well as making model interpretation difficult. Clearly restrictions need to be applied in these cases to reduce the number of association parameters to a manageable amount. When the categorical response has more than two categories, restrictions are particularly relevant.

1.2 Maximum likelihood approaches to correlated categorical data

Many statisticians prefer maximum likelihood approaches (Liang, Zeger and Qaqish, 1992) for analysing multivariate categorical data. One likelihood based approach is

where the effects of the covariates are averaged over all subjects (units) in the dataset. These models are known as population-averaged models. There are also subject-specific (cluster-specific) based approaches. They differ from population-averaged models in the sense that they estimate subject-specific effects. Conditional maximum likelihood is one such approach, but it is limited since it can only estimate within-subject effects (Agresti, 2002). A preferable approach is generalized linear mixed models. This treats the subject effects as random effects and can cope with both between and within-subject effects (for the fixed effects).

Both population-averaged and subject-specific approaches have their use in particular situations. Population-averaged models are preferable if marginal effects are of main interest to the researcher. In contrast, subject-specific models are preferred if the researcher would like to estimate subject-specific effects as well as their associated variability. Agresti (2002) gives a more detailed discussion of when to use each approach. However, the focus of this thesis is on the use of population-averaged models.

1.2.1 Problems with maximum likelihood and the GEE approach

For many years, population-averaged models using maximum likelihood estimation (with the traditional odds ratio as the measure of association) suffered from the drawback of not being able to cope computationally with large cluster sizes. As a general rule, maximum likelihood estimation of cluster sizes greater than five is not feasible with odds ratios (Lesaffre et al. 2000). This lead to the development of the quasi-likelihood (GEE) approach (Liang and Zeger, 1986). In recent years, this method has arguably become the most popular method for the analysis of multivariate categorical data. This is partly due to the fact that it can cope with larger cluster sizes, but also because it is the default method for modelling multivariate categorical data in many statistical packages, such as SPSS.

The GEE approach treats the associations as a nuisance with emphasis placed on the marginal regression and moments of higher-order than two are not specified. It is often referred to as a marginal model since emphasis is placed on the marginal regression. Despite the fact the GEE approach is similar to the multivariate normal distribution in the sense that the first two moments fully describe the joint distribution, many statisticians believe that the association of the responses should be analysed further in order to determine the mechanisms that are generating the associations within a cluster (Lindsey and Lambert, 1998). Consequently, an alternative to the odds ratio association measure for maximum likelihood estimation was sought and this is discussed in Section 1.2.2.

1.2.2 Association measures for maximum likelihood estimation in population-averaged models

A number of association measures have been proposed for analysing multivariate categorical data using population-averaged models and maximum likelihood estimation. This section discusses the traditional odds ratio and the more recently proposed dependence ratio (Ekholm, Smith and McDonald, 1995) which offers a solution to the problem outlined in Section 1.2.1. In order to compare the two measures, consider the bivariate binary response $\mathbf{Y} = (Y_1, Y_2)$. The nature of maximum likelihood approaches means that the joint distribution must be specified. For all population-averaged approaches that use maximum likelihood, the first-order moments are regressed on any relevant explanatory variables (using a appropriate link function). For the bivariate binary case, the first-order moments (assuming no explanatory variables) are specified as:

$$\mu_k = pr(Y_k = 1), \tag{1.1}$$

where $k = 1, 2$.

Moments of second-order or higher inform us about the associations within a cluster and are parameterised differently depending on the association measure considered. The second-order moment for the bivariate binary case is given by:

$$\mu_{12} = pr(Y_1 = 1, Y_2 = 1). \tag{1.2}$$

Table 1.2 shows the relevant 2 by 2 table of cell probabilities for the bivariate binary response with four possible response profiles: $(1,1), (1,0), (0,1), (0,0)$. The joint distribution is completely specified by the first-order moments (μ_1, μ_2) and the second-order moment (μ_{12}). Although μ_{12} represents the association between Y_1 and Y_2, it gives no direct link to the baseline of independence. In contrast, the odds ratio and dependence ratio do relate the higher order moments to independence and they are now discussed in more detail.

Table 1.2: Cell probabilities for a bivariate binary response

	Y_2		
Y_1	1	0	Total
1	μ_{12}	$\mu_1 - \mu_{12}$	μ_1
0	$\mu_2 - \mu_{12}$	$1 - \mu_1 - \mu_2 + \mu_{12}$	$1 - \mu_1$
Total	μ_2	$1 - \mu_2$	1

The odds ratio is the traditional association measure for population-averaged approaches to analysing multivariate categorical responses using maximum likelihood. Numerous different odds ratio parameterisations have been suggested such as the mixed parameterisation by Fitzmaurice and Laird (1993). In this approach, the first-order moments

are combined with conditional log-odds ratios. Fitzmaurice and Laird proposed a two-stage iterative procedure for finding the maximum likelihood estimates. The odds ratio parameterisations will often differ for moments of higher-order than two. However, in the simplest bivariate binary case, the same odds ratio is obtained regardless of the formulation. This is specified as:

$$\chi = \chi_{12} = \frac{\mu_{12}(1-\mu_1-\mu_2+\mu_{12})}{(\mu_2-\mu_{12})(\mu_1-\mu_{12})} = \frac{pr(Y_1=1,Y_2=1)pr(Y_1=0,Y_2=0)}{pr(Y_1=0,Y_2=1)pr(Y_1=1,Y_2=0)}. \tag{1.3}$$

The dependence ratio was introduced by Ekholm et al. (1995) as an advantageous alternative to the odds ratio. For the bivariate binary response, the dependence ratio is given by:

$$\tau = \tau_{12} = \frac{\mu_{12}}{\mu_1\mu_2} = \frac{pr(Y_1=1,Y_2=1)}{pr(Y_1=1)\,pr(Y_2=1)}. \tag{1.4}$$

If $Y_k = 1$ is the probability of a success as opposed to failure ($Y_k = 0$), the dependence ratio is defined as the joint success probability divided by the joint success probability under independence. For both the odds ratio and the dependence ratio, a value of 1 indicates independence. In addition, values less than 1 indicate negative association and values greater than 1 indicate positive association.

The focus of this thesis is on the use of the dependence ratio and in particular extending its use to relevant applications. A key advantage of the dependence ratio approach (discussed in detail in Section 1.2.3) is its ability to cope computationally with large cluster sizes. In addition to putting equal emphasis on modelling the marginal regression and the associations, the dependence ratio approach uses relevant association structures to reduce the number of association parameters (where necessary) and consequently to provide a detailed account of the mechanisms behind the associations. These structures are discussed in detail in Section 1.2.3.1.

1.2.3 The dependence ratio approach

The dependence ratio can be extended from the bivariate case described previously to higher-order moments. For example, for the binary case with three repeated observations, the three-way dependence ratio is defined as:

$$\tau_{123} = \frac{\mu_{123}}{\mu_1\mu_2\mu_3} = \frac{pr(Y_1=1,Y_2=1,Y_3=1)}{pr(Y_1=1)pr(Y_2=1)pr(Y_3=1)}. \tag{1.5}$$

Higher-order dependence ratios can be expressed similarly. Dependence ratios for cases where there are more than two response categories are discussed in paper 1 of this thesis.

The dependence ratio approach combines a marginal regression model and an association model using the profile probability, π_i, which is defined as:

$$\pi_i = pr(Y_{i1} = y_{i1}, \ldots, Y_{iq} = y_{iq}), \tag{1.6}$$

where Y_{ik} represents the response for unit $i = 1, \ldots, n$ at subunit $k = 1, \ldots, q$.

The marginal regression model consists of the first-order moments (univariate means) regressed on explanatory variables using a relevant link function such as the logit link function in the binary response case. The association model uses the second-order and higher moments. A convenient feature of the dependence ratio approach is that these moments can be expressed as the product of the first-order moments and dependence ratios of all orders. This is because, in contrast to the second-order and higher moments, dependence ratios are comparable to the baseline of mutual independence. In addition, in contrast to the odds ratio, the profile probability can be expressed in closed form in terms of the regression and association parameters, where the regression parameters are the first-order moments and the association parameters are the dependence ratios of all orders. This consequently allows the likelihood and log-likelihood functions to be specified for maximum likelihood estimation and also gives the dependence ratio approach a key advantage over the odds ratio approach (see Section 1.3 for more details).

1.2.3.1 Association structures

In many cases, particularly with large cluster sizes, the number of association parameters will need to be reduced to a more manageable number. This is commonly undertaken using a particular association structure such as those defined by Ekholm, Smith and McDonald (2000). Alternatively, constraints may be imposed on the dependence ratios based on the values of the observed dependence ratios or theoretical considerations.

The association structures that are considered in this thesis are now discussed.

Independence (I):

Independence assumes that the observations within a cluster are independent. Consequently, all dependence ratios are equal to one. Independence is rarely applicable to longitudinal or clustered data.

Necessary factor (N):

In certain studies, there may be a particular group of individuals (units) who always give the same response throughout the study. In this situation, a necessary factor association structure is often appropriate. This structure separates the dataset into those individuals that have and those that do not have the factor that is necessary for a (typically) positive response in the binary case or to be (typically) greater than the

smallest response category for a multicategory response. The responses are conditionally independent given the necessary factor. This association structure is discussed further in papers 1 and 2 of this thesis.

Latent binary factor (L):

A latent binary structure is appropriate if every unit has a realisation of a latent binary variable such that the population can be divided into two groups with different response category probabilities (given the same covariate values for the regression model). In addition, responses within a cluster are conditionally independent given L. The structure is often appropriate if an important dichotomous covariate has been omitted (Ekholm et al. 2003). See papers 1 and 2 of this thesis for more details.

Latent binary factor within a necessary factor (NL):

It may be useful to combine the N and L association structures. For example, after fitting N, an improved model fit may be achieved by having a latent binary structure to operate within a necessary factor association structure (denoted NL). In other words, of those individuals that have the necessary factor, there are two groups of individuals with different response probabilities (given the same values for the regression variables). This association structure is considered in this thesis but often had a unsatisfactory fit due to negative profile probabilities (see Section 1.2.4).

Markov structures

Although not relevant to this thesis, Markov structures are discussed. When there is a natural ordering to the responses such as over time, a first-order Markov structure may be appropriate. This is denoted by M_1. The structure has the condition that a particular observation at time t_{k+1} is only influenced by the response at time t_k and is independent of all previous observations, given the response at time t_k. For a binary response, this condition can be expressed as:

$$pr(Y_{it_{k+1}} = 1 \mid Y_{it_1}, \ldots, Y_{it_k}) = pr(Y_{it_{k+1}} = 1 \mid Y_{it_k}), \tag{1.7}$$

where Y_{it_k} represents the response for unit $i = 1, \ldots, n$ at subunit $k = 1, \ldots, q$.

For a first-order Markov structure with a binary response, $(q-1)$ adjacent two-way dependence ratios have to be specified, where q is the number of subunits within a cluster. If we extend this to multicategory responses, $(q-1)(f-1)^2$ two-way adjacent dependence ratios have to be specified, where f is the number of categories in the response. Various constraints can be imposed on the dependence ratios in addition to the first-order Markov structure outlined above such as equality of dependence ratios over time and equality of dependence ratios for certain response categories. The structure is appropriate when there are more than two repeated observations and in order to

satisfy the Markov condition, the 3-way dependence ratios should satisfy the following expression (for a binary response):

$$\tau_{t_{k-1}t_kt_{k+1}} = \tau_{t_{k-1}t_k}\tau_{t_kt_{k+1}}. \tag{1.8}$$

This can be used as a diagnostic check that the first-order Markov structure is adequate. Another structure which can be used is a second-order Markov structure, denoted M_2. For binary responses, $(q-2)$ trivariate dependence ratios are specified (Ekholm et al. 2000).

Finally, another model that is often considered is to have a particular Markov structure (typically first-order) occurring within a necessary factor. In other words, given the individual has the necessary factor, the responses within a cluster follow a first-order Markov structure. This structure is denoted NM and is discussed in detail in Ekholm, Jokinen and Kilpi (2002).

1.2.4 Comparing the odds ratio and the dependence ratio

The following points now aim to summarise in detail the advantages and drawbacks of both the dependence ratio and the odds ratio approaches.

- Interpretation: The dependence ratio does not grow in complexity for the higher-order moments. In contrast, the interpretation of the odds ratio does grow in complexity for these cases since, for example, a 3-way odds ratio is the ratio of two conditional two-way odds ratios (Ekholm, 2003).
- Explicit solution: The odds ratio approach suffers from the drawback of having no closed form solution for the joint distribution (in terms of the first-order moments and odds ratios) when the number of subunits in a cluster (q) is greater than 2 (Ekholm, 2003) since these have to be solved iteratively. In fact, even the simplest bivariate binary case requires the solving of a quadratic in order for the joint distribution to be specified in terms of μ_1, μ_2 and χ. In contrast, the joint distribution can still be expressed in closed form for the dependence ratio approach (in terms of the first-order moments and dependence ratios) when $q > 2$. For the bivariate binary case, the joint distribution is completely specified by μ_1, μ_2 and τ by utilising the transformation $\mu_{12} = \mu_1\mu_2\tau$ so it therefore does not require the solving of a quadratic. This gives the dependence ratio approach a major computational advantage (see Section 1.3 for more details).
- Large cluster sizes: The dependence ratio approach has the distinct advantage of being able to cope computationally with large cluster sizes. In contrast, the

odds ratio approach typically cannot cope with cluster sizes that are larger than approximately five (Lesaffre et al. 2000).

- Event specific association: The dependence ratio measures the strength of the association between specific events whereas the odds ratio focuses on the association between the random variables. The dependence ratio is therefore more similar to the relative risk than the odds ratio since it is based on a ratio of probabilities. The relative risk is often considered to have a more favorable interpretation than the odds ratio (Greenland 1987; Sackett, Deeks and Altman 1996; Davies, Crombie and Tavakoli 1998).

- Association structures: As discussed in Section 1.2.3.1, one of the most useful features of the dependence ratio approach is that, depending on the application, an association structure can often be chosen that represents the underlying mechanism that is generating the responses within a cluster.

- Coding of the response: Unlike the odds ratio, the dependence ratio is influenced by the coding of the response. In other words, if the baseline category of the response is changed, a different set of association parameters is obtained. This issue is due to the fact that the dependence ratio measures event specific association. Some may see this as a disadvantage, but often there is specific interest in coding the response in a specific way. For example, if the response is binary with 1 representing disease and 0 representing no disease, then the researcher will often be more interested in focusing on the probability of disease and therefore having zero as the baseline. In these cases, the dependence ratio measures solely the association of interest.

- Orthogonality: In contrast to the odds ratio approach, the regression and association parameters are not orthogonal in the dependence ratio approach (Ekholm, 2003). Therefore, calculating the correlations of the parameter estimates is recommended (Ekholm et al. 1995). Ekholm (2003) states that high correlations were rarely found in previous analyses.

- Negative profile probabilities: Rather than using the logits of probabilities, the dependence ratio approach uses the profile probabilities (1.6) to combine the regression and association models. Consequently, the dependence ratio approach can produce negative fitted profile probabilities for the unobserved profiles. Positivity constraints are imposed on the observed profiles in the maximum likelihood estimation (discussed in Section 1.3). However, Ekholm (2003) discusses this issue and concludes that it should not be treated as a drawback of the dependence ratio approach since negative profile probabilities act as a way of detecting if the model is incorrectly specified. In other words, they provide a tool for model validation.

- Range: The dependence ratio is constrained by the marginal probabilities (Ekholm, 2003). For example, for a two-way dependence ratio of a binary response, denoted τ, the range is given by:

$$\max\left\{0, \frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1\mu_2}\right\} \leq \tau \leq \min\left\{\frac{1}{\mu_1}, \frac{1}{\mu_2}\right\}. \tag{1.9}$$

This shows further the similarity between the dependence ratio and the relative risk since both are based on the ratio of probabilities and the range of the relative risk is also constrained by the marginal probabilities. In contrast, the odds ratio is not constrained by the marginal probabilities and it has a range from zero to infinity. The range of the dependence ratio has received some criticism due to its varying upper bound. However, the fact the dependence ratio has a finite upper bound is also arguably beneficial over the odds ratio. In addition, research in computer science found the range of the dependence ratio (or the lift/interest statistic as it is known in this area) only to be a problem for small counts (see Section 1.4.6 for more details). As shown in section 4.3.4 of this thesis, when the marginal probabilities are less than approximately 0.5, they are reasonably variation independent of the dependence ratio. Ekholm (2003) shows this for the marginal homogeneity case. In other words, this point advocates modelling the rarer of the response values using the dependence ratio approach. See Section 4.3.4 for more details.

At this point, it is worth discussing the marginal probability dependent Frechet bounds (Frechet, 1951) that impose constraints on the dependence ratio. In fact, for any bivariate distribution function (F) of two discrete random variables (Y_1, Y_2) with corresponding marginal distribution functions F_1 and F_2:

$$\max\left\{0, F_1(y_1) + F_2(y_2) - 1\right\} \leq F(y_1, y_2) \leq \min\left\{F_1(y_1), F_2(y_2)\right\}. \tag{1.10}$$

From (1.10), (1.9) can be obtained by noting the following:

$$F_1(0) = 1 - \mu_1, F_2(1) = 1 - \mu_2, F(0,0) = 1 - \mu_1 - \mu_2 + \mu_{12}.$$

- Conditional probability interpretation: If there is a natural ordering to the subunits within a cluster such as over time, then there is an alternative expression for the dependence ratio in terms of conditional probability (Ekholm, 2003), which is given by (for the bivariate binary case):

$$\tau_{12} = \frac{pr(Y_2 = 1 | Y_1 = 1)}{pr(Y_2 = 1)}. \tag{1.11}$$

In this case, the dependence ratio has the interpretation that given $Y_1 = 1$, how many times more likely is it that $Y_2 = 1$ compared to the corresponding marginal probability. For a more detailed discussion of the dependence ratio in relation to the odds ratio, see Ekholm (2003).

1.2.5 The dependence ratio for case-control studies

Case-control studies are common in health-related applications (Agresti, 2002) for investigating whether exposure to a potential risk factor is associated with a particular disease. In contrast to cohort studies they use past data and thus a retrospective design. Individuals in the population who develop the disease are defined as cases. Each case is typically matched with an individual who did not develop the disease (control) based on characteristics such as age and gender. The cases and controls are denoted as D and $\bar{D}$ respectively with E and $\bar{E}$ denoting the presence and absence of exposure to a risk factor. Consider the following hypothetical example of a case-control study from Wacholder (1996).

Table 1.3: Simple hypothetical example of a case-control study: Data extracted from Table 1 of Wacholder, S (1996).

	E	$\overline{E}$	total
D	10	10	20
$\overline{D}$	5	15	20
total	15	25	40

Suppose the population size is 100 000. In addition, assume that 20 of these individuals develop the disease (cases) and that 20 controls are randomly sampled from all those who do not develop the disease in the population. Furthermore, 50% of the cases and 25% of the controls are found to be exposed, as shown in Table 1.3.

A disadvantage of case-control studies such as those displayed in Table 1.3 is that it is not possible to estimate the probability of disease given exposure status since the marginal distributions of disease status are fixed by design. It is only possible to estimate the probability of exposure given disease status which is not typically of interest to the researcher. Consequently, measures of association that are based on ratios of probabilities such as the relative risk cannot be calculated in the direction of main interest. Similarly, the dependence ratio cannot be calculated for the data in Table 1.3 due to the fact that the marginal distributions of disease status are fixed by design. The odds ratio has the advantage that it can be calculated for case-control studies since it can be determined by conditional distributions in either direction (Agresti, 2002).

However, Wacholder (1996) showed that a case-control study can also be viewed as having data missing by design and that there are advantages to be gained in doing so. For the data in Table 1.3, the population size is known to be 100 000, therefore we know that there are 99 960 individuals with unknown exposure status, all of which are controls. If the missing exposure data is deemed to be missing at random, one would expect 24 990 to be exposed (25% of the 99960) and 74 970 to be not exposed. In order

for the missing at random assumption to hold, the controls need to be randomly drawn. Table 1.4 shows the disease by exposure table at the population level.

Table 1.4: Hypothetical example of case-control study accounting for all 100000 persons in the population, when missing exposure information is inferred: Data extracted from Table 1 of Wacholder, S (1996).

	E	$\overline{E}$	total
D	10	10	20
$\overline{D}$	24 995	74 985	99 980
total	25 005	74 995	100 000

Thus the marginal distributions for disease status are no longer fixed by design and the joint distribution can be estimated. In other words, the dependence ratio can be estimated for case-control studies if the total population who are at risk of the disease is known and the missing at random assumption is met. The dependence ratio estimate is therefore given by:

$$\tau = \frac{\frac{10}{100\,000}}{(\frac{20}{100\,000}) \times (\frac{25\,005}{100\,000})} = 1.9996 \tag{1.12}$$

The odds ratio estimate is given by 3. The fact the dependence ratio is less than the odds ratio is due to the fact that the dependence ratio is always bounded towards zero association (independence) by the odds ratio except for the case of no association (Ekholm, 2003).

1.3 R package drm

A key feature of the dependence ratio approach is the freely available R package drm that was written by Jukka Jokinen (2007). The package can be obtained from:

www.helsinki.fi/∼jtjokine/drm/

In addition, the package can also be obtained from within R using install.packages ("drm"). The full list of features within the drm package can be viewed by clicking on the help pages tab, which is displayed on the above web page. Most notably, this includes details on the key functions (drm and depratio) and information on relevant datasets.

The drm function is the main function since it provides the framework for modelling multivariate categorical data using dependence ratios. This function provides estimates for the regression and association parameters (dependence ratios), fit statistics such

as Akaike's Information Criterion (AIC), and the correlation matrix of the parameter estimates, amongst others. It can cope with binary responses as well as extensions to multicategory responses (ordinal or nominal) through the specification of the relevant link function. However, one of the most convenient aspects of the function is the ability to specify the different association structures from Section 1.2.3.1. If a satisfactory fit cannot be achieved with any of the structures available in drm, alternative constraints can be imposed on the dependence ratios, typically based on the observed dependence ratios or theoretical considerations. If negative profile probabilities are encountered with a particular model, drm warns the user that the model will need to be respecified.

The depratio function is a useful additional function that calculates the observed adjacent w-way dependence ratios for the data in question, where w is specified by the user in advance. Although the depratio function does not allow the calculation of the observed non-adjacent dependence ratios, these can be calculated with some additional R code, as demonstrated in paper 2 of this thesis. Jokinen (2006: PhD Thesis) discusses in detail the way in which the maximum likelihood estimation process works in drm. The key points are now discussed.

As discussed in Section 1.2.3, the profile probability can be expressed in closed form for the dependence ratio approach (in terms of the first-order moments and the dependence ratios of all orders). However, rather than modelling the complete joint distribution, constraints will often need to be applied to the regression and association parameters, especially when there are a large number of subunits within a cluster. Consequently, in these cases the score functions for the dependence ratio approach are nonlinear and need to be solved using iterative methods. Despite this, the fact the profile probability can be expressed in closed form gives the dependence ratio approach a major advantage over the odds ratio approach with regards to the number of iterations required. The profile probabilities cannot (typically) be expressed in closed form in terms of the first-order moments and odds ratios. As a consequence, in order to determine the maximum likelihood estimates for the odds ratio approach, two steps are needed within each round of the iterations for updating the parameter estimates. However, the dependence ratio approach requires only one step in the iterations (within each round) for updating the parameter values.

Jokinen (2006: PhD Thesis) states that the MAREG software package (Kastner et al. 1997), which uses the mixed odds ratio parameterisation and R functions based on the local odds ratio parameterisation (Lang, 2004) are the only other software available for population-averaged modelling of multivariate categorical data, using maximum likelihood estimation. The author is not aware of any other relevant packages being released in recent years. Jokinen (2006: PhD Thesis) does however emphasise that drm is believed to be the only available software for taking into account the handling of missing values (using maximum likelihood). Although this thesis only encounters datasets that have no missing values, the options drm gives for this are now briefly discussed. Within

the drm function, the 'missing' option allows the user to specify a particular structure for the dropout mechanism. The drm package allows for MCAR (missing completely at random), MNAR (missing not at random) and MAR (missing at random).

1.4 The history of the dependence ratio

Although the dependence ratio was not proposed for the purpose of modelling multivariate categorical data until Ekholm et al. (1995), previous research prior to this (and since) has used the dependence ratio formulation for other purposes and in different application areas, using different names for the dependence ratio. For example, the mobility ratio in sociology and the lift statistic in computer science. The dependence ratio concept has proved useful in some application areas whereas others have acknowledged the advantages, but sought different measures. The following sub-sections now describe (chronologically by area) the methods used in the different application areas.

1.4.1 Genetics

1.4.1.1 Coefficient of coincidence

The earliest known use of the dependence ratio formulation was in a genetic context by Muller (1916). Muller proposed the coefficient of coincidence (COC) as a measure of interference with regards to the crossovers that occur at chromosomes during the cell division process of meiosis. Typically, the coefficient of coincidence is based upon the recombination (crossover) rates between three genes. Consider three genes (A, B and C), where B is located between A and C. The two events of interest are whether crossovers occur between A and B (AB), and whether crossovers occur between B and C (BC). The coefficient of coincidence is defined as the joint probability of a double crossover (crossovers at both AB and BC) divided by the joint probability under statistical independence, so it is clearly the same as the dependence ratio formulation (1.4). The coefficient of coincidence is therefore given by:

$$COC = \frac{Pr(AB, BC)}{Pr(AB)Pr(BC)} = \tau_{AB,BC.} \tag{1.13}$$

A COC of 1 represents independence between the crossovers and in this case, the interference is 0 since:

$$\text{Interference} = 1 - COC. \tag{1.14}$$

In other words, the interference is a measure of how a crossover between A and B interferes with a crossover between B and C (and vice versa).

1.4.1.2 Sex ratio

Another use of the dependence ratio formulation in a genetic context was in Kullback (1971). A dataset of 36,536 families was considered with the sex ratio of the first four births analysed for independence and homogeneity. A four-way contingency table was considered and the four marginals analysed. The sex ratio is not the same as the dependence ratio since it is simply the ratio of males to females. In order to test hypotheses concerning independence and homogeneity, a number of test statistics were proposed (generally taking into account the dependence between families). Of particular note for the context of this section is the test statistic for independence presented by Kullback (1971), which is given by:

$$2\sum_i\sum_j\sum_k\sum_l n_{ijkl}\log\left[\frac{\frac{n_{ijkl}}{n}}{\frac{n_{i+++}n_{+j++}n_{++k+}n_{+++l}}{n^4}}\right], \tag{1.15}$$

where n_{ijkl} represents the frequency in the $ijkl$-cell of the four-way contingency table, n_{i+++} represents the marginal distribution of the ith birth and n is the grand total.

It is clear that the (observed) dependence ratio forms part of the test statistic in (1.15):

$$\frac{\frac{n_{ijkl}}{n}}{\frac{n_{i+++}n_{+j++}n_{++k+}n_{+++l}}{n^4}} = \hat{\tau}_{ijkl}. \tag{1.16}$$

It is noted that the work of Kullback (1971) concentrated on the marginal distributions and does not consider the associations between families. However, the relative risk recurrence ratio considers the associations (between relatives) and this is discussed in the next section.

1.4.1.3 Relative recurrence risk

In genetics, the relative recurrence risk is defined as the risk of disease in relatives of affected individuals. This led to the development of the relative recurrence risk ratio (Risch, 1990a), λ_R, which is given by the relative recurrence risk divided by the risk of disease in the general population. For the case of two affected relatives, this is given by:

$$\lambda_R = \frac{pr(D_2=1|D_1=1)}{pr(D_2=1)} = \frac{pr(D_1=1,D_2=1)}{pr(D_1=1)pr(D_2=1)} = \tau_{D_1D_2}, \tag{1.17}$$

where D_1 and D_2 represent the disease status of two relatives (1 = affected by the disease, 0 = unaffected).

Equation (1.17) makes use of the conditional probability interpretation that was described in Section 1.2.4. Following the conditional probability interpretation, $\lambda_R > 1$ means that the relative of an affected individual is at greater risk than the general population.

It is clear that (1.17) is the same as the dependence ratio formulation. The sibling recurrence risk ratio (λ_S, which is λ_R for siblings) is often of particular interest. A common use of λ_S is in exclusion mapping. This involves regions of chromosomes being excluded if they give a λ_S that is less than a specified number. For example, Duffy et al. (2001) used 1.5 as the cut off.

Guo (2000) showed that estimates of λ_R are inflated if relevant environmental factors are ignored. Wallace and Clayton (2003) showed that λ_R can account for these factors. For example, if X_1 and X_2 are the only required environmental factors:

$$\lambda_R = \frac{pr(D_1 = 1, D_2 = 1|X_1, X_2)}{pr(D_1 = 1|X_1)pr(D_2 = 1|X_2)}. \tag{1.18}$$

Wallace and Clayton (2003) focused on leprosy and used a marginal model with logistic margins (to take into account the environmental factors). In terms of dealing with the association between pairs of relatives, they used an extension of Plackett's copula (Plackett, 1965) and this allows for the fitted λ_S values to be obtained (although not directly since Plackett's copula is based upon an odds ratio parameterisation). Copula's were first proposed in Sklar (1959). Consider two random variables (Y_1, Y_2) with corresponding continuous marginal distribution functions $F_1(y_1)$ and $F_2(y_2)$ and joint distribution function $F(y_1, y_2)$. Sklar's theorem states that the joint distribution function F can be expressed as:

$$F(y_1, y_2) = C(F_1(y_1), F_2(y_2)), \tag{1.19}$$

where C is a distribution function with uniform marginals on $[0, 1]$. It is referred to as a copula function that describes the dependence between Y_1 and Y_2.

Copulas are of particular use when the researcher has knowledge of the form of the marginal distributions but not the joint distribution (Wallace and Clayton, 2003). They allow for joint distributions to be developed with given margins. A wide variety of copula functions are possible. Plackett's copula contains a single odds ratio parameter (χ_{pl}) for measuring the association between Y_1 and Y_2 and this satisfies the following:

$$\chi_{pl} = \frac{F(1 - F_1 - F_2 + F)}{(F_1 - F)(F_2 - F)} \tag{1.20}$$

Wallace and Clayton preferred the range of the odds ratio but show a clear preference for using λ_R for interpretation. This is because the odds ratio has the more difficult interpretation of being the ratio of the odds of disease for someone who has an affected relative (of a given type) to the odds of disease for someone who does not have an affected relative (of the same type).

In contrast to Plackett (1965), Wallace and Clayton (2003) used maximum likelihood estimation. They focused on only two relatives and note that the approach does not easily generalise to three or more relatives. In order to cope with a larger number of relatives, they were split in to all possible pairs with robust standard errors used to account for the dependence. Since the dependence ratio approach discussed in this thesis can cope with larger cluster sizes, it may be useful for modelling three or more relatives in a genetic context. In addition, the dependence ratio approach would allow for the λ_R to be calculated directly.

1.4.2 Social mobility

The earliest known use of the dependence ratio formulation in sociology was in the analysis of social mobility tables in the 1940's. As described in Tyree (1973), three sociologists (Natalie Rogoff, David Glass and Gosta Carlsson) in different countries, working on similar datasets, wished to understand the issue of intergenerational occupational mobility. Typically, interest lies in comparing the occupations of respondents to that of their fathers, using mobility tables. Two main problems were found by the three sociologists. Firstly, the selecting of appropriate occupational categories for the fathers and sons was challenging as they were found to differ drastically in size. Secondly, the marginal distributions of the two generations were found to differ greatly. The aim of the three sociologists was to come up with a measure that made the time periods of the two generations comparable with regards to occupational structure (Rogoff, 1953).

Rogoff (1953), Glass (1954) and Carlsson (1958) all developed the same measure to deal with this problem. Rogoff called it the 'social distance mobility' ratio whereas Glass referred to it as the 'index of association' and Carlsson used simply 'c'. In this section, the term mobility ratio is used, as in Tyree (1973). Let n_{ij} represent the count of a particular cell in the two-way mobility table. The mobility ratio (for the (i, j) cell in a k by k mobility table) is specified as:

$$M_{ij} = \frac{n_{ij}n}{n_{i+}n_{+j}}, \tag{1.21}$$

where n_{i+} and n_{+j} are the marginal counts of row i and column j respectively and n is the grand total of the mobility table. In addition, $i, j = 1, \ldots, k$ with k representing the number of occupational categories of the son and father.

Tyree (1973) described the interpretation of the mobility ratio in terms of the cell counts rather than the cell probabilities. Consequently, the mobility ratio is interpreted as the ratio of the observed frequency in a particular cell to the corresponding frequency that is expected under statistical independence. By denoting $\hat{\mu}_{ij}$ as the cell probability of a particular cell, where $\hat{\mu}_{ij} = \frac{n_{ij}}{n}$, it can be shown that (1.21) is the same as the (observed) dependence ratio formulation:

$$M_{ij} = \frac{n_{ij}n}{n_{i+}n_{+j}} = \frac{n^2\hat{\mu}_{ij}}{n_{i+}n_{+j}} = \frac{\hat{\mu}_{ij}}{\hat{\mu}_{i+}\hat{\mu}_{+j}} = \hat{\tau}_{ij}. \tag{1.22}$$

Consequently, the mobility ratio can be interpreted in the same way as the dependence ratio. In other words, the joint probability in a cell divided by the corresponding probability expected under statistical independence.

The mobility ratio received some criticism (Tyree, 1973). Recall from Section 1.2.4 that the dependence ratio (mobility ratio) has a range that is constrained by the marginal probabilities. In other words, different mobility ratios have different maximums. Consequently, the variable upper limit of the mobility ratio makes comparisons between the mobility ratios more difficult.

A solution to the variable upper limit of the mobility ratio is to use a proportion of max measure such that the mobility ratios are divided by their corresponding maximum ($\frac{M_{ij}}{max(M_{ij})}$), where the maximum is the minimum of the reciprocal of the two marginal probabilities (Tyree, 1973). However, this proportion of max measure does not relate to independence since values of independence depend on the marginal probabilities. Tyree (1973) proposed the Yule statistic (Yule, 1912) Q to combat this difficulty. For the simple 2 by 2 mobility table:

$$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}}. \tag{1.23}$$

Q solved some of the difficulties with mobility ratios. For example, Q has a range between -1 and 1 with 0 corresponding to independence. However, Q may be a poor measure for comparing two social mobility tables (Tyree, 1973).

Bibby (1975) also gives a very good discussion on the measures of mobility. The disadvantages in using the mobility ratio to compare different mobility ratios are acknowledged and the proportion of max measure ($\frac{M_{ij}}{max(M_{ij})}$) is proposed, as in Tyree (1973). Bibby calls this proportion of max measure the standardised index.

The Durbin statistic (1955) is also mentioned to combat the fact that the standardised index has no fixed value for independence and is given by:

$$\frac{M_{ij} - 1}{max(M_{ij}) - 1} = \frac{\hat{\tau}_{ij} - 1}{max(\hat{\tau}_{ij}) - 1}. \tag{1.24}$$

In other words, (1.24) is equal to the (observed) dependence ratio minus 1 divided by the maximum dependence ratio minus 1. This measure is used in paper 2 of this thesis. The Durbin statistic has the nice property that zero represents independence and 1 represents the maximum. However, the minimum does depend on the marginal probabilities.

1.4.3 Information theory and linguistics

Information theory refers to the measuring of information and was first introduced by Shannon (1948). It has been used in many disciplines including computational linguistics, which is the focus of this discussion. Shannon and Weaver (1949) introduced the concept of mutual information to information theory. Consider two random variables (X and Y). Mutual information is defined as the amount of information that one of the random variables contains about the other and is given by:

$$I(X;Y) = \sum_{x;y} pr(x,y) \log \frac{pr(x,y)}{pr(x)pr(y)}. \tag{1.25}$$

Theil (1970) proposed the uncertainty coefficient (U), which is found by dividing the mutual information by the entropy (Shannon, 1948). Entropy is defined as $-\sum_{y} pr(y) \log[pr(y)]$. The uncertainty coefficient is defined by:

$$U = -\frac{\sum_{x;y} pr(x,y) \log \frac{pr(x,y)}{pr(x)pr(y)}}{\sum_{y} pr(y) \log[pr(y)]}. \tag{1.26}$$

U takes values between 0 and 1, where 0 corresponds to independence between X and Y, and 1 corresponds to no conditional variation (perfect prediction) such that, for each x, $pr(y|x) = 1$, for a given y.

The numerator in (1.26) is in fact the logarithm of the dependence ratio (for given x and y) weighted by the joint probability of x and y. The denominator (entropy) in (1.26) normalises U so that it takes values between 0 and 1.

In computational linguistics, interest often lies in identifying pairs of events such as word collacations that are strongly associated. Collacations are sequences of words that occur together. It is well known that using the simple co-occurrence probability is not a sufficient measure of association since it does not relate to independence. The pointwise mutual information (PMI) is a popular association measure in linguistics. Church and

Hanks (1990) introduced PMI to the area of linguistics (they call it the association ratio). PMI is defined over specific values of the random variables. In contrast, mutual information is defined as the expected value of the PMI evaluated over all possible values of the random variables. PMI can be positive or negative whereas the mutual information can only be positive. PMI is defined as the logarithmic ratio of the joint probability of the two events (typically words) co-occurring to the joint probability under independence (Church and Hanks, 1990). In other words, the PMI is the log of the dependence ratio. Church and Hanks (1990) proposed the measure with a base of 2 whereas other researchers use the natural logarithm (e.g., Bouma, 2009). The PMI between two events x and y is defined as:

$$PMI = \log \frac{pr(x,y)}{pr(x)pr(y)} = \log(\tau_{xy}). \tag{1.27}$$

The PMI has a well known problem that it tends to over estimate the association between low frequency words (Church and Hanks, 1990; Pantel and Lin, 2002). Same as the dependence ratio, PMI has a fixed value for independence (in this case 0) and the upper bound is constrained by the marginal probabilities. Church and Hanks (1990) removed all pairs with counts ≤ 5 to rectify this problem and noted that the PMI produces very credible results once this has been done. In addition, Bouma (2009) stated that the PMI performs relatively well when the low counts are removed. In other words, these results imply that the lack of a fixed upper bound for the PMI is only an issue in cases when there are low counts.

Rather than removing pairs with low counts, it is often preferable to use statistics that are valid for smaller counts. For example, Bouma (2009) proposed a normalised pointwise mutual information. This statistic has a range between -1 and 1 with 0 corresponding to independence and is given by:

$$i(x,y) = \frac{\log \frac{pr(x,y)}{pr(x)pr(y)}}{-\log(pr(x,y))} = \frac{\log(\tau_{xy})}{-\log(pr(x,y))}. \tag{1.28}$$

In order to prove the range of (1.28) first consider the situation of independence. Under independence, τ_{xy} equals 1 and consequently the numerator in (1.28) equals 0. Therefore 0 corresponds to independence. Next, consider the case where a given two words only occur together. In this situation, $\log(pr(x,y)) = \log(pr(x)) = \log(pr(y))$. Equation (1.28) therefore becomes $\frac{-\log(pr(x,y))}{-\log(pr(x,y))} = 1$, in other words proving that the maximum of (1.28) is 1. Finally, consider the case where a given two words never occur together. Under this scenario, (1.28) approaches -1 when the joint probability $pr(x,y)$ approaches 0 and $pr(x), pr(y)$ are fixed (Bouma, 2009), therefore proving that the minimum of (1.28) is -1.

1.4.4 Bayesian use of the dependence ratio

Paper 3 of this thesis considers the dependence ratio association measure in a Bayesian framework. Although the reader is referred to paper 3 for a more detailed discussion, the work of Good (1956) is discussed here due to the fact the dependence ratio formulation was used. Good (1956) used a Bayesian approach to smooth cell counts in large sparse contingency tables and he referred to dependence ratios as association factors. Log-normal and gamma priors were used in order to estimate association factors for each cell in the table. In contrast to the work of paper 3 and traditional Bayesian approaches, Good (1956) used empirical Bayes approaches whereby the prior parameters are estimated from the data. Good (1956) stated that this method may well be natural to those familiar with PMI and he also discusses the fact that the association factor was called the coefficient of dependence by Keynes and Johnson (Keynes, 1921). Du Mouchel and Pregibon (2001) also proposed an empirical Bayes approach for dealing with small samples, which is discussed in Section 1.4.6.

1.4.5 Correspondence analysis

Goodman (1996) introduced a general method for the analysis of nonindependence between row and column categories in two-way contingency tables. For an I (row) by J (column) contingency table, let p_{ij} represent the probability of an observation being in cell ij. Goodman's method used the Pearson ratios (Goodman, 1996), which are defined as:

$$\psi_{ij} = \frac{p_{ij}}{p_{i+}p_{+j}} = \frac{\mu_{ij}}{\mu_{i+}\mu_{+j}} = \tau_{ij}, \tag{1.29}$$

where $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

It is clear that the Pearson ratios are the same as the dependence ratios. However, the Pearson ratios main use is in correspondence analysis, so consequently this discussion focuses on this application.

Correspondence analysis is a multivariate tool used for graphically displaying the associations between two or more categorical variables in a contingency table. It provides a more intuitive view of the associations between categorical variables than considering numerical summaries alone (Beh, 2008). Correspondence analysis can be applied to two-way or multi-way contingency tables. Two-way refers to the fact that there are two categorical variables and this is also known as simple or classical correspondence analysis. Both methods traditionally assume that the categorical variables have no order to them (nominal) but relevant extensions are available for ordinal variables or a

combination of ordinal and nominal variables (Beh, 2008). This discussion focuses on the simple two-way (nominal) case.

The chi-squared statistic (X^2) is also commonly used for assessing the association between two categorical variables in a contingency table. It can be expressed in terms of the Pearson ratios (Beh, 2004):

$$n\sum_{i=1}^{I}\sum_{j=1}^{J} p_{i+}p_{+j}(\psi_{ij}-1)^2. \tag{1.30}$$

However, in contrast to correspondence analysis, X^2 does not give information on similarities and differences between the categories (Beh, 2004). In addition, the chi-squared statistic (X^2) increases as n increases, consequently affecting tests of association. However, correspondence analysis uses $\frac{X^2}{n}$ (the total inertia) to deal with this issue. In correspondence analysis, this total inertia is decomposed using (typically) the singular value decomposition. Correspondence analysis is achieved by applying the singular value decomposition to the Pearson ratios, such that:

$$\psi_{ij} = \frac{p_{ij}}{p_{i+}p_{+j}} = 1 + \sum_{m=1}^{K} \lambda_m a_{im} b_{jm}, \tag{1.31}$$

where $m = 1, \ldots, K; K = \min(I, J) - 1, a_{im}$ is a singular vector associated with the ith row category, b_{jm} is a singular vector associated with the jth column category and λ_m represent the mth singular value of the Pearson ratio.

1.4.6 Computer science-data mining

Data mining is a branch of computer science that determines patterns in large datasets using a computational process. It is used in artificial intelligence and database systems (amongst others). Agrawal, Imielinski and Swami (1993) introduced the idea of interestingness measures for determining associations in data mining. In contrast to other application areas, data mining uses the term interestingness measures as opposed to association measures.

In data mining, one often wants to determine items that co-occur frequently. Typically this concentrates on pairs of items and this will be the focus of this discussion. For example, in supermarket basket data, one is often interested in determining items that frequently occur together.

Agrawal, Imielinski and Swami (1993) proposed the confidence and support interestingness measures for associations in data mining. Associations were considered interesting if the confidence and support measures exceeded some pre specified cut-off. The support

measure is defined as the proportion of observations in the dataset that contain both of the pair of items under consideration. In other words, the support is equal to the joint probability in the dependence ratio formulation (the numerator). For two items X and Y, the support is given by:

$$Support = pr(x, y). \tag{1.32}$$

In contrast, the confidence statistic is the conditional probability of one item given the other. For two items X and Y:

$$Confidence = \frac{pr(x, y)}{pr(x)}. \tag{1.33}$$

However, the support and confidence measures received criticism for their use in determining interesting associations in the dataset (Aggarwal and Yu, 1998). For example, researchers (Aggarwal and Yu, 1998; Brin, Silverstein and Motwani, 1997) discussed the fact that the confidence statistic does not take into account the marginal probability of Y. In addition, neither the confidence nor the support relates to independence.

This led to the development of the lift (interest) measure by Brin et al. (1997). The lift measure is the same as the dependence ratio since it is defined as the joint probability of both items divided by the joint probability of both items under independence. In other words, in contrast to the confidence and support measures, the lift measure is relative to independence. It is defined as:

$$Lift = \frac{pr(x, y)}{pr(x)pr(y)} = \frac{\text{Confidence}}{pr(y)} = \frac{\text{Support}}{pr(x)pr(y)} = \tau_{xy}. \tag{1.34}$$

DuMouchel (1999) and DuMouchel, Pregibon (2001) noted some limitations with the lift statistic. Most notably, when the support is low, the lift measure becomes unstable. In other words, there appears to be evidence that the variable upper limit of the dependence ratio formulation is only a problem when the counts are small. This coincides with the conclusions for the pointwise mutual information that is used in linguistics (discussed in Section 1.4.3). However, Du Mouchel and Pregibon (2001) proposed an empirical Bayes approach which smoothes the lift statistic (denoted R in their work) and consequently gives more reliable estimates when the support is low. The nature of empirical Bayes approaches means that the prior parameters are estimated from the data. The common prior distribution that is assumed for all R values enables the values of R with smaller support to borrow strength from those with stronger support.

The empirical Bayes approach is conducted separately for different sized item sets (pairwise associations, three-way associations and so on). The focus on this discussion is on

pairwise associations. Let n denote the observed count for a particular set of two items and e denote the corresponding count under independence. The observed counts n are assumed to follow a Poisson distribution with mean $\mu = \lambda e$. Interest lies is determining the posterior distribution of $\lambda = \frac{\mu}{e}$, which is the expected value of R. The λ values are assumed to follow a continuous prior (parametric) density function given by $p(\lambda|\boldsymbol{\theta})$, for fixed $\boldsymbol{\theta}$. This parameter vector $\boldsymbol{\theta}$ is estimated from the data such that every λ is assumed to have a prior distribution given by $p(\lambda|\hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate. The posterior distribution of λ is computed via Bayes rule before quantities such as the posterior mean of λ (denoted δ) can be obtained. To conclude, δ is much more reliable than R when n and e are small but close to R when n and e are sufficiently large (Du Mouchel and Pregibon, 2001).

1.4.7 Use of the dependence ratio approach

The discussion so far has concentrated on uses of the dependence ratio in other contexts to that of the approach proposed by Ekholm et al. (1995). However, the dependence ratio approach introduced by Ekholm et al. (1995) has been used for modelling multivariate categorical data by researchers that were not the original founders of the approach. Most notably, Anatolyev (2008) used the dependence ratio approach to model the dependence across stock markets. The response of interest for each market was binary (successful if the market return went up or not successful if the market return went down). A key difference between this paper and previous modelling approaches using the dependence ratio approach (such as Ekholm et al. 1995) was that Anatolyev (2008) had time periods as units rather than the typical individual. Consequently, rather than using the traditional logistic model for the first-order moments, Anatolyev (2008) proposed a multivariate generalised autoregressive logit model to account for the time ordering and the fact that the units are not independent. Separate analyses were conducted for each of the European, Chinese and Baltic markets. The development of the R package drm (described in Section 1.3) by Jukka Jokinen (2007) should encourage more use of the dependence ratio approach in relevant application areas (such as genetics) since the lack of any available package prior to 2007 made implementation more difficult for those not familiar with the approach.

1.4.8 Conclusion

The dependence ratio formulation has been used in a number of application areas, some of which were proposed before the dependence ratio approach for modelling multivariate categorical data was proposed (Ekholm et al. 1995). The mobility ratio used in social mobility tables (two-way) was shown to be the same as the dependence ratio formulation. This measure received criticism for its use in comparing different mobility ratios since

the upper limit of the measure is constrained by the marginal probabilities. However, researchers in computer science and linguistics found this to only be an issue for small counts. Linguistics used the pointwise mutual information (although this is the log of the dependence ratio) and computer science used the lift measure. In social mobility, relevant measures were proposed to deal with the varying upper limit such as the proportion of max (Tyree, 1973) and Yule's Q (Yule, 1912). Du Mouchel and Pregibon (2001) proposed an empirical Bayes approach which smoothed the values of the dependence ratio in small samples.

With the exception of the empirical Bayes approaches by Good (1956) and Du Mouchel and Pregibon (2001), the dependence ratio formulation has only been considered in an exploratory sense in the different application areas mentioned. When a modelling approach was adopted in genetics (e.g. Wallace and Clayton, 2003), the parameterisation used was based on the odds ratio. Consequently, the association measure of interest ($\lambda_R =$ dependence ratio) cannot be obtained directly. Wallace and Clayton (2003) noted that their odds ratio based approach was limited to two relatives (two subunits in a cluster) due to difficulties with maximum likelihood estimation. In genetics, it would therefore be of benefit to use the dependence ratio approach by Ekholm at al. (1995) and described in Section 1.2.3 when interest lies in modelling three or more relatives since it copes with more than two relatives and λ_R can be obtained directly.

To conclude, the dependence ratio formulation has been used in a number of application areas, some of which were prior to the development of the dependence ratio approach (Ekholm et al. 1995) for modelling multivariate categorical data. These applications used alternative names to the dependence ratio and used the dependence ratio largely in an exploratory sense with modelling approaches using different methods rather than utilising the advantages of the dependence ratio approach.

1.5 Outline of the thesis

As previously discussed, this thesis focuses on the dependence ratio as an association measure for multivariate categorical data in population-averaged models, as opposed to using the traditional odds ratio. Instead of treating the associations as a nuisance and concentrating on the marginal regression (such as the GEE approach), this approach aims to use specific association structures that adequately model the associations within a cluster. This thesis extends the dependence ratio to relevant applications where the dependence ratio has not been applied. This will be undertaken in the form of three papers, which are each outlined below. It should be noted that there is some repetition between the introduction and the papers since it is assumed that some readers may only read certain parts of the thesis and not others.

Paper 1: The dependence ratio association measure for square tables

Paper 1 focuses on how the dependence ratio can be applied to square tables as well as relevant extensions. For a categorical response, whether that be binary, ordinal or nominal, square tables are two-way contingency tables that have the same row and column categories. In other words, square tables refer to multivariate categorical data that is observed only twice. The traditional models for square tables, such as marginal homogeneity and symmetry, are replicated directly in terms of dependence ratios, where possible.

Three datasets are analysed using the dependence ratio approach. Firstly, a straightforward bivariate binary case (2 by 2 table) that assesses bilirubin abnormalities following drug treatment by comparing an active treatment to a placebo treatment (Walker, 2002). In addition, the two well-known eye grade datasets are analysed in detail. These two datasets relate to males (Stuart, 1953) and females (Stuart, 1955). The datasets concern males and females having both their left and right eye graded (on a four level categorical scale). For each dataset, if a satisfactory fit cannot be achieved with one of the traditional models for square tables, relevant constraints are imposed on the dependence ratios.

The second part of the paper focuses on extending the dependence ratios use in square tables to matched sets categorical data. In other words, using the dependence ratio as a measure of association when there are more than two observations within a cluster (multi-way contingency tables). A dataset on the rater agreement between seven pathologists is analysed (Landis and Koch, 1977). The paper uses the R package drm developed by Jokinen (2007) as well as additional R code where necessary.

Paper 2: Applications of the dependence ratio to psoriatic arthritis

In paper 2, the dependence ratio is used as an association measure for analysing the presence of damaged joints in the hands of patients who are affected with psoriatic arthritis. A specific dataset from a study conducted in Toronto, Canada is used for analysis. The patients were followed longitudinally over time. The dataset contains responses for each of the 28 hand joints in a patient (at each clinic visit and excluding the wrist), where the response is a binary indicator of whether the patient's joint is clinically damaged or not (see paper 2 for a definition of clinical damage). However, this analysis focuses on only the last clinic visit since damage is an irreversible process. The dependence ratio approach puts equal emphasis on the marginal regression and the associations within patients. The aim of the marginal regression is to determine if clinical damage differs with regards to explanatory variables such as the age that the patient was diagnosed with psoriatic arthritis and the gender of the patient. A key aim of studying the associations is to determine whether there is stronger association of clinical damage between or within hands, or between certain groups of joints. The drm package is used for analysis along with some additional R code.

Paper 3: Prior and posterior distributions for dependence ratios

Bayesian methods have been developed for multivariate categorical data. However, it is believed that they have not been considered using the dependence ratio association measure and traditional Bayesian approaches. As discussed in Section 1.4.4 and paper 3 of this thesis, Good (1956) used an empirical Bayes approach in order to estimate association factors (dependence ratios). This paper develops a Bayesian approach using dependence ratios and traditional Bayesian approaches whereby prior distributions are put directly on the parameters of interest (first-order moments and dependence ratios) rather than being estimated from the data. The simplest bivariate binary response (2×2 contingency table) is considered in detail. The software R and BUGS are used to fit the models.

1.6 Discussion

This thesis focuses on the use of the dependence ratio for multivariate categorical data with population-averaged coefficients, using maximum likelihood estimation (papers 1 and 2) and Bayesian techniques (paper 3). The work focuses on applying the dependence ratio to applications that have not previously been considered. The dependence ratio approach has some notable advantages over other approaches. A key benefit is that it can cope computationally with larger cluster sizes whereas the odds ratio cannot typically cope with cluster sizes larger than approximately five (Lesaffre et al. 2000). Although the popular quasi-likelihood GEE approach can deal with larger cluster sizes, it focuses on the marginal regression with associations being treated as a nuisance. Many statisticians argue that the whole response profile should be taken into account by determining the mechanisms that are generating the associations within a cluster (Lindsey and Lambert, 1998).

The GEE approach relies on large samples. However, datasets may not have this property. In addition, if there are missing values, the GEE approach may not be adequate for taking into account the dropout mechanism since it relies on the MCAR (missing completely at random) assumption, which may not hold (Jokinen, 2006: PhD Thesis). Consequently, the GEE approach is therefore often only appropriate for datasets with large samples and limited or no missing values. The odds ratio approach is often only appropriate for datasets with small cluster sizes as well as limited or no missing values. In contrast, the dependence ratio approach can cope with larger cluster sizes, and cases where there are missing values. For dealing with missing values, drm allows for MCAR, MAR (missing at random) and MNAR (missing not at random), which permits a more thorough analysis of the dropout mechanism. The reader is referred to Jokinen (2006: PhD Thesis) for a more detailed discussion of these features. Since datasets will often have one or more of: small sample sizes, large cluster sizes and missing values, the dependence ratio approach with the use of drm is often an appropriate method to use.

There are some issues with the dependence ratio approach that should be considered before the approach can be implemented, none of which should be considered serious drawbacks. Firstly, negative profile probabilities can be produced for the unobserved profiles. Instead of treating this as a drawback, the issue should be treated as a method of model validation (Ekholm, 2003), particularly given the fact that drm warns the user that the model is incorrectly specified in these cases. However, it should be noted that negative profile probabilities typically only occur in large cluster sizes. In addition, even in these cases, the GEE approach is the only other method that can typically be used since the odds ratio and other solutions for maximum likelihood run into computational difficulties. Secondly, the range of the dependence ratio has received some criticism due to the fact it is constrained by the marginal probabilities. Ekholm (2003) argued against this being a disadvantage. Interestingly, researchers in linguistics and computer science

(see Section 1.4) found the lack of a fixed upper limit to only be a problem when the counts are small. However, when comparing the magnitude of dependence ratios, it can be sensible to standardise each measure so that they are each a proportion of their max. In addition, further standardisation is desirable to ensure that there is a sensible value for independence and suitable lower and upper bounds. This procedure is demonstrated in paper 2.

Jokinen (2006: PhD Thesis) believed that the MAREG package (Kastner et al. 1997) based on the mixed odds ratio parameterisation and a series of other R functions (Lang, 2004) based on the local odds ratio are the only additional software to drm that are available for modelling multivariate categorical data using maximum likelihood estimation (population-averaged coefficients). The author is not aware of any more recent packages that been released to alter this claim. In addition, drm is believed to be the only software available for modelling the dropout mechanism using maximum likelihood estimation.

To conclude, the three papers in this thesis focus on applications of the dependence ratio to both frequentist (papers 1 and 2) and Bayesian (paper 3) situations. Paper 1 of this thesis focuses on specifying the traditional models for square tables (and extensions) in terms of dependence ratios. Paper 2 uses the dependence ratio approach to analyse a dataset on psoriatic arthritis. Finally, paper 3 demonstrates a Bayesian approach using dependence ratios, which is not thought to have been considered before.

Chapter 2

Paper 1: The dependence ratio association measure for square tables

2.1 Introduction

Multivariate categorical data are commonly found within the Social Sciences, whether that be in a longitudinal sense over time or in a general clustered sense whereby responses are clustered without necessarily having a time element. The former could involve a series of measurements for each respondent over time, whereas the latter may involve family members giving a series of responses such as in social mobility studies. The correlated nature of the data means that applying traditional regression methods are often inadequate since they assume independence. However, by using the response profile of each unit (cluster) as the response variable, multivariate techniques can be used to overcome this problem.

This paper focuses on the analysis of population-averaged models as opposed to subject-specific models. In other words, the regression effects are averaged over the entire population rather than estimating subject-specific effects. In terms of taking into account the associations within a cluster (using maximum likelihood estimation), population-averaged models based on the traditional odds ratio parameterisation suffer from the disadvantage of not being able to adequately cope with cluster sizes that are larger than approximately five (Lesaffre et al. 2000). Although the quasi-likelihood generalised estimating equations (GEE) approach (Liang and Zeger, 1986) helped to solve some of the problems encountered with a odds ratio parameterisation, its emphasis is purely on the marginal regression since it treats the associations within a cluster as a nuisance. The term marginal model is used to refer to cases where only the marginal effects are of interest. However, the present paper gives equal emphasis to both the marginal regression and the associations. Many statisticians argue that equal emphasis should be given to the marginal regression and the associations such as Lindsey and Lambert (1998) who argued that determining the mechanisms that generate the associations within a cluster is vital to obtain a complete analysis of the data. Consequently, alternatives to the odds ratio parameterisation for population-averaged, likelihood-based methods were sought. Ekholm, Smith and McDonald (1995) proposed the dependence ratio as a measure of

association. For the simplest bivariate binary case, $\boldsymbol{Y} = (Y_1, Y_2)$, there are four possible response profiles: (1,1), (0,0), (1,0) and (0,1) with 1 regarded as the response of interest or success and 0 regarded as the baseline or failure. For this case, the dependence ratio is defined as:

$$\tau_{12} = \frac{\mu_{12}}{\mu_1 \mu_2} = \frac{pr(Y_1 = 1, Y_2 = 1)}{pr(Y_1 = 1)\, pr(Y_2 = 1)}. \tag{2.1}$$

The dependence ratio for a bivariate binary response is defined as the joint success probability divided by the joint success probability assuming independence. Similar to the odds ratio, a value of 1 indicates independence, whereas values less than 1 and greater than 1 represent negative and positive associations respectively. Section 2.2 discusses the dependence ratio approach in detail including the case of higher-order dependence ratios and when there are more than two categories in the response. For a full discussion of the advantages and disadvantages of the dependence ratio relative to the odds ratio, the reader is referred to Section 1.2.4 of this thesis or Ekholm (2003). Some of the key points are now discussed:

- Large cluster sizes: The dependence ratio approach can cope computationally with large cluster sizes, using maximum likelihood estimation (as discussed in Section 2.2). In contrast, maximum likelihood estimation is typically not feasible for approaches based on a odds ratio parameterisation when clusters are larger than approximately five (Lesaffre et al. 2000).

- Association structures: Dependence ratios allow for convenient association structures to be imposed in order to reduce the number of association parameters (where necessary) to a more manageable number. These are discussed in Section 2.3.

- Negative fitted profile probabilities: The dependence ratio approach combines a marginal regression model and an association model using the profile probability (see Section 2.2). By parameterising in terms of probabilities as opposed to the logits of probabilities, negative fitted profile probabilities can occur. They can only occur for the unobserved profiles since positivity constraints are imposed on the observed profiles in the maximum likelihood estimation. This should not be treated as a serious disadvantage of the dependence ratio approach since it acts as a method of model validation (Ekholm, 2003).

- Orthogonality: The regression and association parameters (dependence ratios) are not orthogonal in the dependence ratio approach. Consequently, the correlations between these parameters need to be assessed. Ekholm (2003) stated that high correlations were rarely found in previous research. However, when using the odds ratio, the parameters are variation independent and orthogonal.

The dependence ratio approach facilitates the combining of a regression and an association model and this is discussed in Section 2.2. In addition, another convenient aspect

of the approach is the R package drm written by Jukka Jokinen (2007). The software conveniently allows the user to use the dependence ratio approach with certain default association structures. If a satisfactory fit cannot be obtained with these structures, alternative constraints can be imposed on the dependence ratios. These constraints are typically based on the observed dependence ratios as well as theoretical justifications.

The dependence ratio approach has been used to analyse datasets in recent years. Ekholm, Smith and McDonald (1995) introduced the dependence ratio with some initial association structures and used it to analyse two datasets with a binary response. The first dataset concerned pulmonary function and the second concerned the presence of wheeze in young children. Ekholm, McDonald and Smith (2000) then derived more relevant association structures and these were then applied to datasets on rheumatoid arthritis, obesity as well as the wheeze dataset from before, all with a binary response. Ekholm, Jokinen, McDonald and Smith (2003) then extended the dependence ratio to ordinal data and analysed a dataset on the use of the drug Fluvoxamine as a drug for the reduction of psychiatric symptoms. Further work has been undertaken on additional datasets by Ekholm and Skinner (1998); Ekholm, Jokinen and Kilpi (2002); Jokinen, McDonald and Smith (2006) and Jokinen (2006).

One application where the dependence ratio has not been extensively applied to is square tables. Square tables refer to cases where there are only two observations within a cluster. Consequently, they can be represented as a two-way contingency table with the same row and column categories. Square tables are often regarded as a separate area to cases where there are more than two observations in a cluster. Agresti (2002) referred to the square table case as matched pairs data, whereas matched sets (multi-way tables) are used to refer to cases where there are more than two observations in a cluster. Agresti (2002, chapter 10) provided a detailed discussion of the models that are particularly applicable to square tables such as marginal homogeneity and symmetry. He then discusses models relevant to matched sets. However, this discussion either treats the associations as a nuisance or uses the odds ratio as the association measure. As a consequence, this paper introduces the dependence ratio measure as an association measure for square tables and aims to enhance the models discussed in Agresti (2002) in terms of dependence ratios. There is also a section in the present paper that extends the models discussed for square tables to matched sets.

Four datasets are analysed with the dependence ratio approach in the present paper. Firstly, a 2 by 2 contingency table (bivariate binary response) example is considered on bilirubin abnormalities following drug treatment (Walker, 2002). Secondly, the well known eye grade datasets are analysed (Stuart, 1953 and Stuart, 1955), where individuals had their right and left eyes graded according to a four point scale, so the dataset can be represented in a 4 by 4 contingency table. Males and females are analysed separately since they follow different association structures. The final dataset focuses on a rater

agreement study with seven repeated responses, and extends methods for square tables to matched sets.

The outline of the paper is as follows. Section 2.2 discusses the dependence ratio approach for both binary and multicategory responses. Section 2.3 then outlines the main association structures that are considered for the dependence ratio approach. Section 2.4 specifically considers square tables and how the dependence ratio can be applied to existing models. Extensions of these models to matched sets data are discussed in Section 2.5. Section 2.6 applies the dependence ratio approaches discussed to the datasets. The discussion Section 2.7 draws appropriate conclusions.

2.2 The dependence ratio approach

2.2.1 Binary responses

The dependence ratio approach combines a marginal regression model with an association model using the profile probabilities π_i, which are defined as:

$$\pi_i = pr(Y_{i1} = y_{i1}, \ldots, Y_{iq} = y_{iq}), \tag{2.2}$$

where Y_{ik} represents the response for unit $i = 1, \ldots, n$ at subunit $k = 1, \ldots, q$.

The profile probability can be expressed in terms of the univariate marginal probabilities (first-order moments) and dependence ratios of all orders. The first-order moments are given by:

$$pr(Y_{ik} = 1 \mid \boldsymbol{x}_{ik}). \tag{2.3}$$

where $\boldsymbol{x}_{ik}$ is a vector of the explanatory variables.

The marginal regression model is obtained by regressing the first-order moments on relevant explanatory variables, using an appropriate link function. For the binary response case, the logistic link function is typically used and the model is defined as:

$$\text{logit}[pr(Y_{ik} = 1)] = \theta + \boldsymbol{\beta}\boldsymbol{x}_{ik}^T, \tag{2.4}$$

where $\boldsymbol{\beta}$ is a vector of the regression coefficients.

The second-order and higher moments inform us about the association structure within a cluster. However, in contrast to the dependence ratio, they do not relate to the baseline of independence. Consequently, for a more interpretable measure, they are replaced by

the product of first-order moments and dependence ratios. For the case of a binary response, the dependence ratios of all orders are defined as:

$$\begin{aligned} \tau_{12} &= \frac{\mu_{i12}}{\mu_{i1}\mu_{i2}} = \frac{pr(Y_{i1}=1, Y_{i2}=1)}{pr(Y_{i1}=1)\,pr(Y_{i2}=1)} \\ &\vdots \\ \tau_{1\ldots q} &= \frac{\mu_{i1\ldots q}}{\mu_{i1}\ldots\mu_{iq}} = \frac{pr(Y_{i1}=1,\ldots,Y_{iq}=1)}{pr(Y_{i1}=1)\ldots pr(Y_{iq}=1)}. \end{aligned} \tag{2.5}$$

It should be noted that the dependence ratio of a given order is assumed to be the same across all units whereas the moments vary between units. Given q observations within a cluster, constraints will often need to be imposed on the dependence ratios to reduce the number of association measures, particularly when q is large. With association parameters defined by the vector $\boldsymbol{\alpha}$, the association model is defined as:

$$\boldsymbol{\tau} = g(\boldsymbol{\alpha}). \tag{2.6}$$

Consequently the regression model (2.4) and the association model (2.6) are combined using the profile probability in (2.2). The profile probability from (2.2) can also be written as:

$$\pi_i = pr(Y_{i1}=y_{i1},\ldots,Y_{iq}=y_{iq}) = h(\theta,\boldsymbol{\beta},\boldsymbol{\alpha};\boldsymbol{x}_{ik}), \tag{2.7}$$

where $h(\theta,\boldsymbol{\beta},\boldsymbol{\alpha};\boldsymbol{x}_{ik})$ is a function of $\theta,\boldsymbol{\beta},\boldsymbol{\alpha}$ and $\boldsymbol{x}_{ik}$.

In other words, π_i can be expressed in closed form in terms of the regression and association parameters. The likelihood function, and consequently the log-likelihood function, can then be specified for maximum likelihood estimation. In contrast, approaches based on a odds ratio parameterisation generally require iterative procedures to specify the joint distribution.

2.2.2 Multicategory responses

In some situations the response will have more than two categories. This could either be nominal data or if there is an ordering to the response, ordinal data. In each case, the procedure required to combine a regression and association model is more complex than for the binary response case. Ekholm et al. (2003) discuss the approach in detail for an ordinal response. This section summarises this approach as well as incorporating the multinomial case.

The f values taken by the response are defined as $a = 1,\ldots,f$ and it is further assumed that $1 < 2 < \ldots < f$ (if the response is ordinal). In addition, $y_{ik} = a$ is defined as

the response for unit $i = 1, \ldots, n$ at subunit $k = 1, \ldots, q$. Binary indicators are used as dummy variables for the f response values and these are defined as:

$$Y_{ik}^{(a)} = 1 \text{ if } Y_{ik} = a, \text{ else } = 0, \text{ for } a = 1, \ldots, f-1, \tag{2.8}$$

where:

$$Y_{ik}^{(f)} = 1 - Y_{ik}^{(f-1)} - \cdots - Y_{ik}^{(1)}. \tag{2.9}$$

As in the binary case, the profile probability is used to combine the relevant marginal regression and association models and this is given by:

$$\pi_i = pr(Y_{i1} = a_1, \ldots, Y_{iq} = a_q). \tag{2.10}$$

Furthermore, the $1 \times (f^q - 1)$ vector of moment parameters from first-order up to qth-order are defined as:

$$\boldsymbol{\mu}_i = (\mu_{i1}^{(1)}, \ldots, \mu_{iq}^{(f-1)}, \mu_{i12}^{(2,2)}, \ldots, \mu_{i1\ldots q}^{(f,\ldots,f)}), \tag{2.11}$$

where $\mu_{ik}^{(a_k)} = pr(Y_{ik} = a_k)$ and $\mu_{i1\ldots k}^{(a_1,\ldots,a_k)} = pr(Y_{i1} = a_1, \ldots, Y_{ik} = a_k)$.

It was shown by Ekholm et al. (2003) that there is a one-to-one mapping from $\boldsymbol{\mu_i}$ to π_i that allows the joint distribution to be expressed in terms of the moment parameters.

The first-order moments are regressed using a relevant link function on appropriate explanatory variables and form the marginal regression model, as in the binary case. For the ordinal case, where there is an ordering to the response, the proportional odds model is defined as:

$$\log[pr(Y_{ik} \leq a)/(1 - pr(Y_{ik} \leq a))] = \theta_a + \boldsymbol{\beta x}_{ik}^T, \ a = 1, \ldots, f-1, \tag{2.12}$$

where θ_a are the intercept terms.

For the nominal case, where there is no ordering to the response, the traditional multinomial extension of logistic regression is defined as:

$$\log[pr(Y_{ik} = a)/pr(Y_{ik} = f)] = \theta_a + \boldsymbol{\beta}_a \boldsymbol{x}_{ik}^T, \ a = 1 \ldots, f-1. \tag{2.13}$$

The second-order and higher moments are treated differently to those of first-order since these inform us about the associations within a cluster. They are replaced by the

product of first-order moments and dependence ratios, as in the binary case. This is because dependence ratios provide a more useful measure of association than the second order or higher moments since they are comparable to independence. The second-order dependence ratios for both the nominal and ordinal cases are defined as (Jokinen et al. 2006):

$$\tau_{kl}^{(a_k,a_l)} = \frac{\mu_{ikl}^{(a_k,a_l)}}{\mu_{ik}^{(a_k)}\mu_{il}^{(a_l)}} = \frac{pr(Y_{ik}=a_k, Y_{il}=a_l)}{pr(Y_{ik}=a_k)\,pr(Y_{il}=a_l)}, \tag{2.14}$$

for $a_k, a_l = 2, \ldots, f$ and $k, l = 1, \ldots, q; k \neq l$.

Higher-order dependence ratios can be expressed in a similar way. Constraints will typically be imposed on the dependence ratios so that the associations within a cluster can be adequately summarised by a relatively small number of association parameters, particularly when q is large. This forms the association model which is defined as in (2.6). The profile probability (2.10) is used to combine the relevant marginal regression model (ordinal or nominal) with the association model (2.6). As for the binary case, (2.10) can be expressed in closed form in terms of the regression and association parameters, which enables maximum likelihood estimation.

The main interest of this paper is in the application of the dependence ratio to square tables. For square tables, which are considered in detail in Section 2.4, we only have first and second-order moments since there are only two observations within a cluster. However, the response of interest may be either binary or multicategory, as described in this section. The higher-order moments are relevant to situations where there are more than two observations within a cluster and these are discussed in Section 2.5.

2.3 Association structures

As discussed in Section 2.2, it is often necessary to impose constraints on the dependence ratios in order to reduce the number of association parameters. This enables the associations within a cluster to be explained adequately from only a small number of association parameters. Ekholm, McDonald and Smith (2000) proposed a number of association structures that achieve this, some of which are now discussed.

Independence (I):

The most straightforward structure to mention is independence since it assumes that all observations within a cluster are independent. It is therefore rarely appropriate to multivariate categorical data. All dependence ratios are equal to one in this case.

Necessary factor (N):

In some studies, there may be a subset of individuals (units) who will always produce the same response. This can be accounted for by using a necessary factor association structure (N). This structure separates the responses into those individuals that have and those that do not have the factor that is necessary for a response to be (typically) greater than the lowest response category. Responses within a cluster are conditionally independent given the necessary factor. If the necessary factor is present for a given individual, $N_i = 1$ otherwise $N_i = 0$, for $i = 1, \ldots, n$. For individuals with $N_i = 0$, they will always give the lowest response category regardless of the individuals corresponding values for the explanatory variables. As a consequence, the explanatory variables are typically regressed conditional on the individual having the necessary factor (Jokinen et al. 2006). There is one association parameter given by $pr(N_i = 1) = v$. Consequently, $1 - v$ represents the proportion of observations that will always give the lowest response category for each subunit. The dependence ratios for the association structure are defined as:

$$\tau^{(a_1,\ldots,a_w)} = \tau^{(w)} = \gamma^{w-1}, \tag{2.15}$$

where $\gamma = \frac{1}{v}; a_1, \ldots, a_w = 2, \ldots, f$ and $w = 2, \ldots, q$. In other words, the w-way dependence ratios, $\tau^{(w)}$, are equal.

Latent binary factor (L):

It may be appropriate for the population to be divided into two groups such that each has different probabilities for the response categories (given the same values for the explanatory variables in the marginal regression model). This can be accounted for by using a latent binary factor association structure. Each individual either does ($L_i = 1$) or does not have ($L_i = 0$) the latent binary factor (L) and responses within a cluster

are conditionally independent given L. The structure is often appropriate if a relevant dichotomous covariate has not been included in the model (Ekholm et al. 2003). The association model has f parameters, where f is the number of response categories, and is defined as:

$$\boldsymbol{\alpha} = (v_2, \kappa^{(2)}, \ldots, \kappa^{(f)}), \tag{2.16}$$

where $v_2 = pr(L_i = 1)$ and $\kappa^{(a)} = pr(Y_{ik} = a \mid L_i = 0)/pr(Y_{ik} = a \mid L_i = 1)$, for $k = 1, \ldots, q$ and $a = 2, \ldots, f$.

In other words, v_2 represents the proportion of observations with the latent binary factor (latent group 1) and $1 - v_2$ represents the proportion without the factor (latent group 0). The $(f-1)$ different κ parameters are equal to the conditional univariate probability for those in latent group 0 divided by the corresponding probability for those in group 1. The dependence ratios are specified as follows (Jokinen et al. 2006):

$$\tau^{(a_1,\ldots,a_w)} = \frac{v_2 + (1 - v_2)\kappa^{(a_1)} \ldots \kappa^{(a_w)}}{(v_2 + (1 - v_2)\kappa^{(a_1)}) \cdots (v_2 + (1 - v_2)\kappa^{(a_w)})}, \tag{2.17}$$

where $a_1, \ldots, a_w = 2, \ldots, f$ and $w = 2, \ldots, q$.

For given response categories, the w-way dependence ratios are equal. However, these will differ for different response categories. The necessary factor and latent binary association structures may also be combined (as discussed in Section 1.2.3.1 of this thesis).

2.4 Square tables

As described in Sections 2.1 and 2.2, the main focus of this paper is on the application of the dependence ratio to square tables and in particular showing that existing models for square tables can be expressed in terms of dependence ratios. Recall that for square tables, we only have two observations in a cluster, whether that be over time or in a general clustered sense and that data of this form is commonly known as matched pairs data. Hence data of this form can be represented in a two-way table with the same row and column categories. There are two main approaches to the analysis of square tables, as described in Agresti (2002, Chapter 10). Firstly, the marginal model approach which focuses on the marginal distributions. Secondly, the joint distribution can be modelled directly using loglinear models. These approaches are now discussed in Sections 2.4.1 and 2.4.2 respectively before the dependence ratio approach to square tables is considered in Section 2.4.3.

2.4.1 Marginal models for square tables

Consider an $f \times f$ square contingency table with cell counts n_{ab} and corresponding cell probabilities (joint distribution) $\pi_{ab} = pr(Y_1 = a, Y_2 = b)$, where f is the number of categories in the response. Note that the subscripts now refer to the profile and not the ith individual as in π_i from Section 2.2.

In this subsection, the focus is on models for square tables that focus on the marginal distributions rather than the complete joint distribution. They are usually considered when the associations are of secondary interest. One approach is to assume independence between the marginal distributions (Y_1, Y_2) but this is usually not a viable assumption for multivariate categorical data, as discussed in Section 2.1. A preferred approach is to take into account the associations by treating the marginal distributions as dependent. Details of the maximum likelihood estimation for this are discussed at the end of Section 2.4.1 and in Agresti (2002, chapter 11.2.5).

In contrast to subject-specific approaches, the effects in marginal models are population-averaged. One marginal model that is often considered for square tables is marginal homogeneity. For square tables, marginal homogeneity refers to the equality of the marginal distributions for Y_1 and Y_2. Depending on the data in question, the researcher may be interested in testing for marginal homogeneity. For example, consider a social mobility study where each cluster refers to a family with two observations on social class status, one for the father and one for the son. In this case, it is reasonable to test for marginal homogeneity in order to determine if the distribution changes from one generation to another.

2.4.1.1 Binary responses

Let $\boldsymbol{Y} = (Y_1, Y_2)$ represent the binary responses. The following marginal model is defined for square tables:

$$\text{logit}[pr(Y_k = 1)] = \theta + \beta x_k, \tag{2.18}$$

where 1 denotes the category of interest and 0 denotes the baseline; $k = 1, 2$; $x_1 = 0$ and $x_2 = 1$. The parameter β is interpreted as the log odds of a response (=1) for Y_2 relative to Y_1. Marginal homogeneity is equivalent to:

$$pr(Y_1 = 1) = pr(Y_2 = 1). \tag{2.19}$$

Consequently, marginal homogeneity corresponds to the case where $\beta = 0$. Model (2.18) is saturated for the marginal probabilities since the two marginal probabilities are specified by two parameters (Agresti, 2002). However, it is not saturated for the

square table since it does not specify the complete joint distribution, as there is no parameter for the associations. Table 2.1 gives the cell counts for a 2 by 2 square table (bivariate binary response), where:

Table 2.1: Cell counts for a bivariate binary response

	Y_2		
Y_1	1	0	Total
1	n_{11}	n_{10}	n_{1+}
0	n_{01}	n_{00}	n_{0+}
Total	n_{+1}	n_{+0}	n

$$n_{1+} = n_{10} + n_{11} \text{ and } n_{+1} = n_{01} + n_{11}.$$

Marginal homogeneity can be tested using McNemar's score test (McNemar, 1947):

$$Z = \frac{n_{10} - n_{01}}{\sqrt{n_{10} + n_{01}}}, \tag{2.20}$$

where n_{01} and n_{10} are the off-diagonal counts of the 2 by 2 table.

The square of Z is asymptotically chi-squared with one degree of freedom.

2.4.1.2 Multicategory responses

Similarly, responses with more than two categories can be generalised to only two observations in a cluster. Marginal homogeneity can be thought of as:

$$pr(Y_1 = a) = pr(Y_2 = a), \tag{2.21}$$

for $a = 1, \ldots, f$.

If there is no ordering to the response, the nominal model for square tables is defined as:

$$\log[pr(Y_k = a)/pr(Y_k = f)] = \theta_a + \beta_a x_k, \tag{2.22}$$

where $k = 1, 2,\ a = 1, \ldots, f-1$ and $x_1 = 0$.

Marginal homogeneity is specified by $\beta_1 = \cdots = \beta_{f-1} = 0$. The β parameters are interpreted in a similar way to the single β parameter from the binary model (2.18) with each representing the log odds of a response relative to the baseline (the highest category) for Y_2 relative to Y_1. Model (2.22) is saturated for the marginal probabilities since there are $2(f-1)$ parameters for the $2(f-1)$ unconstrained marginal probabilities

(Agresti, 2002). Since there are no parameters specified for the associations, the model is not saturated for the joint distribution.

Although McNemar's test of marginal homogeneity is only applicable to bivariate binary responses, Stuart (1955) proposed a generalised version of McNemar's test that is suitable for nominal responses with more than two levels. In addition, Bhapkar (1966) also proposed a Wald test of marginal homogeneity. The reader is referred to Agresti (2002, Chapter 10) for more details.

If the response is ordinal, the cumulative probabilities are used as opposed to the marginal probabilities. The proportional odds model from (2.12) can be specified for square tables as follows:

$$\text{logit}[pr(Y_k \leq a)] = \theta_a + \beta x_k, \tag{2.23}$$

where $k = 1, 2,\ a = 1, \ldots, f-1$ and $x_1 = 0$.

In contrast to the nominal model, the proportional odds model has the same covariate effect for each level of the response. This is known as the proportional odds assumption and needs to be satisfied for this model to be valid. In terms of interpreting the β parameter, the log odds of $Y_2 \leq a$ is equal to β times the log odds of $Y_1 \leq a$. This model is not saturated for the marginal probabilities since there are f parameters for the $2(f-1)$ unconstrained marginal probabilities (Agresti, 2002). However, it is nested within the saturated model for the multinomial case (2.22) so hence there are $(f-2)$ degrees of freedom for testing fit. In addition, $\beta = 0$ represents marginal homogeneity.

In order to fit the marginal models discussed for binary and multicategory responses in this section (and treat the marginal distributions as dependent), the maximum likelihood approach maximises the multinomial likelihood for the joint distribution. However, as discussed in detail in Agresti (2002, chapter 11.2.5), this is not straightforward. This is because only the marginal distributions are specified, which cannot be substituted in the log-likelihood since it refers to the joint distribution π_{ab}. Although this method makes no assumptions about the joint distribution, the marginal effects are consistent provided the marginal model is correctly specified (Agresti, 2002). The method suffers from similar disadvantages to approaches based on the odds ratio association measure in the sense that maximum likelihood estimation is generally not feasible for cases where there are large cluster sizes, too many predictors or continuous predictors (Agresti, 2002). However, none of these issues are relevant for square tables.

2.4.2 Loglinear approaches to square tables

An alternative approach to analysing square tables is to directly model the joint distribution using loglinear models. It provides a more thorough analysis than the marginal

approaches discussed in Section 2.4.1 since parameters for the associations are specified. Marginal homogeneity cannot be expressed in this form. However, some of the other models that use this approach have marginal homogeneity as a special case. This section discusses the most common of these models with the reader being referred to Agresti (2002, Chapter 10) for a more detailed discussion.

2.4.2.1 Symmetry

The simplest model to consider using this approach is symmetry. For a particular response, the joint distribution is said to satisfy symmetry if:

$$pr(Y_1 = a, Y_2 = b) = pr(Y_1 = b, Y_2 = a) \text{ or equivalently } \pi_{ab} = \pi_{ba}, \tag{2.24}$$

for all $a \neq b$.

The symmetry model can be expressed in both logit and loglinear form. The loglinear form, as defined in Agresti (2002), is given by:

$$\log(\mu_{ab}) = \lambda + \lambda_a + \lambda_b + \lambda_{ab}, \tag{2.25}$$

where $\lambda_{ab} = \lambda_{ba}$, $\mu_{ab} = n\pi_{ab}$ and $\log(\mu_{ab}) = \log(\mu_{ba})$.

The main effects are equal for the symmetry model. When symmetry occurs, marginal homogeneity also occurs and when the response is binary, symmetry is equivalent to marginal homogeneity (Agresti, 2002). However, when the response has more than two categories, it is possible for marginal homogeneity to occur without symmetry. It should be noted that the symmetry model is often not applicable, particularly when the marginal distributions are substantially different. However, as demonstrated in Section 2.6.2.2 when relevant datasets are analysed, it is possible for the symmetry model to give a good fit.

2.4.2.2 Quasi-symmetry

One model that often fits much better than symmetry is quasi-symmetry (Caussinus, 1966). This allows the main effects from the symmetry model to differ. The model takes a loglinear form that is defined as:

$$\log(\mu_{ab}) = \lambda + \lambda_a^1 + \lambda_b^2 + \lambda_{ab}, \tag{2.26}$$

where $\lambda_{ab} = \lambda_{ba}$ for all $a < b$.

As discussed previously, marginal homogeneity cannot be expressed as a loglinear model. However, if marginal homogeneity and quasi-symmetry both hold, then this is equivalent to symmetry (Caussinus, 1966). In other words: quasi symmetry + marginal homogeneity = symmetry.

Marginal homogeneity can therefore be tested by comparing the fit statistics of the symmetry and quasi-symmetry models using a likelihood ratio test (degrees of freedom $= f - 1$). In other words, although marginal homogeneity cannot be expressed as a loglinear model, it can be tested using this approach as opposed to the tests described in Section 2.4.1.

2.4.2.3 Quasi-independence

The final model to mention is quasi-independence (for the off-diagonal cells). The quasi-independence model for square tables can be parameterised in different ways, one of which has a perfect fit on the main diagonal with all non-diagonal cells satisfying quasi-independence. The rationale behind this form of the model is that square tables will often have larger counts on the main diagonal compared to the independence model. In addition, for all 2 by 2 tables within the overall square table (excluding diagonal elements), the odds ratios are consequently equal to 1. The model is only applicable to responses with three or more levels (Agresti, 2002). Quasi-independence has a loglinear model form given by:

$$\log(\mu_{ab}) = \lambda + \lambda_a^1 + \lambda_b^2 + \delta_a I(a = b), \tag{2.27}$$

where $I(a = b) = 1$ if $a = b$, else $= 0$.

The three models described in Section 2.4.2 along with marginal homogeneity are the most common models for square tables. There are numerous other models that have been proposed for square tables, typically extensions of the quasi-symmetry model. For example, the ordinal quasi-symmetry model is particularly applicable when the response is ordered since the loglinear models discussed so far do not use any ordinal information. Agresti (2002) discussed this model in detail as well as ordinal extensions for symmetry (conditional symmetry, McCullagh 1978) and quasi-independence (quasi-uniform association, Goodman 1979a). The models discussed in Sections 2.4.1 and 2.4.2 focused on treating the associations as a nuisance and using odds ratios as the association measure respectively. Section 2.4.3 discusses these models for the dependence ratio approach and expresses the models directly in terms of dependence ratios where possible.

2.4.3 The dependence ratio approach to square tables

In Section 2.2, the dependence ratio approach was discussed in detail for both binary and multicategory responses. However, this related to the general case as opposed to square tables where there are only two observations within a cluster. This section describes the approach for square tables and discusses the models mentioned in Sections 2.4.1 and 2.4.2.

2.4.3.1 Marginal homogeneity

Marginal homogeneity was discussed in Section 2.4.1 and was shown to be a feature of the marginal models described for both binary (2.18) and multicategory responses (2.22, 2.23), when the β coefficient(s) are set equal to 0. Recall that for the dependence ratio approach, the profile probabilities (joint distribution) can be expressed in closed form in terms of the first-order moments (marginal regression model) and dependence ratios (association model) of all orders. For this case, marginal homogeneity is specified by combining the relevant marginal regression model (with β coefficient(s) set equal to zero) from Section 2.4.1 with a relevant association model, using the profile probability. One form of the marginal homogeneity model is one that has no constraints on the association parameters (the dependence ratios). In other words, it assumes a saturated association structure. However, marginal homogeneity is still present with constraints on the dependence ratios.

The maximum likelihood estimates under marginal homogeneity can be calculated. This is now discussed for the simplest situation of a bivariate binary response $\boldsymbol{Y} = (Y_1, Y_2)$. Recall from Section 2.1 that there are four possible response profiles: (1,1), (0,0), (1,0) and (0,1) with 1 regarded as the response of interest or success and 0 regarded as the baseline or failure. These can consequently be represented in a 2 by 2 table. Assuming a multinomial sampling scheme, the joint distribution is completely specified by μ_1, μ_2 and τ. It is the saturated model. However, marginal homogeneity assumes $\mu_1 = \mu_2 = \mu$ so consequently has one fewer parameter than the saturated model. Table 2.2 shows the cell probabilities for marginal homogeneity and the corresponding counts are as in Table 2.1.

Table 2.2: Cell probabilities for a bivariate binary response under marginal homogeneity

	Y_2		
Y_1	1	0	Total
1	μ_{12}	$\mu - \mu_{12}$	μ
0	$\mu - \mu_{12}$	$1 - 2\mu + \mu_{12}$	$1 - \mu$
Total	μ	$1 - \mu$	1

The second-order moment is given by $\mu_{12} = pr(Y_1 = 1, Y_2 = 1)$. Under marginal homogeneity, constraints on the association parameters are not relevant since there is only a single dependence ratio parameter, which is as specified in (2.28).

$$\tau_{12} = \frac{\mu_{12}}{\mu^2}. \tag{2.28}$$

Assuming a multinomial sampling scheme, the likelihood function for (μ, μ_{12}), for a sample of size n, is given by:

$$Ł(\mu, \mu_{12}) \propto (1 - 2\mu + \mu_{12})^{n_{00}} \ (\mu - \mu_{12})^{n_{01}} \ (\mu - \mu_{12})^{n_{10}} \ (\mu_{12})^{n_{11}}. \tag{2.29}$$

The likelihood expressed in terms of τ_{12} (denoted τ) rather than μ_{12} is given by:

$$Ł(\mu, \tau) \propto (1 - 2\mu + \mu^2\tau)^{n_{00}} \ (\mu - \mu^2\tau)^{n_{01}} \ (\mu - \mu^2\tau)^{n_{10}} \ (\mu^2\tau)^{n_{11}}. \tag{2.30}$$

It can be shown (see Appendix for proof) that the maximum likelihood estimates of μ and τ, under marginal homogeneity, for a bivariate binary response, are given by:

$$\hat{\mu} = \frac{\frac{n_{1+}}{n} + \frac{n_{+1}}{n}}{2} = \frac{n_{1+} + n_{+1}}{2n}, \tag{2.31}$$

$$\hat{\tau} = \frac{\hat{\mu}_{12}}{\hat{\mu}^2} = \frac{\frac{n_{11}}{n}}{\hat{\mu}^2} = \frac{4nn_{11}}{(2n_{11} + n_{10} + n_{01})^2}. \tag{2.32}$$

Ekholm (2003) also discussed the issue of marginal homogeneity for a bivariate binary response. He showed that μ is reasonably variation independent of τ when μ is less than 0.5. However, as μ approaches 1, τ also approaches 1. The expected Fisher information for these estimates is also calculated in the Appendix at the end of the current paper.

2.4.3.2 Symmetry

Symmetry is equivalent to marginal homogeneity for 2 by 2 square tables. Clearly when 3 by 3 square tables or larger are considered, constraints can be imposed on the dependence ratios and marginal homogeneity can occur without symmetry. However, symmetry cannot occur without marginal homogeneity. Given marginal homogeneity, the symmetry model can be directly expressed in terms of dependence ratios. The proof is now discussed.

Given (2.14) and (2.24), we know that for the symmetry model:

$$\tau_{12}^{(a,b)}\mu_1^{(a)}\mu_2^{(b)} = \tau_{12}^{(b,a)}\mu_1^{(b)}\mu_2^{(a)}, \tag{2.33}$$

where $\mu_k^{(a)} = pr(Y_k = a)$.

In addition, in Section 2.4.2 it was stated that symmetry implies marginal homogeneity. Consequently, for the symmetry model:

$$\mu_1^{(a)} = \mu_2^{(a)} \text{ for all } a,\ \mu_1^{(b)} = \mu_2^{(b)} \text{ for all } b. \tag{2.34}$$

From (2.33) and (2.34), the symmetry model can therefore be specified in terms of dependence ratios as follows:

$$\tau_{12}^{(a,b)} = \tau_{12}^{(b,a)},\ \text{for all } a \neq b. \tag{2.35}$$

The symmetry model can be fitted in drm by specifying (2.35) as constraints in the marginal homogeneity model, where the marginal homogeneity model has no constraints on the dependence ratios.

2.4.3.3 Quasi-symmetry and Quasi-independence

In contrast to the symmetry model, quasi-symmetry and quasi-independence cannot be naturally expressed in terms of dependence ratios. However, by using the fitted values from the loglinear models in Section 2.4.2, the dependence ratio parameters can be estimated.

In some cases, a satisfactory fit may not be achieved with any of the common models for square tables. In such cases, it is appropriate to apply relevant constraints to the dependence ratios based on the observed dependence ratios or theoretical justifications.

2.5 Extensions to matched sets

The models described for square tables in Section 2.4 do not allow for the inclusion of additional explanatory variables. In contrast, the models described in Section 2.2 are applicable to more than two observations within a cluster and do allow for the inclusion of such variables. This section extends the models from Section 2.4 to matched sets data (multi-way contingency tables) so in these cases explanatory variables are not included. However, the datasets that are typically analysed with these models normally have limited covariate information available anyway. In addition, given the fact that males and females often follow different association patterns, they are often analysed separately, if such information is available. To be consistent with Section 2.4, existing approaches are discussed first before the dependence ratio is considered.

2.5.1 Marginal models for matched sets data

The marginal models described in Section 2.4.1 can be extended to matched sets. Here the extension for just binary responses is presented since the nominal and ordinal responses can be extended similarly. The binary response model for square tables (2.18) can be extended to matched sets data (Agresti 2002, chapter 11) as follows:

$$\text{logit}[pr(Y_k = 1)] = \theta + \beta_k, \tag{2.36}$$

where $k = 1, \ldots, q$ and $\beta_1 = 0$.

Model (2.36) is saturated for the marginal probabilities since the q marginal probabilities are represented by q parameters, with no association parameter specified for the associations. Marginal homogeneity is achieved by $\beta_2 = \cdots = \beta_q = 0$. Under marginal homogeneity:

$$pr(Y_1 = 1) = \cdots = pr(Y_q = 1). \tag{2.37}$$

Marginal homogeneity can be tested with a likelihood ratio test that compares model (2.36) to the marginal homogeneity model. In this case there are $(q - 1)$ degrees of freedom. As discussed in Section 2.4.1, maximum likelihood estimation that treats the marginal distributions as dependent is not straightforward. Although the difficulties encountered were not an issue for square tables, they are more problematic for matched sets data with larger cluster sizes. Some alternative methods have been proposed that specify odds ratio parameters for the associations such as the mixed parameterisation by Fitzmaurice and Laird (1993) in which the joint distribution is specified in terms of the marginal distributions and conditional log-odds ratios. However, maximum likelihood estimation is still often not feasible for large cluster sizes. Lesaffre et al. (2000) stated that maximum likelihood estimation with odds ratios is not feasible for cluster sizes larger than five.

In terms of dealing with large cluster sizes (as well as a large number of predictor variables and continuous variables), some statisticians would advocate the use of the GEE approach. For example, Agresti (2002) discussed the GEE approach in detail for dealing with this issue. However, as discussed in Section 2.1, the GEE approach treats the associations as a nuisance whereas the focus of this thesis is on putting equal emphasis on the marginal regression and the associations. The dependence ratio approach discussed in Section 2.2 achieves this as well as being able to cope with large cluster sizes. Although not relevant to the matched sets discussion in Section 2.5.3, the dependence ratio approach can often cope with a larger number of predictor variables and continuous variables.

2.5.2 Loglinear approaches to matched sets data

In Section 2.4.2, loglinear approaches were discussed for square tables. This section focuses on the extensions for matched sets data and concentrates on the extension of symmetry. For this case, $(Y_1, \ldots, Y_q)$ is defined as the q responses in a cluster, for subunit $k = 1, \ldots, q$. Agresti (2002) also discusses extending quasi-symmetry for matched sets data.

2.5.2.1 Complete symmetry

The symmetry model for square tables can be extended to matched sets data through the complete symmetry model, which is specified as:

$$\pi_{\boldsymbol{i}} = \pi_{\boldsymbol{j}}, \tag{2.38}$$

for any permutation $\boldsymbol{j} = (j_1, \ldots, j_q)$ of $\boldsymbol{i} = (i_1, \ldots, i_q)$, where $\pi_{\boldsymbol{i}}$, $\pi_{\boldsymbol{j}}$ represent the profile probabilities.

For example, for a binary response and three observations within a cluster, $\pi_{(1,0,0)} = \pi_{(0,1,0)} = \pi_{(0,0,1)}$. The complete symmetry model can be expressed in loglinear form (as given in Agresti, 2002):

$$\log(\mu_{\boldsymbol{i}}) = \lambda_{ab\ldots m}, \tag{2.39}$$

where a is the minimum in $\boldsymbol{i}$, b is the second smallest and m is the largest. In addition, $\mu_{\boldsymbol{i}} = n\pi_{\boldsymbol{i}}$.

2.5.3 The dependence ratio approach to matched sets

2.5.3.1 Marginal Homogeneity

As for the discussion on square tables in Section 2.4.3.1, marginal homogeneity is specified by combining a relevant marginal model (with β coefficients set to zero) with an association model. Consider the simplest scenario for matched sets data in which there is a binary response with 3 repeated observations, denoted by $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$. There are consequently 8 possible response profiles: (1,1,1),(1,1,0),(1,0,1),(0,1,1),(0,0,1),(0,1,0),(1,0,0) and (0,0,0). As discussed in Section 2.2, the profile probabilities (joint distribution) can be expressed in closed form in terms of the marginal probabilities (first-order moments) and dependence ratios of all orders.

Assuming a multinomial sampling scheme, 7 parameters are required in order to completely specify the joint distribution of a 2^3 response since one is redundant. Firstly, the marginal regression model consists of 3 marginal probabilities: $\mu_1 = pr(Y_1 = 1), \mu_2 = pr(Y_2 = 1), \mu_3 = pr(Y_3 = 1)$. Under marginal homogeneity, these are set equal. The association model consists of 3 dependence ratios of order 2: $\tau_{12} = \frac{pr(Y_1=1,Y_2=1)}{pr(Y_1=1)pr(Y_2=1)}, \tau_{13}, \tau_{23}$ and one of order 3: $\tau_{123} = \frac{pr(Y_1=1,Y_2=1,Y_3=1)}{pr(Y_1=1)pr(Y_2=1)pr(Y_3=1)}$. For marginal homogeneity, additional constraints may be imposed for the association model (if appropriate). For example, the 3 second-order dependence ratios of order 2 may be constrained to be equal.

2.5.3.2 Complete Symmetry

In addition to the symmetry model, the complete symmetry model can also be specified directly in terms of dependence ratios. The proof is now discussed for three repeated observations and three response values: a, b, c. Given the discussion on the symmetry model from Section 2.4.3.2, we know that:

$$\tau_{123}^{(a,b,c)}\mu_1^{(a)}\mu_2^{(b)}\mu_3^{(c)} = \tau_{123}^{(a,c,b)}\mu_1^{(a)}\mu_2^{(c)}\mu_3^{(b)} = \cdots = \tau_{123}^{(c,b,a)}\mu_1^{(c)}\mu_2^{(b)}\mu_3^{(a)}, \tag{2.40}$$

where $\mu_k^{(a)} = pr(Y_k = a)$, $\mu_k^{(b)} = pr(Y_k = b)$ and $\mu_k^{(c)} = pr(Y_k = c)$.

Complete symmetry implies second-order marginal symmetry (Agresti, 2002). Second-order marginal symmetry states that $Pr(Y_k = a, Y_l = b)$ for all k and l and for all pairs (a, b). Second-order marginal symmetry also implies marginal homogeneity. In other words, for the complete symmetry model:

$$\mu_1^{(a)} = \mu_2^{(a)} = \mu_3^{(a)}, \;\; \mu_1^{(b)} = \mu_2^{(b)} = \mu_3^{(b)}, \;\; \mu_1^{(c)} = \mu_2^{(c)} = \mu_3^{(c)}. \tag{2.41}$$

From (2.40) and (2.41), the complete symmetry model can therefore be specified in terms of dependence ratios as follows:

$$\tau_{123}^{(a,b,c)} = \tau_{123}^{(a,c,b)} = \tau_{123}^{(b,a,c)} = \tau_{123}^{(b,c,a)} = \tau_{123}^{(c,a,b)} = \tau_{123}^{(c,b,a)} \text{ for all } a \neq b \neq c. \tag{2.42}$$

The complete symmetry model will rarely fit well, but it has been discussed due to its convenient formulation in terms of dependence ratios. Quasi-symmetry and quasi-independence may give a satisfactory fit, although they cannot be directly expressed in terms of dependence ratios. However, they can be fitted in terms of dependence ratios by using the fitted values from the corresponding loglinear model.

2.6 Application to datasets

In this section, four relevant datasets are analysed with the dependence ratio approach, using the relevant models from Sections 2.3, 2.4 and 2.5. The first three datasets relate to square tables so use the models described in Section 2.4. The final dataset considers a rater agreement dataset with seven repeated observations within a cluster so models from Section 2.5 are considered for this. In each case, if a satisfactory fit cannot be obtained with one of the models described in the previous sections, appropriate constraints are imposed on the dependence ratios. Association structures as described in Section 2.3 are also used where applicable. The R package drm is used, with additional calculations being performed where necessary.

2.6.1 Bilirubin abnormalities dataset

The following dataset from Example 18.1 of Walker (2002) assesses bilirubin abnormalities following drug treatment. Each of the 86 patients in the study had their bilirubin levels measured before the start of the study (pre-treatment) and following three months of treatment with an experimental drug (post-treatment). In each case, the response of interest is binary with 1 representing an abnormally high bilirubin rate that is higher than the upper limit of the normal range and 0 representing values below this limit. The aim of the study was to determine if there is a significant difference between the pre and post bilirubin abnormality rates. In other words, testing for marginal homogeneity. Since the study has a bivariate binary response, the data can be represented in a 2 by 2 table, as shown in Table 2.3.

In order to specify the joint distribution using the profile probability, two univariate probabilities and one dependence ratio are required. The estimates of the saturated model with these three parameters are given in Table 2.4.

Table 2.3: Bilirubin abnormality: pre and post drug treatment

	Post-Treatment		
Pre-Treatment	1	0	Total
1	6	6	12
0	14	60	74
Total	20	66	86

For marginal distributions Y_1 (pre-treatment) and Y_2 (post-treatment), the single dependence ratio parameter is given by:

$$\tau_{12} = \frac{pr(Y_1 = 1, Y_2 = 1)}{pr(Y_1 = 1)\, pr(Y_2 = 1)}. \tag{2.43}$$

Similarly, the two univariate marginal probabilities are given by:

$$\mu_1 = pr(Y_1 = 1),\ \mu_2 = pr(Y_2 = 1). \tag{2.44}$$

Table 2.4: Parameter estimates for the saturated model

Sub model	Parameter	Estimate	Standard error
Regression model	$\hat{\mu}_1$	0.1395	0.0374
	$\hat{\mu}_2$	0.2326	0.0456
Association model	$\hat{\tau}_{12}$	2.1500	0.5757

Table 2.4 shows that the bilirubin rate is higher at post-treatment (0.2326) than pre-treatment (0.1395), thus suggesting that the drug is actually a hindrance. Testing for marginal homogeneity will enable an assessment of this claim. The dependence ratio parameter shows that the probability of having two abnormally high bilirubin rates is 2.15 times higher than the corresponding probability under independence. Although the standard error for the dependence ratio was naturally available from drm, the standard errors for the univariate marginal probabilities were not since the logistic link function was used for the marginal regression model. The standard errors for the marginal probabilities were obtained by paramterising the Fisher information matrix in terms of the marginal probabilities and the dependence ratio and then inverting this matrix (see Appendix).

As a comparison, the odds ratio is also calculated. For a bivariate binary response, all odds ratio formulations are the same and are given by:

$$\chi_{12} = \frac{\mu_{12}(1 - \mu_1 - \mu_2 + \mu_{12})}{(\mu_2 - \mu_{12})(\mu_1 - \mu_{12})} = \frac{pr(Y_1 = 1, Y_2 = 1)pr(Y_1 = 0, Y_2 = 0)}{pr(Y_1 = 0, Y_2 = 1)pr(Y_1 = 1, Y_2 = 0)}. \tag{2.45}$$

The odds ratio estimate for the saturated model is 4.29. In other words, the odds of an abnormally high bilirubin rate are 4.29 times higher for patients at post-treatment than pre-treatment.

The main aim of the study was to test for marginal homogeneity between the treatments. This model has one fewer parameter than the saturated model since $\mu_1 = \mu_2 = \mu$. The likelihood ratio test statistic for testing marginal homogeneity is 3.291 (1df) compared to the saturated model, yielding a p-value of 0.0697. Similarly, McNemar's test gives a test statistic of 3.2 and a p-value of 0.0736. In other words, there is some weak evidence that marginal homogeneity between the pre and post treatments does not hold at the 5% significance level. Table 2.5 shows the parameter estimates for this model.

Table 2.5: Parameter estimates under marginal homogeneity

Sub model	Parameter	Estimate	Standard error
Regression model	$\hat{\mu}$	0.1860	0.0329
Association model	$\hat{\tau}_{12}$	2.0156	0.5590

The conclusion to make is that the probability of an abnormally high bilirubin rate is 0.1860, for each of the pre and post treatments. In addition, the dependence ratio parameter tells us that the probability of having two abnormally high bilirubin rates is approximately two times the corresponding probability under independence. In contrast, the odds ratio estimate is given by $\frac{14}{6} = 2.33$ (Breslow and Day, 1980 pg 165).

Due to the smaller counts that make the asymptotic chi-squared assumption more questionable, McNemar's exact test (Fay, 2011) was also used to test for marginal homogeneity since it is more applicable to these cases. This test gave a p-value of 0.1153 so thus not altering the previous conclusions (5% significance level). The conclusions obtained with regards to marginal homogeneity coincide with Sun and Yang (2008). However, Sun and Yang did not consider McNemar's exact test or a dependence ratio approach.

It can be shown that $\hat{\tau}^{MH} \leq \hat{\tau}^{S}$, where $\hat{\tau}^{MH}$ represents the dependence ratio under marginal homogeneity and $\hat{\tau}^{S}$ represents the dependence ratio for the saturated case.

In order to prove this, first recall that the arithmetic mean (AM) of a list of non-negative real numbers is greater than or equal to the geometric mean (GM) of the same list. In addition, the two means are equal if and only if every number in the list is identical.

Consider two non-negative numbers x and y:

$(x-y)^2 = x^2 - 2xy + y^2 \geq 0 \implies x^2 + y^2 \geq 2xy.$

Let $x = \sqrt{a}$ and $y = \sqrt{b}$:

$(\sqrt{a} - \sqrt{b})^2 = \sqrt{a}^2 - 2\sqrt{a}\sqrt{b} + \sqrt{b}^2 \geq 0 \implies \sqrt{a}^2 + \sqrt{b}^2 \geq 2\sqrt{a}\sqrt{b}.$

Therefore, $a + b \geq 2\sqrt{ab} \implies \frac{a+b}{2} \geq \sqrt{ab}.$

Hence, AM $\geq$ GM with equality if and only if $a = b$.

Or alternatively:

$$\frac{1}{ab} \geq \frac{1}{\left(\frac{a+b}{2}\right)^2}. \tag{2.46}$$

Recall:

$$\widehat{\tau}^S = \frac{\widehat{\mu}_{12}}{\widehat{\mu}_1\,\widehat{\mu}_2} = \frac{\frac{n_{11}}{n}}{\frac{n_{1+}}{n}\,\frac{n_{+1}}{n}},$$

$$\widehat{\tau}^{MH} = \frac{\widehat{\mu}_{12}}{\widehat{\mu}_1^{MH}\,\widehat{\mu}_2^{MH}} = \frac{\frac{n_{11}}{n}}{\frac{\left(\frac{n_{1+}}{n}+\frac{n_{+1}}{n}\right)}{2}\,\frac{\left(\frac{n_{1+}}{n}+\frac{n_{+1}}{n}\right)}{2}} = \frac{\frac{n_{11}}{n}}{\left(\frac{\frac{n_{1+}}{n}+\frac{n_{+1}}{n}.}{2}\right)^2} \leq \tau^S.$$

From (2.46):

$$\frac{1}{\left(\frac{\frac{n_{1+}}{n}+\frac{n_{+1}}{n}}{2}\right)^2} \leq \frac{1}{\frac{n_{1+}}{n}\,\frac{n_{+1}}{n}} = \widehat{\tau}^S.$$

In other words, the dependence ratio estimate assuming marginal homogeneity is less than or equal to the dependence ratio estimate under the saturated model.

2.6.2 Eye grade datasets

The following section focuses on 4 by 4 square tables by considering the commonly analysed eye grade datasets. Females (Stuart, 1955) and males (Stuart, 1953) had each of their right and left eyes graded, according to a four point scale ranging from lowest (1) to highest (4). Note that the right eye is treated as the first measurement and the left eye as the second measurement.

2.6.2.1 Eye grade females dataset

This dataset concerns females who were aged 30 to 39 and worked in British factories from 1943 to 1946. This was first analysed by Stuart (1955) and has since been analysed by numerous statisticians including Caussinus (1966), Bishop et al. (1975, p.284), McCullagh (1978), Goodman (1979b), Agresti (1983), Tomizawa (1989) and Bergsma (1997). Table 2.6 shows the data and Table 2.7 shows the estimates for the saturated model. The saturated model has six regression parameters (assuming the highest marginal probability category to be the baseline) and nine dependence ratios

(association parameters), with the dependence ratios as expressed in (2.14). These are highlighted in bold in Table 2.7. Since this is the saturated model, which replicates the observed data, estimates for the baseline categories are also included.

Table 2.6: Unaided distance vision of British women (Stuart, 1955)

	Left Eye Grade				
Right Eye Grade	Highest (4)	Second (3)	Third (2)	Worst (1)	Total
Highest (4)	1520	266	124	66	1976
Second (3)	234	1512	432	78	2256
Third (2)	117	362	1772	205	2456
Worst (1)	36	82	179	492	789
Total	1907	2222	2507	841	7477

Table 2.7: Parameter estimates for the saturated model

	Left Eye Grade				
Right Eye Grade	Highest (4)	Second (3)	Third (2)	Worst (1)	Marginal
Highest (4)	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(4,4)}}$= 3.0160	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(4,3)}}$= 0.4530	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(4,2)}}$= 0.1872	$\hat{\tau}_{12}^{(4,1)}$= 0.2970	$\hat{\mu}_1^{(4)}$=0.2643
Second (3)	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(3,4)}}$= 0.4067	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(3,3)}}$= 2.2553	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(3,2)}}$= 0.5711	$\hat{\tau}_{12}^{(3,1)}$= 0.3074	$\hat{\boldsymbol{\mu}}_{\mathbf{1}}^{\mathbf{(3)}}$= 0.3017
Third (2)	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(2,4)}}$= 0.1868	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(2,3)}}$= 0.4960	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(2,2)}}$= 2.1518	$\hat{\tau}_{12}^{(2,1)}$= 0.7421	$\hat{\boldsymbol{\mu}}_{\mathbf{1}}^{\mathbf{(2)}}$= 0.3285
Worst (1)	$\hat{\tau}_{12}^{(1,4)}$=0.1789	$\hat{\tau}_{12}^{(1,3)}$=0.3497	$\hat{\tau}_{12}^{(1,2)}$= 0.6766	$\hat{\tau}_{12}^{(1,1)}$= 5.5440	$\hat{\boldsymbol{\mu}}_{\mathbf{1}}^{\mathbf{(1)}}$= 0.1055
Marginal	$\hat{\mu}_2^{(4)}$= 0.2550	$\hat{\boldsymbol{\mu}}_{\mathbf{2}}^{\mathbf{(3)}}$= 0.2972	$\hat{\boldsymbol{\mu}}_{\mathbf{2}}^{\mathbf{(2)}}$= 0.3353	$\hat{\boldsymbol{\mu}}_{\mathbf{2}}^{\mathbf{(1)}}$= 0.1125	1

As a starting point, marginal homogeneity between left-eye vision and right-eye vision is considered as a potential model, with three fewer parameters than the saturated model since the three marginal probabilities for the right eye are constrained to be equal to the corresponding marginal probabilities on the left eye. The model does not give a satisfactory fit (likelihood ratio test statistic = 11.987, degrees of freedom = 3, p-value = 0.0075). The symmetry model imposes the following additional constraints on the association parameters:

$$\tau_{12}^{(3,4)} = \tau_{12}^{(4,3)}, \tau_{12}^{(3,2)} = \tau_{12}^{(2,3)}, \tau_{12}^{(4,2)} = \tau_{12}^{(2,4)}. \tag{2.47}$$

The symmetry model also gives a poor fit when compared to the saturated model (likelihood ratio test statistic = 19.25, degrees of freedom = 6, p-value < 0.001). From the literature, quasi-independence also gives a poor fit (p-value < 0.001). Quasi-symmetry gave a satisfactory fit with a p-value of 0.0637. However, a better fit in terms of dependence ratios is desirable. Given the ordinal nature of the data, the proportional odds model (2.23) is now considered for the regression model (the first-order moments) since the models considered so far do not exploit the ordinal nature of the data. The model has two fewer parameters than the saturated model. In contrast to marginal homogeneity, the probabilities vary between eyes though of course the proportional odds model is based upon cumulative probabilities. The parameter estimates are given in Table 2.8 with the regression model as described in (2.23) and the association model consisting of

the nine dependence ratios (no constraints on the dependence ratios). The model gives a satisfactory fit (likelihood ratio test statistic = 0.394, degrees of freedom = 2, p-value = 0.8212). However, it is advantageous to reduce the number of association parameters, to aid the interpretation of the model and also increase efficiency. In order to do this, the observed dependence ratios were assessed to determine appropriate constraints that could be imposed on the dependence ratios. Table 2.9 shows the observed dependence ratios.

Table 2.8: Parameter estimates for the proportional odds regression model with an unconstrained association model (denoted PO unconstrained)

Sub model	Parameter	Estimate	Standard error
Regression model	$\hat{\theta}_1$	-2.1278	0.0337
	$\hat{\theta}_2$	-0.2645	0.0227
	$\hat{\theta}_3$	1.0212	0.0257
	Left eye	-0.0539	0.0158
Association model	$\hat{\tau}_{1,2}^{(2,2)}$	2.1522	0.0305
	$\hat{\tau}_{12}^{(3,3)}$	2.2547	0.0347
	$\hat{\tau}_{12}^{(4,4)}$	3.0164	0.0523
	$\hat{\tau}_{12}^{(2,3)}$	0.4966	0.0207
	$\hat{\tau}_{12}^{(3,2)}$	0.5704	0.0208
	$\hat{\tau}_{12}^{(2,4)}$	0.1855	0.0160
	$\hat{\tau}_{12}^{(4,2)}$	0.1886	0.0157
	$\hat{\tau}_{12}^{(3,4)}$	0.4032	0.0220
	$\hat{\tau}_{12}^{(4,3)}$	0.4569	0.0226

Table 2.9: Observed dependence ratios

Observed dependence ratio	Value
$\tau_{12}^{(2,2)}$	2.1518
$\tau_{12}^{(3,3)}$	2.2553
$\tau_{12}^{(4,4)}$	3.0160
$\tau_{12}^{(2,3)}$	0.4960
$\tau_{12}^{(3,2)}$	0.5711
$\tau_{12}^{(2,4)}$	0.1868
$\tau_{12}^{(4,2)}$	0.1872
$\tau_{12}^{(3,4)}$	0.4067
$\tau_{12}^{(4,3)}$	0.4530

Table 2.9 shows which of the observed dependence ratios are similar to one another. A stepwise procedure was used in the sense that all potential constraints were applied to the model as a starting point: $\tau_{12}^{(2,2)} = \tau_{12}^{(3,3)}$, $\tau_{12}^{(2,3)} = \tau_{12}^{(3,2)}$, $\tau_{12}^{(2,4)} = \tau_{12}^{(4,2)}$, $\tau_{12}^{(3,4)} = \tau_{12}^{(4,3)}$ with fewer constraints being applied (if required), until a satisfactory fit is obtained. Applying the constraints: $\tau_{12}^{(3,4)} = \tau_{12}^{(4,3)}$ and $\tau_{12}^{(2,4)} = \tau_{12}^{(4,2)}$ to the model from Table 2.8 was found to give the best fit (likelihood ratio test statistic = 5.599, degrees of freedom = 4, p-value = 0.2311) and is the chosen model for the females, denoted POWC (proportional odds with constraints). Including the additional constraint $\tau_{12}^{(2,3)} = \tau_{12}^{(3,2)}$, denoted PO Symmetry, gave a poor fit (likelihood ratio test statistic = 12.432, degrees of freedom = 5, p-value = 0.0293). The parameter estimates for the POWC model are given in Table 2.10.

Table 2.10: Parameter estimates for the proportional odds regression model with constraints on the association model (POWC)

Sub model	Parameter	Estimate	Standard error
Regression model	$\hat{\theta}_1$	−2.1124	0.0328
	$\hat{\theta}_2$	−0.2505	0.0215
	$\hat{\theta}_3$	1.0354	0.0248
	Left eye	−0.0271	0.0085
Association model	$\hat{\tau}_{12}^{(3,4)} = \hat{\tau}_{12}^{(4,3)}$	0.4304	0.0182
	$\hat{\tau}_{12}^{(2,4)} = \hat{\tau}_{12}^{(4,2)}$	0.1873	0.0117
	$\hat{\tau}_{12}^{(2,3)}$	0.5148	0.0190
	$\hat{\tau}_{12}^{(3,2)}$	0.5514	0.0186
	$\hat{\tau}_{12}^{(2,2)}$	2.1528	0.0305
	$\hat{\tau}_{12}^{(3,3)}$	2.2544	0.0347
	$\hat{\tau}_{12}^{(4,4)}$	3.0166	0.0523

The regression model can be used to estimate the univariate cumulative probabilities with right eye treated as the baseline for the covariate effect. The main conclusion to make is that the right eye is, on average, better than the left eye. The main point to make for the association model is that the probability of a female rating both her right and left eyes highest is 3.02 times higher than if the two events were independent. Table 2.11 shows a comparison of the models considered. The PO Unconstrained and POWC models clearly give a substantially better fit over the traditional models for square tables (marginal homogeneity, symmetry, quasi-symmetry and quasi-independence). However, it is of interest to see how they compare to other models that have been proposed for the eye grade females dataset, some of which are also shown in Table 2.11. The conditional symmetry ordinal extension of the symmetry model (McCullagh, 1978) yields a satisfactory fit but not as satisfactory as the PO Unconstrained and POWC models. The

diagonals-parameter symmetry model (Goodman, 1979b) did however yield a superior fit.

Table 2.11: Comparison of model fit for the females

Model	G^2	df	p-value
Marginal homogeneity	11.987	3	0.0075
Symmetry	19.25	6	< 0.001
Quasi-independence	199.1	5	< 0.001
PO Symmetry	12.432	5	0.0293
Quasi-symmetry	7.271	3	0.0637
Conditional Symmetry	7.23	5	0.2041
POWC	**5.599**	**4**	**0.2311**
PO Unconstrained	0.394	2	0.8212
Diagonals-parameter symmetry	0.5	3	0.9978

2.6.2.2 Eye grade males dataset

The eye grade dataset for the males from Stuart (1953) is now analysed. The dataset concerns 3242 males having each of their right and left eyes rated according to the same four point scale as the females. Table 2.12 shows the data and Table 2.13 gives the parameter estimates for the saturated model.

Table 2.12: Unaided distance vision of British men (Stuart, 1953)

	Left Eye Grade				
Right Eye Grade	Highest (4)	Second (3)	Third (2)	Worst (1)	Total
Highest (4)	821	112	85	35	1053
Second (3)	116	494	145	27	782
Third (2)	72	151	583	87	893
Worst (1)	43	34	106	331	514
Total	1052	791	919	480	3242

Table 2.13: Parameter estimates for the saturated model

	Left Eye Grade				
Right Eye Grade	Highest (4)	Second (3)	Third (2)	Worst (1)	Marginal
Highest (4)	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(4,4)}}$=2.4027	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(4,3)}}$=0.4359	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(4,2)}}$=0.2848	$\hat{\tau}_{12}^{(4,1)}$=0.2245	$\hat{\mu}_1^{(4)}$=0.3249
Second (3)	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(3,4)}}$=0.4571	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(3,3)}}$=2.5891	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(3,2)}}$=0.6541	$\hat{\tau}_{12}^{(3,1)}$=0.2332	$\hat{\boldsymbol{\mu}}_{\mathbf{1}}^{\mathbf{(3)}}$=0.2412
Third (2)	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(2,4)}}$=0.2485	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(2,3)}}$=0.6930	$\hat{\boldsymbol{\tau}}_{\mathbf{12}}^{\mathbf{(2,2)}}$=2.3031	$\hat{\tau}_{12}^{(2,1)}$=0.6580	$\hat{\boldsymbol{\mu}}_{\mathbf{1}}^{\mathbf{(2)}}$=0.2754
Worst (1)	$\hat{\tau}_{12}^{(1,4)}$=0.2578	$\hat{\tau}_{12}^{(1,3)}$=0.2711	$\hat{\tau}_{12}^{(1,2)}$=0.7275	$\hat{\tau}_{12}^{(1,1)}$=4.3495	$\hat{\boldsymbol{\mu}}_{\mathbf{1}}^{\mathbf{(1)}}$=0.1585
Marginal	$\hat{\mu}_2^{(4)}$=0.3244	$\hat{\boldsymbol{\mu}}_{\mathbf{2}}^{\mathbf{(3)}}$=0.2440	$\hat{\boldsymbol{\mu}}_{\mathbf{2}}^{\mathbf{(2)}}$=0.2835	$\hat{\boldsymbol{\mu}}_{\mathbf{2}}^{\mathbf{(1)}}$=0.1481	1

In contrast to the females, marginal homogeneity (with no constraints on the dependence ratios) gives a satisfactory fit for males when compared to the saturated model (likelihood ratio test statistic = 3.68, degrees of freedom = 3, p-value = 0.2982). In other words, there is no significant difference between the vision of the right and left eye in males.

The parameter estimates for this model are given in Table 2.14. It is also of interest to reduce the number of association parameters using appropriate constraints. Table 2.15 shows the observed dependence ratios.

Table 2.14: Parameter estimates under marginal homogeneity with an unconstrained association model

Sub model	Parameter	Estimate	Standard error
Regression model	$\hat{\theta}_1$	-0.7502	0.0496
	$\hat{\theta}_2$	-0.1498	0.0410
	$\hat{\theta}_3$	-0.2912	0.0418
Association model	$\hat{\tau}_{12}^{(2,2)}$	2.3025	0.0574
	$\hat{\tau}_{12}^{(3,3)}$	2.5887	0.0719
	$\hat{\tau}_{12}^{(4,4)}$	2.4032	0.0533
	$\hat{\tau}_{12}^{(2,3)}$	0.6903	0.0427
	$\hat{\tau}_{12}^{(3,2)}$	0.6565	0.0421
	$\hat{\tau}_{12}^{(2,4)}$	0.2505	0.0254
	$\hat{\tau}_{12}^{(4,2)}$	0.2826	0.0260
	$\hat{\tau}_{12}^{(3,4)}$	0.4626	0.0337
	$\hat{\tau}_{12}^{(4,3)}$	0.4309	0.0328

Table 2.15: Observed dependence ratios

Observed dependence ratio	Value
$\tau_{12}^{(2,2)}$	2.3031
$\tau_{12}^{(3,3)}$	2.5891
$\tau_{12}^{(4,4)}$	2.4027
$\tau_{12}^{(2,3)}$	0.6930
$\tau_{12}^{(3,2)}$	0.6541
$\tau_{12}^{(2,4)}$	0.2485
$\tau_{12}^{(4,2)}$	0.2848
$\tau_{12}^{(3,4)}$	0.4571
$\tau_{12}^{(4,3)}$	0.4359

Based on the observed dependence ratios and using the same stepwise procedure as the females, a good fitting model was found by applying the following constraints to the marginal homogeneity model from Table 2.14:

$$\tau_{12}^{(2,4)} = \tau_{12}^{(4,2)}, \tau_{12}^{(2,3)} = \tau_{12}^{(3,2)}, \tau_{12}^{(3,4)} = \tau_{12}^{(4,3)}. \tag{2.48}$$

This is the symmetry model. Recall that marginal homogeneity is still present after applying constraints to the association parameters of the model in Table 2.14. The symmetry model gives a good fit (likelihood ratio statistic = 4.77, 6 degrees of freedom, p-value = 0.5736) when compared to the saturated model, with parameter estimates given in Table 2.16. For the regression model, the marginal probabilities are assumed equal for both the left and the right eyes. The highest category (the baseline) has the largest marginal probability (0.3246). In terms of the association model, the probability of a male rating both their right and left eyes as highest is 2.40 times higher than independence. However, the probability of a male rating both their eyes as second highest is 2.59 times higher than independence. Table 2.17 shows a comparison of the models considered.

Table 2.16: Parameter estimates for the symmetry model

Sub model	Parameter	Estimate	Standard error
Regression model	$\hat{\theta}_1$	−0.7504	0.0496
	$\hat{\theta}_2$	−0.1499	0.0410
	$\hat{\theta}_3$	−0.2913	0.0418
Association model	$\hat{\tau}_{12}^{(2,3)}=\hat{\tau}_{12}^{(3,2)}$	0.6734	0.0359
	$\hat{\tau}_{12}^{(2,4)}=\hat{\tau}_{12}^{(4,2)}$	0.2669	0.0204
	$\hat{\tau}_{12}^{(3,4)}=\hat{\tau}_{12}^{(4,3)}$	0.4465	0.0278
	$\hat{\tau}_{12}^{(4,4)}$	2.4027	0.0533
	$\hat{\tau}_{12}^{(3,3)}$	2.5891	0.0719
	$\hat{\tau}_{12}^{(2,2)}$	2.3026	0.0574

Table 2.17: Comparison of model fit for the males

Model	G^2	df	p-value
Marginal homogeneity	3.68	3	0.2982
Symmetry	**4.77**	**6**	**0.5736**

2.6.3 Pathologist dataset

The following dataset is an example of matched sets data. One area where the dependence ratio approach has not been applied to is rater agreement studies. Rater agreement is particularly applicable to this paper since, for example, it is typically of interest to test for marginal homogeneity between the raters. In addition, the dependence ratio approach also provides an assessment of the associations. This dataset has seven pathologists as the raters, each of which classified 118 slides according to the presence or absence of carcinoma of the uterine cervix. In other words, the response is binary and there are seven observations in a cluster. The dataset was first presented in Landis and Koch (1977). The dependence ratio approach is particularly applicable for this dataset since existing approaches using maximum likelihood estimation and based on a odds ratio parameterisation typically cannot cope with large cluster sizes such as this. Table 2.18 shows the data with 1 representing presence and 0 representing absence.

Table 2.18: Diagnosis of carcinoma by seven pathologists

Pathologist							
A	B	C	D	E	F	G	Count
0	0	0	0	0	0	0	34
0	0	0	0	1	0	0	2
0	1	0	0	0	0	0	6
0	1	0	0	0	0	1	1
0	1	0	0	1	0	0	4
0	1	0	0	1	0	1	5
1	0	0	0	0	0	0	2
1	0	1	0	1	0	1	1
1	1	0	0	0	0	0	2
1	1	0	0	0	0	1	1
1	1	0	0	1	0	0	2
1	1	0	0	1	0	1	7
1	1	0	0	1	1	1	1
1	1	0	1	0	0	1	1
1	1	0	1	1	0	1	2
1	1	0	1	1	1	1	3
1	1	1	0	1	0	1	13
1	1	1	0	1	1	1	5
1	1	1	1	1	0	1	10
1	1	1	1	1	1	1	16
							118

Since we have seven repeated responses, $2^7-7-1=120$ dependence ratios and 7 marginal probabilities are required to specify the complete joint distribution. The number of association parameters is therefore much larger than the datasets considered previously for square tables. Hence, substantial constraints need to be imposed on the dependence ratios to reduce the number of association parameters. In order to do this, the most appropriate approach is to consider the association structures described in Section 2.3

since these describe the association structure in a relevant way, using only a small number of parameters. The regression model in (2.36) with pathologist as the covariate effect will be combined with association structures from Section 2.3. The best fitting association structure will be taken forward to assess the fit of marginal homogeneity. Models such as complete symmetry will not be considered for this dataset but may be appropriate in other situations. Since the association structures from Section 2.3 are not nested, Akaike's Information Criterion (AIC) is used to compare the models as opposed to the likelihood ratio test. Table 2.19 shows the AIC for the association models considered. The Markov structures described in Section 2.3 are not appropriate for this dataset since there is no ordering to the repeated responses.

Table 2.19: Comparison of model fit for association structures using AIC, with pathologist covariate

Association Structure	AIC
Independence (I)	1062.93
Necessary factor (N)	772.519
Latent binary factor (L)	689.149

Of the 118 slides, 34 were rated zero by all seven pathologists (28.8%). This suggests that the necessary factor association structure N may be appropriate. In other words, the pathologist's classifications are conditionally independent given the necessary factor. Although N gave a superior fit compared to independence, the latent binary structure gives the best fit of the association structures considered. The latent binary factor within a necessary factor association structure (described in Section 1.2.3.1 of this thesis) produced negative profile probabilities, which is unsatisfactory. Given the latent binary factor association structure, a likelihood ratio test was used to compare the regression models with and without the pathologist covariate. Marginal homogeneity gave a poor fit (likelihood ratio test statistic = 170.426, degrees of freedom = 6, p-value < 0.001). Table 2.20 gives the parameter estimates for the chosen combined regression and association model (denoted PLB=pathologist latent binary), where the regression model is as described in (2.36) with pathologist as the covariate effect and the association model is the latent binary association structure (L).

The conclusion for the regression model is that pathologists 2, 5 and 7 are associated with higher probabilities (although not significantly) of presence than pathologist 1 (the baseline). In contrast, pathologists 3, 4 and 6 are associated with (significantly, at the 5% level) lower probabilities of presence than pathologist 1. For the association model, the parameter v_2 indicates that approximately 56% of observations (slides) are in latent group 1. Consequently, approximately 44% of observations are in the latent group 0. The κ parameter shows that the probability of presence in group 0 is approximately 9% of those in group 1. In other words, the latent binary factor is showing that there appears to be two groups, one that constitutes 44% of the population with a low presence

Table 2.20: Parameter estimates for model PLB

Sub model	Parameter	Value	Standard error	z-value
Regression model	intercept	0.2420	0.1780	1.3597
	pathologist 2	0.1547	0.0916	1.6898
	pathologist 3	−0.6544	0.1750	−3.7388
	pathologist 4	−1.1896	0.2100	−5.6634
	pathologist 5	0.1155	0.0990	1.1668
	pathologist 6	−1.5267	0.2337	−6.5336
	pathologist 7	0.1063	0.1012	1.0513
Association model	v_2	0.5649	0.0473	11.9533
	κ	0.0939	0.0204	4.5979

rate and the other that accounts for 56% of the population with a much higher presence rate. In order to come up with an explanation for this, groups of the population that satisfy these criteria will need to be found (Jokinen et al. 2006). Table 2.21 shows the estimated probability of presence for each pathologist conditional on the latent group. From Table 2.21, latent group 1 appears to contain slides where there is strong agreement between at least pathologists A, B , E and G on the presence of carcinoma. Latent group 0 appears to contain slides where the pathologists largely agree on the absence of carcinoma.

Table 2.21: Estimated presence probabilities conditional on the latent group

	Pathologist						
Estimated Probability	A	B	C	D	E	F	G
Pr(Presence $\mid L = 1$)	0.9248	0.9870	0.6576	0.4612	0.9714	0.3578	0.9677
Pr(Presence $\mid L = 0$)	0.0868	0.0926	0.0617	0.0433	0.0912	0.0336	0.0908

As a check of the adequacy of the model, the fitted marginal probabilities were compared to the observed marginal probabilities, for each pathologist. These are shown in Table 2.22. The conclusion to make from this is that the marginal probabilities are largely very similar with the exception of pathologist 2 who has a slightly lower fitted value than what was observed.

Table 2.22: Comparison of observed and fitted marginal probabilities for model PLB

Pathologist	Observed	Fitted
1	0.5593	0.5602
2	0.6695	0.5979
3	0.3814	0.3983
4	0.2712	0.2794
5	0.6017	0.5884
6	0.2119	0.2168
7	0.5593	0.5862

The association parameters of a latent variable structure, such as the latent binary, are not observable. As a consequence, there is no clear way to assess the adequacy of the association structure. Jokinen (2006: PhD Thesis) discusses the issue and concludes that all latent structures should be treated with caution and also notes that this issue is not just relevant to the dependence ratio approach.

This analysis has shown that the dependence ratio approach can be used to analyse rater agreement. The approach is particularly relevant to datasets with large cluster sizes that may not be feasible to analyse with a odds ratio parameterisation. Although only two latent classes were considered for the latent binary association structure in this analysis, the dependence ratio approach can also be extended to 3 or more latent classes. For the pathologist dataset, considering 3 latent classes led to negative fitted profile probabilities, an unsatisfactory fit.

2.7 Discussion

When analysing multivariate categorical data using population-averaged models and maximum likelihood, odds ratios are often used as the association measure. However, odds ratios suffer from the drawback of not being able to cope computationally with large cluster sizes. Although the GEE approach helped to solve this issue, the method focuses on the marginal regression with associations being treated as a nuisance. The rationale of the GEE approach is that it is difficult to model the higher-order moments (Diggle et al. 2002). However, the dependence ratio approach can cope computationally with larger cluster sizes.

The focus of this paper was on square tables and extensions to matched sets. Common models from the analysis of square tables using existing methods were discussed in terms of dependence ratios. Marginal homogeneity was discussed in detail. The unconstrained marginal homogeneity model in terms of dependence ratios was shown to be equivalent to combining a saturated association model with a marginal model (no explanatory variables). Marginal homogeneity can be still present when the association structure is not saturated such as in the symmetry model where additional constraints are imposed on the dependence ratios. Of course, for the simplest bivariate binary case, marginal homogeneity and symmetry are equivalent and no constraints on the association model are necessary. It was also shown that the symmetry model can be expressed directly in terms of dependence ratios for square tables. A convenient feature of the dependence ratio approach is that constraints can easily be imposed on the dependence ratios in order to reduce the number of parameters and aid the interpretability of the model. For example, although marginal homogeneity with a saturated association structure gave a satisfactory fit for the eye grade males dataset, additional constraints were still imposed on the dependence ratios to reduce the number of association parameters. This can also lead to more efficient estimates of the effects.

The secondary aim of the paper was to extend the common models for square tables to matched sets, using the dependence ratio approach. The dependence ratio approach offers a clear advantage for matched sets data since it can cope with large cluster sizes as opposed to the odds ratio approaches which typically encounter difficulties with maximum likelihood estimation. Marginal homogeneity and complete symmetry were discussed in detail. Complete symmetry will rarely give a satisfactory fit, although it was discussed because of its convenient formulation in terms of dependence ratios. Marginal homogeneity is typically of more interest to the researcher and is more likely to give a satisfactory fit. In contrast to square tables, determining the saturated association structure can be a difficult task. Consequently, a reasonable approach to follow (as demonstrated with the pathologist dataset in Section 2.6.3) is to firstly determine the best fitting association structure based on the AIC, since the models will typically not be nested. The best fitting association structure can then be taken forward to assess the

assumption of marginal homogeneity (if relevant). Once the most appropriate model is found, the regression model can be assessed by comparing the observed and marginal probabilities and the association model can be assessed if the model is not based on a latent structure. Some datasets may have additional covariate(s) to include. In such cases, marginal homogeneity can be assessed given the presence of the covariate(s).

Appendix

Proof of maximum likelihood estimates from Section 2.4.3.1

In Section 2.4.3.1, the maximum likelihood estimates ($\hat{\mu}$ and $\hat{\tau}_{12}$) for a bivariate binary response $\boldsymbol{Y} = (Y_1, Y_2)$ under marginal homogeneity were presented. The proof that these are the maximum likelihood estimates is now given.

Consider a sample of size n from $\boldsymbol{Y}=(Y_1, Y_2)$. This can be classified in a 2 by 2 contingency table with cell counts $(n_{00}, n_{01}, n_{10}, n_{11})$ and corresponding cell probabilities $(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$. Under the multinomial sampling scheme, the likelihood function for $(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$ is given by:

$$L(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}) \propto \pi_{00}^{n_{00}}\ \pi_{01}^{n_{01}}\ \pi_{10}^{n_{10}}\ \pi_{11}^{n_{11}}.$$

Consequently, the log-likelihood function can be expressed as:

$$l(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}) = n_{00} \log(\pi_{00}) + n_{01} \log(\pi_{01}) + n_{10} \log(\pi_{10}) + n_{11} \log(\pi_{11}) + c,$$

where $c = \log\left\{\frac{n!}{n_{00}!n_{01}!n_{10}!n_{11}!}\right\}$.

Table 2.2 gives the cell probabilities for the bivariate binary case in terms of μ and μ_{12}. The profile probabilities above can be replaced by these cell probabilities, as shown below:

$$l(\mu, \mu_{12}) = n_{00} \log(1 - 2\mu + \mu_{12}) + n_{01} \log(\mu - \mu_{12}) + n_{10} \log(\mu - \mu_{12}) + n_{11} \log(\mu_{12}).$$

In order to find the maximum likelihood estimates of μ and τ_{12}, the log-likelihood is differentiated with respect to μ and μ_{12} respectively since it is easier to work with μ_{12} for this proof. These are then set equal to zero.

$$\frac{\partial l}{\partial \mu} = \frac{n_{10} + n_{01}}{\mu - \mu_{12}} - \frac{2n_{00}}{1 - 2\mu + \mu_{12}},$$
$$\frac{\partial l}{\partial \mu_{12}} = \frac{n_{11}}{\mu_{12}} - \frac{n_{10} + n_{01}}{\mu - \mu_{12}} + \frac{n_{00}}{1 - 2\mu + \mu_{12}}.$$

However, rather than solving the above equations for μ and μ_{12}, it is easier to substitute the appropriate maximum likelihood estimates from Section 2.4.3.1 and show that the above equations are equal to zero. These maximum likelihood estimates came from prior belief but are proven to be correct by this method.

From the above:

$$\frac{n_{10}+n_{01}}{\hat{\mu}-\hat{\mu}_{12}} = \frac{n_{10}+n_{01}}{\frac{2n_{11}+n_{10}+n_{01}}{2n}-\frac{n_{11}}{n}},$$
$$= \frac{n_{10}+n_{01}}{\frac{n_{10}+n_{01}}{2n}} = 2n.$$

In addition:

$$\frac{n_{00}}{1-2\hat{\mu}+\hat{\mu}_{12}} = \frac{n_{00}}{1-\frac{(n_{10}+n_{01}+n_{11})}{n}},$$
$$= \frac{n(n_{00})}{n-n_{10}-n_{01}-n_{11}} = \frac{n(n-n_{10}-n_{01}-n_{11})}{n-n_{10}-n_{01}-n_{11}},$$
$$= n.$$

Consequently:

$$\frac{\partial l}{\partial \mu} = 2n-2n = 0,$$
$$\frac{\partial l}{\partial \mu_{12}} = \frac{n_{11}}{\frac{n_{11}}{n}} - 2n + n = n - 2n + n = 0.$$

Therefore proving that the maximum likelihood estimates of μ and μ_{12} are:

$$\hat{\mu} = \frac{\frac{n_{1+}}{n}+\frac{n_{+1}}{n}}{2} = \frac{n_{1+}+n_{+1}}{2n},$$
$$\hat{\mu}_{12} = \frac{n_{11}}{n}.$$

Therefore,

$$\hat{\tau} = \frac{\hat{\mu}_{12}}{\hat{\mu}^2} = \frac{\frac{n_{11}}{n}}{\frac{(n_{1+}+n_{+1})^2}{4n^2}} = \frac{4nn_{11}}{(2n_{11}+n_{10}+n_{01})^2}.$$

Therefore proving that the maximum likelihood estimates of μ and τ_{12} are as stated in (2.31) and (2.32).

Proof of the expected Fisher information for estimates from Section 2.4.3.1

The expected Fisher information for a bivariate binary response under marginal homogeneity is now found. For this case, the focus is on τ_{12} (denoted τ) as opposed to μ_{12}. Hence the expected Fisher Information is given by:

$$I_E = -E\begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \tau} \\ \frac{\partial^2 l}{\partial \tau \partial \mu} & \frac{\partial^2 l}{\partial \tau^2} \end{pmatrix}.$$

The log-likelihood with τ as opposed to μ_{12} is given by:

$$l(\mu,\tau) = n_{00}\log(1-2\mu+\tau\mu^2) + n_{01}\log(\mu-\tau\mu^2) + n_{10}\log(\mu-\tau\mu^2) + n_{11}\log(\tau\mu^2).$$

In order to prove the above, the first stage is to derive the first order derivatives with respect to μ and τ. These are given by:

$$\frac{\partial l}{\partial \mu} = \frac{2n_{11}}{\mu} + \frac{(1-2\mu\tau)(n_{10}+n_{01})}{\mu-\mu^2\tau} + \frac{(2\mu\tau-2)n_{00}}{1-2\mu+\mu^2\tau},$$
$$\frac{\partial l}{\partial \tau} = \frac{n_{11}}{\tau} - \frac{\mu(n_{01}+n_{10})}{1-\mu\tau} + \frac{\mu^2 n_{00}}{1-2\mu+\mu^2\tau}.$$

The second order derivatives are consequently given by:

$$\frac{\partial^2 l}{\partial \mu^2} = \frac{2\tau n_{00}}{(1-2\mu+\mu^2\tau)} - \frac{(2\mu\tau-2)^2 n_{00}}{(1-2\mu+\mu^2\tau)^2} - \frac{2\tau(n_{10}+n_{01})}{\mu(1-\mu\tau)} - \frac{(1-2\mu\tau)^2(n_{10}+n_{01})}{(\mu-\mu^2\tau)^2} - \frac{2n_{11}}{\mu^2},$$
$$\frac{\partial^2 l}{\partial \tau^2} = -\frac{n_{11}}{\tau^2} - \frac{(n_{10}+n_{01})\mu^2}{(1-\mu\tau)^2} - \frac{n_{00}\mu^4}{(1-2\mu+\mu^2\tau)^2},$$
$$\frac{\partial^2 l}{\partial \mu \partial \tau} = \frac{\partial^2 l}{\partial \tau \partial \mu} = \frac{2\mu(1-\mu)n_{00}}{(1-2\mu+\mu^2\tau)^2} - \frac{(n_{10}+n_{01})}{(1-\mu\tau)^2}.$$

The expected values for the above quantities are specified as:

$$E\left\{\frac{\partial^2 l}{\partial \mu^2}\right\} = \frac{2\tau E[n_{00}]}{(1-2\mu+\mu^2\tau)} - \frac{(2\mu\tau-2)^2 E[n_{00}]}{(1-2\mu+\mu^2\tau)^2} - \frac{2\tau E[(n_{10}+n_{01})]}{\mu(1-\mu\tau)} - \frac{(1-2\mu\tau)^2 E[n_{10}+n_{01}]}{(\mu-\mu^2\tau)^2} - \frac{2E[n_{11}]}{\mu^2},$$

$$E\left\{\frac{\partial^2 l}{\partial \tau^2}\right\} = -\frac{E[n_{11}]}{\tau^2} - \frac{\mu^2 E[n_{10}+n_{01}]}{(1-\mu\tau)^2} - \frac{\mu^4 E[n_{00}]}{(1-2\mu+\mu^2\tau)^2},$$

$$E\left\{\frac{\partial^2 l}{\partial \tau \partial \mu}\right\} = E\left\{\frac{\partial^2 l}{\partial \mu \partial \tau}\right\} = \frac{2\mu(1-\mu)E[n_{00}]}{(1-2\mu+\mu^2\tau)^2} - \frac{E[n_{10}+n_{01}]}{(1-\mu\tau)^2}.$$

From Tables 2.1 and 2.2 in Section 2.4, the expected values for each of the four cells of the table can be specified:

$$E[n_{11}] = \tau\mu^2,\ E[n_{10}] = E[n_{01}] = \mu(1-\mu\tau),\ E[n_{00}] = 1 - 2\mu + \mu^2\tau.$$

Consequently:

$$\begin{aligned}
E\left\{\frac{\partial^2 l}{\partial \mu^2}\right\} &= -4\tau - \frac{(2\mu\tau - 2)^2}{1 - 2\mu + \mu^2\tau} - \frac{2(1-2\mu\tau)^2}{\mu(1-\mu\tau)},\\
E\left\{\frac{\partial^2 l}{\partial \tau^2}\right\} &= -\frac{\mu^2}{\tau} - \frac{2\mu^3}{(1-\mu\tau)} - \frac{\mu^4}{(1-2\mu+\mu^2\tau)},\\
E\left\{\frac{\partial^2 l}{\partial \tau \partial \mu}\right\} &= E\left\{\frac{d^2 l}{d\mu d\tau}\right\} = \frac{2\mu(1-\mu)}{1-2\mu+\mu^2\tau} - \frac{2\mu}{1-\mu\tau}.
\end{aligned}$$

The expected Fisher information for a bivariate binary response under marginal homogeneity is therefore given by:

$$I_E = \begin{pmatrix} 4\tau + \frac{(2\mu\tau-2)^2}{1-2\mu+\mu^2\tau} + \frac{2(1-2\mu\tau)^2}{\mu(1-\mu\tau)} & \frac{2\mu}{1-\mu\tau} - \frac{2\mu(1-\mu)}{1-2\mu+\mu^2\tau} \\ \frac{2\mu}{1-\mu\tau} - \frac{2\mu(1-\mu)}{1-2\mu+\mu^2\tau} & \frac{\mu^2}{\tau} + \frac{2\mu^3}{(1-\mu\tau)} + \frac{\mu^4}{(1-2\mu+\mu^2\tau)} \end{pmatrix}.$$

Proof of the expected Fisher information for the saturated case

For completeness, the expected Fisher information for the saturated bivariate binary case is now derived. There is one additional parameter to the marginal homogeneity case since the two marginal probabilities are no longer constrained to be equal. The cell probabilities for this case are shown below:

Cell probabilities for a bivariate binary response

	Y_2		
Y_1	1	0	Total
1	$\tau\mu_1\mu_2$	$\mu_1 - \tau\mu_1\mu_2$	μ_1
0	$\mu_2 - \tau\mu_1\mu_2$	$1 - \mu_1 - \mu_2 + \tau\mu_1\mu_2$	$1 - \mu_1$
Total	μ_2	$1 - \mu_2$	1

The log-likelihood is given by:

$$l(\mu_1, \mu_2, \tau) = n_{00}\log(1 - \mu_1 - \mu_2 + \tau\mu_1\mu_2) + n_{10}\log(\mu_1 - \tau\mu_1\mu_2) + n_{01}\log(\mu_2 - \tau\mu_1\mu_2) + n_{11}\log(\tau\mu_1\mu_2).$$

Consequently:

$$\begin{aligned}
\frac{\partial l}{\partial \mu_1} =& \frac{(-1+\tau\mu_2)n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)} + \frac{(1-\tau\mu_2)n_{10}}{\mu_1-\tau\mu_1\mu_2} - \frac{(\tau\mu_2)n_{01}}{\mu_2-\tau\mu_1\mu_2} + \frac{n_{11}}{\mu_1},\\
\frac{\partial l}{\partial \mu_2} =& \frac{(-1+\tau\mu_1)n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)} - \frac{(\tau\mu_1)n_{10}}{\mu_1-\tau\mu_1\mu_2} + \frac{(1-\tau\mu_1)n_{01}}{\mu_2-\tau\mu_1\mu_2} + \frac{n_{11}}{\mu_2},\\
\frac{\partial l}{\partial \tau} =& \frac{(\mu_1\mu_2)n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)} - \frac{(\mu_1\mu_2)n_{10}}{\mu_1-\tau\mu_1\mu_2} - \frac{(\mu_1\mu_2)n_{01}}{\mu_2-\tau\mu_1\mu_2} + \frac{n_{11}}{\tau},\\
\frac{\partial^2 l}{\partial \mu_1^2} =& \frac{-(\tau\mu_2-1)^2 n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)^2} - \frac{(1-\tau\mu_2)^2 n_{10}}{(\mu_1-\tau\mu_1\mu_2)^2} - \frac{(-\tau\mu_2)^2 n_{01}}{(\mu_2-\tau\mu_1\mu_2)^2} - \frac{n_{11}}{(\mu_1)^2},\\
\frac{\partial^2 l}{\partial \mu_2^2} =& \frac{-(\tau\mu_1-1)^2 n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)^2} - \frac{(-\tau\mu_1)^2 n_{10}}{(\mu_1-\tau\mu_1\mu_2)^2} - \frac{(1-\tau\mu_1)^2 n_{01}}{(\mu_2-\tau\mu_1\mu_2)^2} - \frac{n_{11}}{(\mu_2)^2},\\
\frac{\partial^2 l}{\partial \tau^2} =& \frac{-(\mu_1\mu_2)^2 n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)^2} - \frac{(-\mu_1\mu_2)^2 n_{10}}{(\mu_1-\tau\mu_1\mu_2)^2} - \frac{(-\mu_1\mu_2)^2 n_{01}}{(\mu_2-\tau\mu_1\mu_2)^2} - \frac{n_{11}}{(\tau)^2}.
\end{aligned}$$

The second-order mixed derivatives are given by:

$$\begin{aligned}
\frac{\partial^2 l}{\partial \mu_1 \partial \mu_2} = \frac{\partial^2 l}{\partial \mu_2 \partial \mu_1} =& \frac{(\tau)n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)} - \frac{[(\tau\mu_2-1)(\tau\mu_1-1)]n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)^2} - \frac{(\tau)n_{10}}{\mu_1-\tau\mu_1\mu_2}\\
&+ \frac{(1-\tau\mu_2)(\tau\mu_1)n_{10}}{(\mu_1-\tau\mu_1\mu_2)^2} - \frac{(\tau)n_{01}}{(\mu_2-\tau\mu_1\mu_2)} + \frac{(1-\tau\mu_1)(\tau\mu_2)n_{01}}{(\mu_2-\tau\mu_1\mu_2)^2},\\
\frac{\partial^2 l}{\partial \tau \partial \mu_1} = \frac{\partial^2 l}{\partial \mu_1 \partial \tau} =& \frac{(\mu_2)n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)} - \frac{[(\tau\mu_2-1)(\mu_1\mu_2)]n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)^2} - \frac{(\mu_2)n_{10}}{\mu_1-\tau\mu_1\mu_2}\\
&+ \frac{(1-\tau\mu_2)(\mu_1\mu_2)n_{10}}{(\mu_1-\tau\mu_1\mu_2)^2} - \frac{(\mu_2)n_{01}}{(\mu_2-\tau\mu_1\mu_2)} - \frac{(\mu_1\mu_2)(\tau\mu_2)n_{01}}{(\mu_2-\tau\mu_1\mu_2)^2},\\
\frac{\partial^2 l}{\partial \tau \partial \mu_2} = \frac{\partial^2 l}{\partial \mu_2 \partial \tau} =& \frac{(\mu_1)n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)} - \frac{[(\tau\mu_1-1)(\mu_1\mu_2)]n_{00}}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)^2} - \frac{(\mu_1)n_{10}}{\mu_1-\tau\mu_1\mu_2}\\
&- \frac{(\tau\mu_1)(\mu_1\mu_2)n_{10}}{(\mu_1-\tau\mu_1\mu_2)^2} - \frac{(\mu_1)n_{01}}{(\mu_2-\tau\mu_1\mu_2)} + \frac{(\mu_1\mu_2)(1-\tau\mu_1)n_{01}}{(\mu_2-\tau\mu_1\mu_2)^2}.
\end{aligned}$$

In addition:

$$
\begin{aligned}
E[n_{00}] =& 1-\mu_1-\mu_2+\tau\mu_1\mu_2,\\
E[n_{10}] =& \mu_1-\tau\mu_1\mu_2,\\
E[n_{01}] =& \mu_2-\tau\mu_1\mu_2,\\
E[n_{11}] =& \tau\mu_1\mu_2.
\end{aligned}
$$

The expected values of the second-order derivatives are given by:

$$
\begin{aligned}
E\left\{\frac{\partial^2 l}{\partial\mu_1^2}\right\} =& -\frac{(\tau\mu_2-1)^2}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)}-\frac{(1-\tau\mu_2)^2}{(\mu_1-\tau\mu_1\mu_2)}-\frac{(-\tau\mu_2)^2}{(\mu_2-\tau\mu_1\mu_2)}-\frac{\tau\mu_2}{\mu_1},\\
E\left\{\frac{\partial^2 l}{\partial\mu_2^2}\right\} =& -\frac{(\tau\mu_1-1)^2}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)}-\frac{(-\tau\mu_1)^2}{(\mu_1-\tau\mu_1\mu_2)}-\frac{(1-\tau\mu_1)^2}{(\mu_2-\tau\mu_1\mu_2)}-\frac{\tau\mu_1}{\mu_2},\\
E\left\{\frac{\partial^2 l}{\partial\tau^2}\right\} =& -\frac{(\mu_1\mu_2)^2}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)}-\frac{(-\mu_1\mu_2)^2}{(\mu_1-\tau\mu_1\mu_2)}-\frac{(-\mu_1\mu_2)^2}{(\mu_2-\tau\mu_1\mu_2)}-\frac{\mu_1\mu_2}{\tau},\\
E\left\{\frac{\partial^2 l}{\partial\mu_1\partial\mu_2}\right\} =& E[\frac{\partial^2 l}{\partial\mu_2\partial\mu_1}] = -\frac{(\tau\mu_2-1)(\tau\mu_1-1)}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)}-\frac{(1-\tau\mu_2)(-\tau\mu_1)}{\mu_1-\tau\mu_1\mu_2}-\frac{(1-\tau\mu_1)(-\tau\mu_2)}{\mu_2-\tau\mu_1\mu_2}-\tau,\\
E\left\{\frac{\partial^2 l}{\partial\mu_1\partial\tau}\right\} =& E[\frac{\partial^2 l}{\partial\tau\partial\mu_1}] = -\frac{(\tau\mu_2-1)(\mu_1\mu_2)}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)}-\frac{(1-\tau\mu_2)(-\mu_1\mu_2)}{\mu_1-\tau\mu_1\mu_2}-\frac{(-\mu_1\mu_2)(-\tau\mu_2)}{\mu_2-\tau\mu_1\mu_2}-\mu_2,\\
E\left\{\frac{\partial^2 l}{\partial\mu_2\partial\tau}\right\} =& E[\frac{\partial^2 l}{\partial\tau\partial\mu_2}] = -\frac{(\tau\mu_1-1)(\mu_1\mu_2)}{(1-\mu_1-\mu_2+\tau\mu_1\mu_2)}-\frac{(-\tau\mu_1)(-\mu_1\mu_2)}{\mu_1-\tau\mu_1\mu_2}-\frac{(1-\tau\mu_1)(-\mu_1\mu_2)}{\mu_2-\tau\mu_1\mu_2}-\mu_1.
\end{aligned}
$$

The expected Fisher information for the saturated bivariate binary case is given by:

$$I_E = -E \begin{pmatrix} \frac{\partial^2 l}{\partial \mu_1^2} & \frac{\partial^2 l}{\partial \mu_1 \partial \mu_2} & \frac{\partial^2 l}{\partial \mu_1 \partial \tau} \\ \frac{\partial^2 l}{\partial \mu_2 \partial \mu_1} & \frac{\partial^2 l}{\partial \mu_2^2} & \frac{\partial^2 l}{\partial \mu_2 \partial \tau} \\ \frac{\partial^2 l}{\partial \tau \partial \mu_1} & \frac{\partial^2 l}{\partial \tau \partial \mu_2} & \frac{\partial^2 l}{\partial \tau^2} \end{pmatrix},$$

where the expectations of the partial second derivatives are given above. Interestingly, by substituting $\mu_1 = \mu_2 = \mu$ in to $\frac{\partial^2 l}{\partial \tau^2}$ from the saturated case, $\frac{\partial^2 l}{\partial \tau^2}$ from the marginal homogeneity case is obtained.

R code

In order to calculate the standard errors of the marginal probabilities in the bilirubin dataset, firstly the Fisher information matrix was parameterised in terms of μ_1, μ_2 and τ, as shown previously in the Appendix. This matrix can then be inverted to find the variance-covariance matrix and consequently the standard errors. The R code below shows this for both the saturated and marginal homogeneity cases.

```
The following R function calculates the expected Fisher information
for the saturated case and the variance-covariance matrix.

#Expected Fisher information for 3 by 3 table for the saturated case,
# parameters mu1, mu2 and tau, sample of size n.

VarCovSaturated<-function(mu1hat,mu2hat,tauhat,n){
# part1
p1<-(tauhat*mu2hat)-1
# part2
p2<-1-mu1hat-mu2hat+(tauhat*mu1hat*mu2hat)
# part3
p3<-1-(tauhat*mu2hat)
# part4
p4<-mu1hat-(tauhat*mu1hat*mu2hat)
# part5
p5<-tauhat*mu2hat
# part6
p6<-mu2hat-(tauhat*mu1hat*mu2hat)
# part7
p7<-(tauhat*mu1hat)-1
# part8
p8<-tauhat*mu1hat
# part 9
p9<-1-(tauhat*mu1hat)
# part 10
p10<-mu1hat*mu2hat
# Calculating the negative expected value of the second-order derivative
# of the log-likelihood with respect to mu1
a<-((n*(p1^{2}))/(p2))+((n*(p3^{2}))/(p4))+((n*(p5^{2}))/(p6))+(n*p5)/(mu1hat)
# Calculating the negative expected value of the second-order derivative
# of the log-likelihood with respect to mu2
e<-((n*(p7^{2}))/(p2))+((n*(p8^{2}))/(p4))+((n*(p9^{2}))/(p6))+((n*p8)/(mu2hat))
# Calculating the negative expected value of the second-order derivative
# of the log-likelihood with respect to tau
i<-((n*(p10^{2}))/(p2))+((n*(p10^{2}))/(p4))+((n*(p10^{2}))/(p6))+((n*p10)/(tauhat))
# Calculating the negative expected value of the second-order mixed derivative
# of the log-likelihood with respect to mu1 and mu2
b<-((n*p1*p7)/(p2))-((n*p8*p3)/(p4))+(n*tauhat)-((n*p9*p5)/(p6))
# Calculating the negative expected value of the second-order mixed derivative
# of the log-likelihood with respect to mu1 and tau
c<-((n*p1*p10)/(p2))-((n*p10*p3)/(p4))+(n*mu2hat)+((n*p10*p5)/(p6))
```

```
# Calculating the negative expected value of the second-order mixed derivative
# of the log-likelihood with respect to mu2 and tau
f<-((n*p7*p10)/(p2))+((n*p10*p8)/(p4))+(n*mu1hat)-((n*p10*p9)/(p6))
FisherInformation<-matrix(c(a,b,c,b,e,f,c,f,i),nrow=3,ncol=3,byrow = TRUE)
VarCovMatrix<-solve(FisherInformation)
return(VarCovMatrix)
}

# Specifying the parameter values for the bilirubin dataset saturated model.
VarCovSat<-VarCovSaturated(12/86,20/86,2.15,86)

#Standard error of mu1hat
sqrt(VarCovSat[1])
#Standard error of mu2hat
sqrt(VarCovSat[5])
#Standard errof of tauhat
sqrt(VarCovSat[9])

# The following R function calculates the expected Fisher information
# for the marginal homogeneity case and the variance-covariance matrix.

# Expected Fisher Information for 2 by 2 table under marginal homogeneity,
# parameters mu and tau, sample of size n.

VarCovMH<-function(muhat,tauhat,n){
# Calculating the negative expected value of the second-order derivative
# of the log-likelihood  with respect to tau
d<-(n*(muhat)^4)/(1-2*muhat+(tauhat*(muhat^2)))+((2*n*(muhat^3))/(1-muhat*tauhat))
+((n*(muhat^2))/(tauhat))
# Calculating the negative expected value of the second-order derivative
# of the log-likelihood with respect to mu
a<-(n*4*tauhat)+((n*((2*muhat*tauhat)-2)^2)/(1-2*muhat+tauhat*(muhat^2)))
+((2*n*(1-2*muhat*tauhat)^2)/(muhat*(1-muhat*tauhat)))
# Calculating the negative expected value of the second-order mixed derivative
# of the log-likelihood  with respect to mu and tau
b<-((-2*n*muhat)*(1-muhat))/(1-2*muhat+(tauhat*(muhat^2)))
+((2*n*muhat)/(1-muhat*tauhat))
# Specifying the Fisher information matrix
FisherInformation<-matrix(c(a,b,b,d),nrow=2,ncol=2,byrow = TRUE)
#Inverting the Fisher information matrix
VarCovMatrix<-solve(FisherInformation)
return(VarCovMatrix)
}

Specifying the parameter values for the bilirubin dataset marginal homogeneity model.
VarCov<-VarCovMH(16/86,129/64,86)

#Standard Errror of muhat
sqrt(VarCov[1])
#Standard Errror of tauhat
sqrt(VarCov[4])
```

Chapter 3

Paper 2: Applications of the dependence ratio to psoriatic arthritis

3.1 Introduction

Arthritis is a common chronic condition that results in inflammation and pain within a joint and in some cases permanent joint damage. The condition often causes a great deal of suffering to those affected so a better understanding of the disease process is of huge benefit. The focus of the current paper is on psoriatic arthritis in the hand joints (excluding the wrist). Psoriatic arthritis is an inflammatory arthritis which is associated with the skin condition psoriasis and is commonly found in the hand joints. However, only some individuals with psoriasis will develop psoriatic arthritis. Both males and females are equally likely to be affected (Gladman et al. 2005; O'Keeffe et al. 2011).

Data for this paper were jointly provided by the MRC Biostatistics Unit in Cambridge and the Toronto Psoriatic Arthritis Clinic. The clinic was established in 1978 by Professor Dafna Gladman and contains the largest cohort of patients with psoriatic arthritis in the world. A detailed explanation of the data used is given in Section 3.1.2.

Although not of direct interest for this paper, another common form of arthritis that often affects the hand joints is rheumatoid arthritis. It is discussed due to its contrasting features compared to psoriatic arthritis. Rheumatoid arthritis is an autoimmune disease since the body's immune system attacks the joints, leading to inflammation and pain. In contrast to psoriatic arthritis, rheumatoid arthritis more commonly affects females than males (Gladman et al. 2005; O'Keeffe et al. 2011). Although both conditions can be difficult to diagnose, individuals with psoriatic arthritis will typically test negative for rheumatoid factor (a blood test for rheumatoid arthritis), whereas individuals with rheumatoid arthritis will typically test positive for rheumatoid factor.

Figure 3.1 shows a diagram of the 14 joints in a hand. There are therefore 28 individual joints across the two hands. The numbering assigned to the joints in Figure 3.1 (1-14) corresponds to the right hand throughout the paper. Joint numbers 15-28 are used to

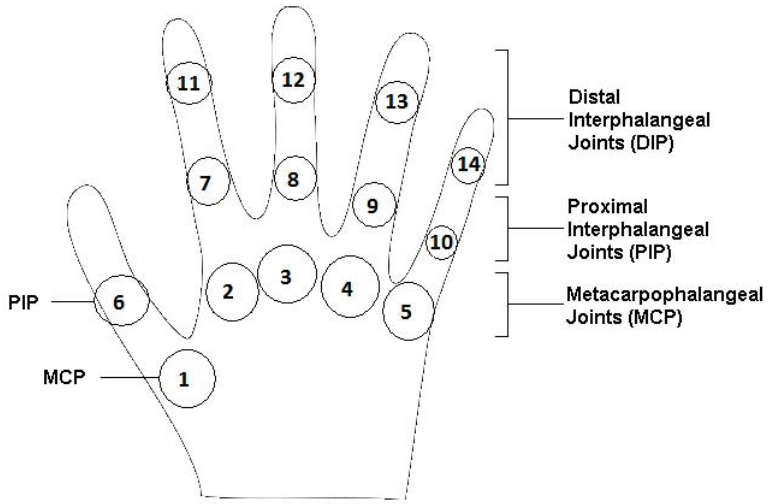

Figure 3.1: Location of the hand joints (Outline of hand reprinted with permission from Aidan O'Keeffe)

refer to the corresponding joints in the left hand. For example, joints 1 and 15 refer to the metacarpophalangeal (MCP) thumb joint on the right and left hands respectively.

The joints within a hand can be split into three different types, as shown in Figure 3.1. The distal interphalangeal (DIP) joints are the closest to the fingertips and are present in all digits except the thumb. They are commonly affected in psoriatic arthritis (Gladman et al. 2005).

The proximal interphalangeal (PIP) joints and the MCP joints are present in all digits. The MCP joints (excluding the thumb), commonly known as knuckles, are the largest joints. Table 3.1 shows a breakdown of joint numbers by hand, joint type and digit.

The patterns of affected joints can be used to determine which of psoriatic arthritis and rheumatoid arthritis an individual is more likely to have. The DIP joints (closest to the fingernails) are commonly affected in psoriatic arthritis. In contrast, the MCP and PIP joints are more commonly affected in rheumatoid arthritis (Gladman et al. 2006). The joints in patients with rheumatoid arthritis are often affected in a symmetrical pattern (Gladman et al. 2005). In other words, the same joints in each hand are affected.

In contrast, patients with psoriatic arthritis often have an asymmetrical pattern (different joints affected in each hand). In particular, asymmetrical patterns often exist with regards to the DIP joints (closest to the fingertips) and the larger MCP joints (O'Keeffe et al. 2011). In addition, individuals with psoriatic arthritis will often have their joints affected in a ray pattern such that joints of the same digit are affected (Gladman et al. 2005).

Table 3.1: Joint numbers by hand, joint type and digit

Right hand	Left hand	Joint type	Digit
1	15	MCP	thumb
2	16	MCP	index
3	17	MCP	middle
4	18	MCP	ring
5	19	MCP	little
6	20	PIP	thumb
7	21	PIP	index
8	22	PIP	middle
9	23	PIP	ring
10	24	PIP	little
11	25	DIP	index
12	26	DIP	middle
13	27	DIP	ring
14	28	DIP	little

It should be noted that once a hand joint is considered to be affected with psoriatic arthritis (following diagnosis), it can then reach two further states (as defined in O'Keeffe et al. 2011): disease activity and clinical damage. Disease activity is a reversible process and is defined on a binary scale such that a joint is described as active if it is either tender (stress pain and/or tenderness) or effused (swelling and/or tenderness). In contrast, clinical damage is an irreversible process meaning that once a joint is damaged, it will remain damaged. It is also defined on a binary scale with a joint defined as clinically damaged if there is a limitation in its movement of more than 20% of the range not related to joint swelling (Gladman et al. 1990; Cresswell and Farewell, 2010). It is also considered damaged if it ankylosed (Siannis et al. 2006), deformed, flail or if the joint is a replacement (Cresswell and Farewell, 2010). Note that a joint can reach the clinically damaged state without ever having disease activity present. Once a patient is diagnosed with psoriatic arthritis (it will usually only affect a small subset of joints at any time), every effort should be made to find a relevant treatment that reduces the possibility of permanent joint damage in the future.

Only clinical damage in patients with psoriatic arthritis is considered in this analysis. This is performed using the dependence ratio association measure (Ekholm, Smith and McDonald, 1995). The dependence ratio is an advantageous alternative to the odds ratio for analysing multivariate categorical data (such as the psoriatic arthritis data described in Section 3.1.2), using maximum likelihood based population-averaged models. Population-averaged refers to the fact that the regression coefficients are averaged over the whole population. In contrast, conditional (or subject-specific) approaches estimate subject-specific effects. The dependence ratio approach puts equal emphasis on the marginal regression and the associations within a cluster (patient) as opposed to the

popular generalised estimating equations (GEE) approach which concentrates on the marginal regression with the associations of secondary interest. The current paper consequently has two research questions. Firstly, which explanatory variables significantly predict clinical damage? Secondly, what are the association patterns between joints? The dependence ratio approach is discussed in detail in Section 3.2.

3.1.1 Previous research on psoriatic arthritis in the hand joints

Clinical damage is typically the measure of primary interest, with previous research on psoriatic arthritis in the hand joints (such as Cresswell and Farewell, 2010; O'Keeffe et al. 2011) largely concentrating on symmetrical damage. In the context of arthritis of the hand joints, the term symmetry is used to refer to the same joints on each hand being damaged. The relationship between disease activity and clinical damage has also been considered in previous research (for both psoriatic arthritis and rheumatoid arthritis), with evidence found of an association between disease activity and progression to clinical damage, regardless of the activity measure used (such as Gladman et al. 1995; O'Keeffe et al. 2011). For example, the total number of tender/effused joints may be used as a measure of disease activity or alternatively biochemical markers (O'Keeffe et al. 2011).

The work of this paper is similar to that of Cresswell and Farewell (2010) in the sense that only clinical damage is considered. In addition, only the last clinic visit for each patient, for each joint is analysed (as with some of the methods in Cresswell and Farewell). This is a reasonable approach when only clinical damage is considered since the process is irreversible. The dependence ratio approach adopted here focuses on looking for association patterns in the damaged joints, both symmetrical and non-symmetrical. Although symmetry is often of primary interest, determining other patterns of clinical damage such as asymmetrical tendencies is also useful, particularly given that psoriatic arthritis in the hands is recognised for demonstrating asymmetrical patterns. In contrast to the work of O'Keeffe et al. (2011), longitudinal patterns of damaged joints over time and disease activity are not considered in this paper.

Since clinical damage is being considered in the current paper as opposed to simply whether the joint is affected with psoriatic arthritis or not (following diagnosis), patterns of the clinically damaged joints can consequently differ from those described previously for affected joints. For example, although males and females are equally likely to be affected with psoriatic arthritis, females have been found to progress to clinical damage quicker (Gladman et al. 2005). In addition, although at least 50% of patients with psoriatic arthritis have DIP joint(s) affected (Gladman, 2006), the percentage is much lower when clinical damage is considered (as shown in Section 3.6).

There is strong agreement among rheumatologists that clinical damage in patients with rheumatoid arthritis is highly symmetric (Gladman et al. 2005; O'Keeffe et al. 2011).

Recent studies have demonstrated that symmetrical patterns of clinical damage do also exist in the hand joints of individuals with psoriatic arthritis (Cresswell and Farewell, 2010; O'Keeffe et al. 2011). However, there is some debate among rheumatologists with regards to the extent of this symmetry (Gladman et al. 2005; Gladman et al. 2006) since symmetry is generally regarded as less common in psoriatic arthritis than in rheumatoid arthritis (Gladman et al. 2006).

Work in the last few years has helped to dispute this claim. At this point, it is worth discussing measures of symmetric damage that have been proposed in recent years for psoriatic arthritis. Cresswell and Farewell (2010) considered a number of different methods, all of which demonstrated evidence of symmetric damage in psoriatic arthritis. For example, one method is based upon the definition of symmetry from Helliwell et al. (2000) which defines symmetric damage as present if 50% or more of the damaged joints form symmetric pairs (a particular joint and its corresponding joint on the other hand are damaged). This is done separately for each possible number of damaged joints (N) with the observed proportion of patients that meet Helliwell's definition being compared to an expected proportion. The expected proportions assume that joints progress independently to damage. Symmetrical damage is therefore indicated by greater observed proportions than expected. Although no formal significance test is presented, Cresswell and Farewell noted that observed proportions are consistently higher than expected proportions, with the exception of $N = 9$. In the present paper, symmetric damage is assessed using the observed dependence ratios (see Section 3.5).

A notable paper by Bond et al. (2007) assessed which variables significantly predict the progression to clinical damage in the hand and feet joints. The response of interest was the increase in the total number of clinically damaged joints between clinic visits with a negative binomial regression model used. The GEE approach was used in order to take into account the within-patient correlation and provide robust standard errors. Sex was found to be non-significant whereas the initial ESR (erythrocyte sedimentation rate) value was found to significantly predict progression to clinical damage with a low value protective of damage. The ESR is a test that is used to indirectly measure inflammation in the body. In addition, time in the clinic was found to significantly predict progression to clinical damage.

Bond et al. (2007) concluded that the same predictors of clinical damage are seen regardless of how damage is detected. Since the analysis by Bond et al. (2007) included the feet joints (which are not of interest in this analysis), it is of interest to see if similar conclusions are obtained for the analyses in this paper (using the GEE approach in Section 3.4 and the dependence ratio approach in Section 3.6). These analyses use the binary formulation of clinical damage (as defined in Section 3.1) so a logistic regression model is used instead for the marginal regression in future sections. In addition, Bond et al. (2007) considered longitudinal patterns over time as opposed to the last clinic visit. However, the explanatory variables considered in the current analysis reflect the

fact that only the last clinic visit is used. For example, disease activity is not relevant to include in this paper since it is a reversible process over time so including it as an explanatory variable for the last clinic visit is inappropriate. It is also worth noting that the GEE analysis conducted by Bond et al. (2007) has the association structure as a secondary interest so the dependence ratio approach provided in this paper provides a more thorough analysis.

3.1.2 Data used

Data for 386 patients who enrolled in the Toronto Psoriatic Arthritis Clinic between 1978 and 2000 were made available. As part of the consent process, the dataset will not be reproduced in the current paper. Of the 386 patients, there were 214 males (55.4%) and 172 females (44.6%). The mean age at entry to the clinic was 51 years and 5 months whereas the mean age of diagnosis with psoriatic arthritis was 34 years and 10 months. The current paper only considers the last clinic visit of each patient (for each each joint) with the number of visits per patient varying from 2 to 47. Each of the 386 patients therefore gives 28 responses. Table 3.2 shows the gender patterns by age, where age refers to when the patient was diagnosed with psoriatic arthritis.

Table 3.2: Age of diagnosis and gender distributions of the Toronto PsA dataset

	Gender		
Age Group	Females (%)	Males (%)	Total (%)
10-19	22 (12.8%)	10 (4.70%)	32 (8.30%)
20-29	46 (26.7%)	63 (29.4%)	109 (28.2%)
30-39	43 (25.0%)	71 (33.2%)	114 (29.5%)
40-49	38 (22.1%)	42 (19.6%)	80 (20.7%)
50+	23 (13.4%)	28 (13.1%)	51 (13.2%)
Total	172	214	386

Table 3.2 shows that the diagnosis of psoriatic arthritis is less common in the youngest and oldest age groups (in both males and females). This is not surprising since the onset of psoriatic arthritis is most common in middle age. Gladman et al. (2005) state that the onset is most common in those in their fourth decade of life.

In order to make the patients comparable, all patients in the dataset provided had no clinical damage in any of their 28 joints at entry to the clinic. In addition, only patients with a minimum of two clinic visits were included. The patients were assessed at regular intervals (approximately every 6-12 months) during this time with each of their joints being assessed for clinical damage at each visit. There were no missing values. Recent investigations using the data (Cresswell and Farewell, 2010; O'Keeffe et al. 2011) had a larger cohort of patients since they use data up to 2006. Data from 2001 to 2006 were

not available for the current paper. Explanatory variables were available such as the age that the patient was diagnosed with psoriatic arthritis and gender. The effect of these variables on clinical damage (the response) will be assessed in subsequent sections.

The structure of the paper is as follows. In Section 3.2, the dependence ratio approach is discussed in detail. In Section 3.3, some initial exploratory analysis of the marginal patterns of clinical damage is conducted, irrespective of the associations. This is then built on in Section 3.4 which contains a detailed GEE analysis in order to assess in more detail the marginal patterns of clinical damage. Exploratory analysis of the association patterns of clinical damage using the observed dependence ratios is then performed in Section 3.5. Using information from Section 3.5, Section 3.6 uses the dependence ratio approach such that the association structure is taken into account. Section 3.7 concludes the paper.

Previous work that has looked at clinical damage in patients with psoriatic arthritis has largely focused on symmetrical patterns. In the current paper, association patterns within a hand as well as asymmetrical patterns are also considered. It is also of interest to determine how the patterns of damaged joints in this analysis differ from the patterns described in the literature for joints that are simply affected by psoriatic arthritis. It should be noted that the main aim of this paper is to advocate the use of the dependence ratio approach for analysing multivariate categorical data, such as the psoriatic arthritis data, rather than determining any key medical conclusions.

3.2 The dependence ratio approach

Due to the correlated nature of multivariate categorical data, standard univariate regression techniques cannot be used because they assume independence. However, by treating all the patients responses within a cluster as the response, multivariate techniques can be used to overcome this issue. This paper focuses on the use of population-averaged based models, as opposed to conditional (or subject-specific) models, for analysing the psoriatic arthritis data outlined in Section 3.1.2, using maximum likelihood estimation. The focus of this paper is on multivariate binary data since the response of interest (clinical damage) has only two categories.

The analysis of a multivariate categorical response is more complicated than the traditional multivariate normal response. This is because, unlike the multivariate normal distribution, the first two moments do not specify the complete joint distribution of a multivariate categorical response (unless there are only two responses in a cluster). Consequently, constraints will often need to be imposed on the higher-order moments.

In terms of taking into account the associations within a cluster (or patient in the psoriatic arthritis data), the odds ratio is typically used for population-averaged based

maximum likelihood estimation. Although the odds ratio has some convenient properties, such as the orthogonality between the regression and association parameters, it suffers from the inability to cope computationally with cluster sizes larger than approximately five (Lesaffre et al. 2000). The quasi-likelihood GEE approach helps to deal with this issue. However, this approach focuses on the marginal regression and treats the associations as a secondary interest since moments of higher-order than two are not specified. Lindsey and Lambert (1998) argued that determining the mechanisms that generate the associations within a cluster is vital to obtain a complete analysis of the data. This is particularly important if interest lies in the associations, such as the psoriatic arthritis data in this paper. As a consequence, the dependence ratio was proposed (Ekholm et al. 1995) as an alternative to the odds ratio for population-averaged based maximum likelihood estimation. As well as being able to cope computationally with larger cluster sizes than the odds ratio, the dependence ratio also allows for relevant association structures to be imposed, which is not available with the GEE approach.

In order to demonstrate the dependence ratio, the simplest bivariate binary case is considered ($\boldsymbol{Y} = (Y_1, Y_2)$). For this case, there are four possible response profiles: (1,1), (0,0), (1,0) and (0,1) with 1 regarded as the response of interest and 0 as the baseline. For this case, the dependence ratio is defined as:

$$\tau_{12} = \frac{\mu_{12}}{\mu_1 \mu_2} = \frac{pr(Y_1 = 1, Y_2 = 1)}{pr(Y_1 = 1)\, pr(Y_2 = 1)}. \tag{3.1}$$

The dependence ratio for a bivariate binary response is defined as the joint success probability divided by the joint success probability assuming independence. A value of 1 indicates independence, whereas values less than 1 and greater than 1 represent negative and positive associations respectively. In the context of the psoriatic arthritis data, a two-way dependence ratio refers to the probability of a patient having both joints damaged divided by the joint probability of damage under independence. Higher-order dependence ratios as well as dependence ratios for when there are more than two response categories can be expressed similarly and are discussed in detail in paper 1 of this thesis. The third-order dependence ratio for a binary response is given by:

$$\tau_{123} = \frac{\mu_{123}}{\mu_1 \mu_2 \mu_3} = \frac{pr(Y_1 = 1, Y_2 = 1, Y_3 = 1)}{pr(Y_1 = 1)\, pr(Y_2 = 1)\, pr(Y_3 = 1)}. \tag{3.2}$$

An advantage of the dependence ratio is that it does not grow in complexity for the higher-order dependence ratios. In contrast, the interpretation of the odds ratio grows in complexity for the higher-order moments. For example, a three-way odds ratio is the ratio of two conditional two-way odds ratios (Ekholm, 2003).

The dependence ratio approach facilitates the combining of a marginal regression model with an association model, using the profile probabilities π_i. For a binary response, the profile probabilities are defined as:

$$\pi_i = pr(Y_{i1} = y_{i1}, \ldots, Y_{iq} = y_{iq}), \tag{3.3}$$

where Y_{ik} represents the response for unit $i = 1, \ldots, n$ at subunit $k = 1, \ldots, q$.

For the psoriatic arthritis data, unit is patient and subunit is joint. The profile probability can be expressed in closed form in terms of the univariate marginal probabilities (first-order moments) and dependence ratios of all orders (association model).

The first-order moments are given by:

$$pr(Y_{ik} = 1 \mid \boldsymbol{x}_{ik}). \tag{3.4}$$

where $\boldsymbol{x}_{ik}$ is a vector of the explanatory variables.

For a binary response, the marginal regression model is obtained by regressing the first-order moments on relevant explanatory variables, using the logistic link function, which is given by:

$$\text{logit}[pr(Y_{ik} = 1)] = \theta + \boldsymbol{\beta}\boldsymbol{x}_{ik}^T, \tag{3.5}$$

where $\boldsymbol{\beta}$ is a vector of the regression coefficients.

The methods for dealing with the first-order moments do not typically differ between the different approaches for population-averaged based methods that use maximum likelihood estimation. However, the second-order or higher moments inform us about the association structure within a cluster, so consequently the methods differ depending on the association measure considered. For the dependence ratio paramterisation, the second-order or higher moments are replaced by the product of first-order moments and dependence ratios of the appropriate order. This is because the dependence ratio provides a more interpretable measure of association than the second-order or higher moments since it is relative to independence.

Given q observations within a cluster, constraints will often need to be imposed on the dependence ratios to reduce the number of association parameters, particularly when q is large. This is discussed in detail in Section 3.2.1. With association parameters defined by the vector $\boldsymbol{\alpha}$, the association model is defined as:

$$\boldsymbol{\tau} = g(\boldsymbol{\alpha}). \tag{3.6}$$

Consequently, the marginal regression model (3.5) and the association model (3.6) are combined using the profile probability in (3.3). The profile probability from (3.3) can also be written as:

$$\pi_i = pr(Y_{i1} = y_{i1}, \ldots, Y_{iq} = y_{iq}) = h(\theta, \boldsymbol{\beta}, \boldsymbol{\alpha}; \boldsymbol{x}_{ik}), \tag{3.7}$$

where $h(\theta, \boldsymbol{\beta}, \boldsymbol{\alpha}; \boldsymbol{x}_{ik})$ is a function of $\theta, \boldsymbol{\beta}, \boldsymbol{\alpha}$ and $\boldsymbol{x}_{ik}$.

In others words, the profile probability can be expressed in closed form, in terms of the regression and association parameters. The likelihood function, and consequently the log-likelihood function can then be specified. This gives the dependence ratio approach a key advantage over odds ratio approaches for maximum likelihood estimation since the odds ratio generally requires iterative procedures to specify the joint distribution (Jokinen, 2006: PhD Thesis). For more details, see Section 1.3 of the introduction to this thesis.

3.2.1 Association structures

Constraints will often need to be imposed on the association parameters (dependence ratios). For example, for a binary response with 6 observations in a cluster, $2^6 - 6 - 1 = 57$ dependence ratios need to be specified. Clearly this is too many to be estimated so relevant constraints will need to be imposed on the dependence ratios. In the psoriatic arthritis data, each patient has 28 joints which leads to a very large number of possible association parameters (268,435,427).

A nice feature of the dependence ratio approach is that association structures can easily be constructed that adequately represent the association structure using only a small number of association parameters. In addition to computational issues with estimating a large number of association parameters from the data, a small number is also desirable for the interest of model parsimony and interpretation. Ekholm, Smith and McDonald (2000) specified a series of useful association structures, some of which are relevant to the psoriatic arthritis data in this paper. Constraints may also be imposed on the dependence ratios based on the values of the observed dependence ratios or theoretical considerations.

Some of the association structures from Ekholm et al. (2000) are now discussed.

Independence (I):

Independence assumes that the observations within a cluster are independent. Consequently, all dependence ratios are equal to one. Independence is rarely applicable to multivariate categorical data.

Exchangeable structures (E):

Association structures from Ekholm et al. (2000) that are relevant to the context of the current paper are exchangeable structures. In contrast to Markov structures that were discussed in the introduction of this thesis, they do not assume a time ordering to the responses. The exchangeable structures considered in this paper are formulated through the use of latent variables.

Necessary factor (N):

In some studies, there may be a particular group of individuals (units) who will always give the same response throughout the study. This feature can be accounted for by using a necessary factor association structure. For a binary response with 1 as the response of interest and 0 as the baseline, there will often be a group of individuals who always respond 0. For the psoriatic arthritis data, 1 represents a clinically damaged joint for a given individual at a given joint and 0 represents no damage present. In this case, the structure separates the individuals that have the factor from those that do not have the factor that is necessary for the response of interest.

If a given individual has the necessary factor then $N_i = 1$, otherwise $N_i = 0$, for $i = 1, \ldots, n$. There is one association parameter given by $v = pr(N_i = 1)$. Consequently, $1-v$ represents the proportion of observations that will always respond 0 for each subunit. In the case of the psoriatic arthritis data, v represents the proportion of individuals that are susceptible to clinical damage, whereas $1 - v$ represents the proportion of individuals that are not susceptible to clinical damage in any of their joints.

The responses are conditionally independent given the necessary factor. Under this structure, all w-way dependence ratios, $\tau^{(w)}$, are equal, such that:

$$\tau^{(w)} = \gamma^{w-1}, \tag{3.8}$$

where $\gamma = 1/v$ and $w = 2, ..., q$.

This structure is particularly relevant to the psoriatic arthritis data (see Section 3.6) since 64.5% of patients have no clinical damage in any of their joints, where clinical damage is the response of interest. In addition, the explanatory variables for the regression model are (typically) regressed conditional on $N_i = 1$ (those susceptible to damage). It may also be beneficial to have the necessary factor varying by a between subject factor such as gender or age. This consequently increases the number of association parameters.

Latent binary factor (L):

A latent binary structure is appropriate if the population can be divided into two groups such that each group has different response category probabilities (given the same values for the explanatory variables in the regression model). Each individual either does ($L_i = 1$) or does not have ($L_i = 0$) the latent binary factor. Responses within a cluster are conditionally independent given L. The structure is often appropriate if an important dichotomous covariate has been omitted (Jokinen, McDonald and Smith, 2006).

For a binary response, this association model has 2 parameters, and is defined as:

$$\boldsymbol{\alpha} = (v_2, \kappa), \tag{3.9}$$

where $v_2 = pr(L_i = 1)$ and $\kappa = pr(Y_{ik} = 1 \mid L_i = 0)/pr(Y_{ik} = 1 \mid L_i = 1)$, for $k = 1, \ldots, q$.

In other words, v_2 represents the proportion of observations with the latent binary factor (latent group 1) and $1 - v_2$ represents the proportion without the factor (latent group 0). The κ parameter represents the conditional univariate probability for those in latent group 0 divided by the corresponding probability for those in group 1. All the w-way dependence ratios ($\tau^{(w)}$) are equal for this structure and are specified as follows:

$$\tau^{(w)} = \frac{v_2 + (1 - v_2)\kappa^w}{(v_2 + (1 - v_2)\kappa)^w}, \tag{3.10}$$

where $w = 2, ..., q$.

Latent binary factor within a necessary factor (NL):

The N and L exchangeable structures can also be extended further since it may be appropriate for a latent binary structure to operate within a necessary factor association structure (denoted NL). In other words, of those individuals that have the necessary factor, there are two groups of individuals with different response probabilities (given the same values for the regression variables).

3.2.2 Comparing the odds ratio and the dependence ratio

It has been noted that the dependence ratio has some key benefits over the odds ratio. Most notably, the fact it can cope computationally with much larger cluster sizes and allows for relevant association structures to be imposed. However, there are some issues with the dependence ratio approach that should be fully understood before the approach can be adequately implemented. These points are summarised below.

- **Orthogonality**: In contrast to the odds ratio, the regression and association parameters are not orthogonal to each other in the dependence ratio approach

(Ekholm, 2003). Therefore, estimating the correlations between the parameter estimates is recommended (Ekholm et al. 1995). Ekholm (2003) states that high correlations were rarely found in previous analyses and they are not found to be an issue for modelling the psoriatic arthritis data using the dependence ratio approach in Section 3.6.

- **Negative profile probabilities**: One of the main differences between the odds ratio and dependence ratio approaches is that the dependence ratio approach uses the profile probabilities to combine the regression and association models. Although the dependence ratio approach imposes positivity constraints on the observed profiles in the maximum likelihood estimation, negative fitted profile probabilities can occur for the unobserved profiles. However, this should not be treated as a drawback of the dependence ratio approach since negative profile probabilities act as a way of detecting if the model is incorrectly specified (Ekholm, 2003).

- **Range**: It is well known that the range of the odds ratio is from 0 to infinity with 1 representing independence. In contrast, the dependence ratio is constrained by the marginal probabilities (Ekholm, 2003). For example, for a two-way dependence ratio, for a binary response, denoted τ_{12}, the range is given by:

$$\max\left\{0, \frac{1}{\mu_1}+\frac{1}{\mu_2}-\frac{1}{\mu_1\mu_2}\right\} \leq \tau_{12} \leq \min\left\{\frac{1}{\mu_1}, \frac{1}{\mu_2}\right\}. \tag{3.11}$$

  The range of the dependence ratio has consequently received some criticism. However, as discussed in the introduction to this thesis, the fact the dependence ratio has a finite upper bound is arguably beneficial over the odds ratio. In addition, research in computer science found the range of the dependence ratio to only be a problem for small counts (see Section 1.4.6 for more details). The range of the dependence ratio is discussed again in Section 3.5 with a proportion of max measure proposed for the psoriatic arthritis data to assess the impact of the marginal probabilities on the conclusions obtained. For a full discussion of the range of the dependence ratio in comparison to the odds ratio, see Ekholm (2003).

3.2.3 R package drm

Prior to 2007 there was no statistical package readily available that allowed for the use of the dependence ratio approach. Hence, despite the advantages of the dependence ratio for population-averaged based approaches to multivariate categorical data using maximum likelihood estimation, the approach was not directly available for use in applied work. However, in 2007 the R package drm was released by Jukka Jokinen. Consequently, the dependence ratio can now be easily applied to relevant applications. The package is available free through CRAN or from the author's website:

http://www.helsinki.fi~jtjokine/drm/

The development of the package should help advance the use of the dependence ratio in applied work, particularly in cases where there are a large number of subunits in a cluster, such as the psoriatic arthritis data. The dependence ratio has recently been used in the paper by Anatolyev (2008) which uses the dependence ratio approach for looking at the dependence across European, Chinese and Baltic stock markets.

The package conveniently allows for the use of all the association structures from Section 3.2.1, amongst others. In addition, relevant fit statistics such as the deviance and Akaike's Information Criterion (AIC) are produced for model comparison. In cases where there are negative profile probabilities, the user is warned that the model needs to be respecified. For a full discussion of the features of the package, the reader is referred to the help pages on the website above. The psoriatic arthritis data are analysed using the dependence ratio approach in Section 3.6 and this is performed using drm with some additional code being used where necessary.

3.3 Marginal patterns of clinical damage: an exploratory analysis

This section performs some basic exploratory analysis on the Toronto psoriatic arthritis data in order to gain a better understanding of the marginal patterns of clinical damage (ignoring the associations).

In total there are 10,808 binary responses of clinical damage (where 1 represents damage and 0 represents no damage) since each of the 386 patients have 28 joints across their two hands. Recall that only the last clinic visit per patient is included since clinical damage is an irreversible process. Of the 10,808 joints, only 745 were damaged (6.89%). Splitting this further by gender in Table 3.3 shows a slightly higher proportion of damaged joints in females.

It has been well documented that males and females are equally likely to be affected with psoriatic arthritis. However, it should be noted that being affected simply relates to

Table 3.3: Clinical damage by gender

	Gender		
Damage	Females	Males	Total
no	4393	5670	10063
yes	423	322	745
Total	4816	5992	10,808
Proportion damaged	8.78%	5.37%	6.89%

being diagnosed with the disease and therefore differs from the issue of clinical damage discussed in this paper. In fact, females tend to progress quicker to damage than males (Gladman et al. 2005) and this is supported by the results shown in Table 3.3 although the significance of this is not assessed until the GEE analysis of Section 3.4.

It is also of interest to look at the probabilities of clinical damage at each individual joint since they form the basis of the GEE approach in Section 3.4 and the regression side of the models in the dependence ratio approach of Section 3.6. Table 3.4 shows this for the right hand and Table 3.5 for the left hand with marginal patterns by gender also included. The probabilities are given in percentage form for ease of interpretation.

Table 3.4: Right hand marginal percentages by gender

Joint Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Joint Type	**MCP**	**MCP**	**MCP**	**MCP**	**MCP**	**PIP**	**PIP**	**PIP**	**PIP**	**PIP**	**DIP**	**DIP**	**DIP**	**DIP**
Overall	10.9	4.7	3.9	2.1	2.3	4.4	7.8	8.0	4.4	8.0	10.9	9.3	7.8	11.4
Males	9.8	3.7	3.3	1.4	0.9	3.3	4.2	5.6	2.8	3.7	8.4	9.4	5.6	8.9
Females	12.2	5.8	4.7	2.9	4.1	5.8	12.2	11.1	6.4	13.4	14.0	9.3	10.5	14.5

Table 3.5: Left hand marginal percentages by gender

Joint Number	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Joint Type	**MCP**	**MCP**	**MCP**	**MCP**	**MCP**	**PIP**	**PIP**	**PIP**	**PIP**	**PIP**	**DIP**	**DIP**	**DIP**	**DIP**
Overall	9.8	4.9	4.7	2.9	2.6	5.4	8.0	7.8	6.5	6.5	11.9	9.1	8.6	8.6
Males	9.4	4.2	3.7	1.9	0.9	4.7	6.1	4.2	3.7	4.7	12.2	9.4	7.9	6.5
Females	10.5	5.8	5.8	4.1	4.7	6.4	10.5	12.2	9.9	8.7	11.6	8.7	9.3	11.1

Females have higher marginal probabilities of clinical damage in 25 of the 28 joints. The marginal patterns also seem to be very similar for each hand. In general, the MCP joints have the smallest marginal probabilities of damage (with the exception of the thumb) and the eight joints closest to the fingertip (the DIP joints) have the largest. This is not surprising since previous research has shown that the joints closest to the fingertip are commonly affected in psoriatic arthritis (Gladman et al. 2005). The above tables are limited in the sense that they do not take into account the effect of all explanatory variables. However, the significance of explanatory variables such as gender, hand and joint type is assessed in detail in the GEE approach of Section 3.4. The GEE approach

also takes the within-patient correlation into account. The marginal probabilities give no indication of the patterns of association in the data. In other words, a joint with a higher marginal probability of damage may not be substantially associated with any other joints. The associations are considered in detail in Sections 3.5 and 3.6.

3.4 GEE analysis

Before considering the association patterns using the dependence ratio approach, a GEE analysis of all 28 joints across both hands was conducted in order to determine which explanatory variables significantly predict the probability of clinical damage (using a 5% significance level). This is firstly done for all patients and secondly for patients who had at least one joint damaged, for reasons discussed in due course. The GEE approach (Liang and Zeger, 1986) concentrates on the marginal regression with the associations of secondary interest since moments of higher-order than two are not specified. Hence the full likelihood is not available and the approach is a quasi-likelihood based method. The GEE approach specifies a working correlation matrix for dealing with the second-order moments and has the advantage that even if this matrix is misspecified, the parameter estimates are consistent (Liang and Zeger, 1986). However, robust standard errors based on the sandwich estimator are preferred to the model based standard errors since they guard against misspecification of the working correlation matrix.

A disadvantage of using correlations for the pairwise associations of binary data in the GEE approach is that they are constrained by the marginal probabilities. As a consequence, they may yield estimates that are outside the range of possible values. This issue is discussed in detail in Chaganty and Joe (2004). One way to combat this is to use the independence working correlation matrix with robust standard errors, as recommended in Hanley et al. (2000). This approach is reasonable provided the correlations are not too strong. Liang and Zeger (1986) stated that the independence working correlation matrix has good efficiency provided the actual correlations are weak to moderate. Agresti (2002) also stated that all working correlation matrices should yield similar results when the correlations are modest. This is what was found for the analyses in this section so the choice therefore does not seem overly crucial. However, to avoid any potential difficulties with the correlation bounds, the independence working correlation matrix with robust standard errors was chosen.

3.4.1 GEE analysis of all patients

The explanatory variables available for this analysis were sex, age (in years) that the patient was diagnosed with psoriatic arthritis, hand (right or left), finger (5 level categorical variable with little finger as the baseline), joint type (MCP, PIP or DIP, with DIP

as the baseline) and time (in years) since diagnosis, with the response of interest being the presence of clinical damage (for each joint of each patient at their last clinic visit). The logistic link function was used for the binary response of clinical damage. Recall that the thumb digit has only the MCP and PIP joint types present whereas the other four fingers have each of the MCP, PIP and DIP present. In other words, the finger and joint type variables are not perfectly crossed. The covariates age since diagnosis and time since diagnosis were both mean centered to aid interpretability. Table 3.6 shows the results from a model with all the main effects and no interactions.

Table 3.6: Parameter estimates for the GEE analysis of all patients: main effects model

Effects		Estimate	Standard error
Intercept		−2.565	0.188
Joint type	MCP	−0.882	0.165
	PIP	−0.526	0.130
Sex	Female	0.2720	0.262
Age at diagnosis (centered)	(in years)	0.0150	0.010
Finger	Thumb	0.4500	0.166
	Index	0.2340	0.114
	Middle	0.0940	0.117
	Ring	−0.231	0.107
Hand	Left	0.0150	0.075
Time since diagnosis (centered)	(in years)	0.0680	0.009

Although sex is non-significant, females are associated with a higher probability of damage than males. This coincides with previous research on psoriatic arthritis that has acknowledged that females progress quicker to damage than males even though males and females are equally likely to be affected with psoriatic arthritis (Gladman et al. 2005). Gender is of specific interest in this paper so will consequently be included throughout regardless of its significance. Hand and age of diagnosis were found to be non-significant. Joint type was found to be significant with the MCP and PIP joints (significantly) associated with lower probabilities of damage than the DIP joints. Finger is also significant with the thumb and index finger both being (significantly) associated with higher probabilities of clinical damage than the little finger. In addition, the ring finger is (significantly) associated with lower probabilities of clinical damage than the little finger. Time since diagnosis is also significant with increased time since diagnosis being (significantly) associated with a higher probability of clinical damage.

Since the full likelihood is not available in the GEE approach, likelihood based model selection methods such as likelihood ratio tests and the AIC are not available. Consequently, Pan (2001) proposed an alternative to suit the GEE approach (Quasi Information Criterion = QIC). The QIC is used throughout the GEE analyses to confirm

conclusions obtained by the Wald test. Stepwise selection procedures using the QIC revealed the same conclusions as those shown in Table 3.6. However, the main effects model presented in Table 3.6 does not take into account interactions between variables. This is discussed in the following section.

3.4.1.1 GEE analysis of all patients: interactions

Following the assessment of the main effects, the age and hand explanatory variables were excluded from the further analysis of the interactions, but gender was included due to the theoretical reasons described previously. A detailed analysis of all the two-way interactions was conducted. The QIC was used to determine which, if any, interactions should to be included. This led to the inclusion of the interactions for joint type with each of gender and finger. The final chosen model for the GEE analysis of all patients across all 28 joints therefore has the main effects for sex ($sex_i = 1$ (female), 0 = (male)), joint type ($j = 1$ (MCP), 2 (PIP) or 3 (DIP = baseline)), finger ($f = 1$(thumb), 2 (index), 3 (middle), 4 (ring) or 5 (little finger = baseline)) and time since diagnosis (centered), as well as the mentioned interactions. This model can be expressed as:

$$\text{logit}[pr(Y_{ijf} = 1)] = \beta_0 + \beta_1 sex_i + \beta_{2j} + \beta_{3f} + \beta_4 time_i + \beta_{12j} sex_i + \beta_{23jf}, \tag{3.12}$$

where $i = 1, \ldots, 386, j = 1, 2, 3$ and $f = 1, \ldots, 5$.

Table 3.7 summaries the parameter estimates for the main effects in the model. Tables 3.8 and 3.9 show the parameter estimates for the interactions. Interaction plots are displayed in Figures 3.2 and 3.3.

Table 3.7: Parameter estimates for the chosen model of all patients: main effects

Effects		Estimate	Standard error
Intercept		−2.385	0.189
Joint type	MCP	−1.626	0.293
	PIP	−0.687	0.184
Sex	Female	0.0580	0.278
Finger	Thumb	−0.442	0.211
	Index	0.1590	0.130
	Middle	−0.095	0.141
	Ring	−0.234	0.136
Time since diagnosis (centered)	(in years)	0.0630	0.008

Table 3.8: Parameter estimates for the chosen model of all patients: sex by joint type interaction with standard errors in parentheses

	JOINT TYPE		
SEX	MCP	PIP	DIP
Female	0.138 (0.308)	0.575 (0.212)	0
Male	0	0	0

Table 3.9: Parameter estimates for the chosen model of all patients: finger by joint type interaction with standard errors in parentheses

	JOINT TYPE		
FINGER	MCP	PIP	DIP
Thumb	2.044 (0.337)	0	0
Index	0.563 (0.219)	−0.059 (0.178)	0
Middle	0.691 (0.229)	0.1950 (0.171)	0
Ring	0.234 (0.212)	−0.095 (0.173)	0
Little	0	0	0

Table 3.7 shows that time since diagnosis is again significant with an increase associated with a higher probability of damage. In fact, for every one year increase in the time since diagnosis, the odds of clinical damage are 1.065 times greater (holding other variables constant). The significance of the interactions means that interpretations regarding the sex, joint type and finger variables should be in terms of the interactions in order to obtain sensible conclusions.

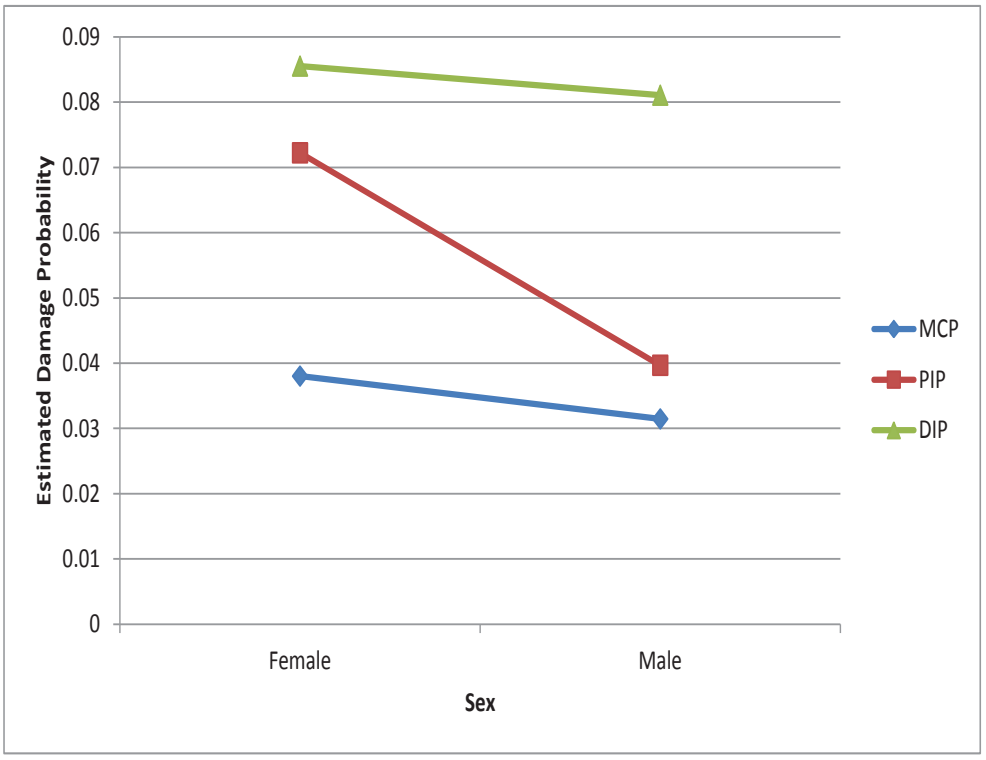

Figure 3.2: Sex by joint type interaction: all patients

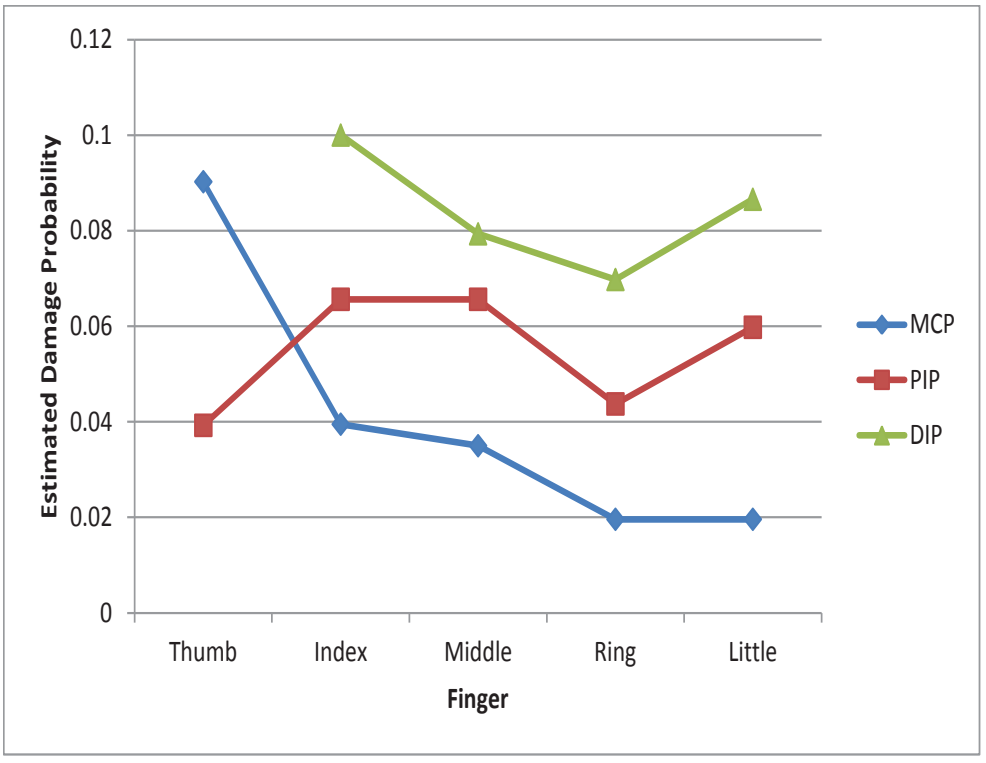

Figure 3.3: Finger by joint type interaction: all patients

The estimated marginal damage probabilities in Figures 3.2 and 3.3 take into account the other terms in the model. The time since diagnosis variable is fixed at its mean value (0) since it is mean centered. The mean of the non-centered time since diagnosis variable is 16.58 years. Figure 3.2 shows that the DIP joints have the highest probability of damage in both males and females (approximately 0.08) and the MCP joints have the lowest probabilities of damage in both males and females (approximately 0.03). However, the significant interaction between sex and joint type appears to be due to the differing patterns with regards to the PIP joints. This was confirmed by post hoc tests with a 5% significance level (controlling the overall type 1 error). Post hoc tests revealed that males had significantly higher probabilities of damage in their DIP joints than both the PIP and MCP joints. In contrast, there was no significant difference between the probability of damage in the DIP and PIP joints of females. In addition, the DIP and PIP joints both have significantly higher probabilities of damage than the MCP joints in females.

Figure 3.3 shows that the interaction between finger and joint type appears to be due to the differing patterns of the thumb in comparison to the other fingers since the lines are approximately parallel when the thumb is ignored. Although the thumb has no DIP joint, its patterns with regards to the MCP and PIP joints differ in relation to the other fingers. The thumb has a larger estimated probability of damage for the MCP joint (0.09) than it does for the PIP joint (0.04) whereas other fingers see the reverse pattern. Post hoc tests again confirm the conclusions. The MCP thumb joint has a significantly higher probability of damage than the PIP thumb joint (as well as the MCP joints in the other four fingers). In contrast, the MCP joint has a lower probability than the PIP joint in the other four fingers although only the little finger was significantly lower. Excluding the thumb which has no DIP joint, the DIP joints have significantly higher probabilities of damage than the MCP joints in all four fingers. Although the DIP joints had consistently larger probabilities of damage than the PIP joints at every finger, none were significantly greater.

3.4.2 Further GEE analysis

Of the 386 patients in the dataset, 249 had no damage in any of their 28 joints (64.5%). This acts as a strong reason for considering the necessary factor association structure discussed in Section 3.2.1. This structure is considered in detail in Section 3.6 when the dependence ratio approach is considered for taking into account the association structure. However, one of the features of the necessary factor association structure is that the marginal regression model is (typically) estimated conditional on the patients having the necessary factor for damage (being susceptible to damage). This therefore brings about the thought that the patients with no damage in any of their 28 joints are in fact not susceptible to clinical damage. In other words, only the 35.5% who have

at least one joint damaged are susceptible. Consequently, this section applies the GEE approach to just this 35.5% in order to see if the conclusions differ.

Of the 137 patients who had at least one joint damaged, 69 were male (50.4%) and 68 were female (49.6%). This shows that females are associated with greater clinical damage since in the analysis with all 386 patients, 214 were male (55.4%) and 172 were female (44.6%).

Table 3.10 shows the results from a model with just the main effects from Table 3.6. The conclusions were the same as those obtained in Table 3.6 when all patients were included. Finger, joint type and time since diagnosis (mean centered) were significantly associated with predicting clinical damage. Although sex was non-significant, females were associated with a greater probability of damage than males. Hand and age of diagnosis (mean centered) were again non-significant.

Table 3.10: Parameter estimates for the GEE analysis of patients with at least one joint damaged: main effects model

Effects		Estimate	Standard error
Intercept		−1.114	0.187
Joint type	MCP	−0.997	0.181
	PIP	−0.604	0.147
Sex	Female	0.1660	0.246
Age of diagnosis (centered)	(in years)	-0.003	0.010
Finger	Thumb	0.5040	0.186
	Index	0.2700	0.132
	Middle	0.1080	0.134
	Ring	−0.258	0.119
Hand	Left	0.0180	0.086
Time since diagnosis (centered)	(in years)	0.0370	0.009

Prior to considering the two-way interactions, hand and age of diagnosis were removed from the model, but sex was kept in the model for theoretical reasons. A further investigation of the two-way interactions using the QIC led to the inclusion of the interactions between joint type and each of sex and finger, as before. The final chosen model for the subset of patients with at least one joint damaged is therefore the same as when all patients were included (3.12). Table 3.11 shows the parameter estimates for the main effects in this model. Tables 3.12 and 3.13 show the parameter estimates for the interactions with Figures 3.4 and 3.5 showing the corresponding interaction plots.

Table 3.11: Parameter estimates for the chosen model of patients with at least one joint damaged: main effects

Effects		Estimate	Standard error
Intercept		−0.909	0.195
Joint type	MCP	−1.796	0.318
	PIP	−0.798	0.213
Sex	Female	−0.115	0.280
Finger	Thumb	−0.491	0.233
	Index	0.1970	0.161
	Middle	−0.115	0.170
	Ring	−0.278	0.160
Time since diagnosis (centered)	(in years)	0.0390	0.009

Table 3.12: Parameter estimates for the chosen model of patients with at least one joint damaged: sex by joint type interaction

	JOINT TYPE		
SEX	MCP	PIP	DIP
Female	0.177 (0.354)	0.665 (0.247)	0
Male	0	0	0

Table 3.13: Parameter estimates for the chosen model of patients with at least one joint damaged: finger by joint type interaction

	JOINT TYPE		
FINGER	MCP	PIP	DIP
Thumb	2.253 (0.356)	0	0
Index	0.561 (0.238)	−0.082 (0.211)	0
Middle	0.738 (0.246)	0.230 (0.201)	0
Ring	0.278 (0.230)	−0.090 (0.198)	0
Little	0	0	0

Time since diagnosis is again significant with an increase associated with a higher probability of damage. For a one year increase in the time since diagnosis, the odds of clinical damage are 1.040 times greater (holding other variables constant). The significance

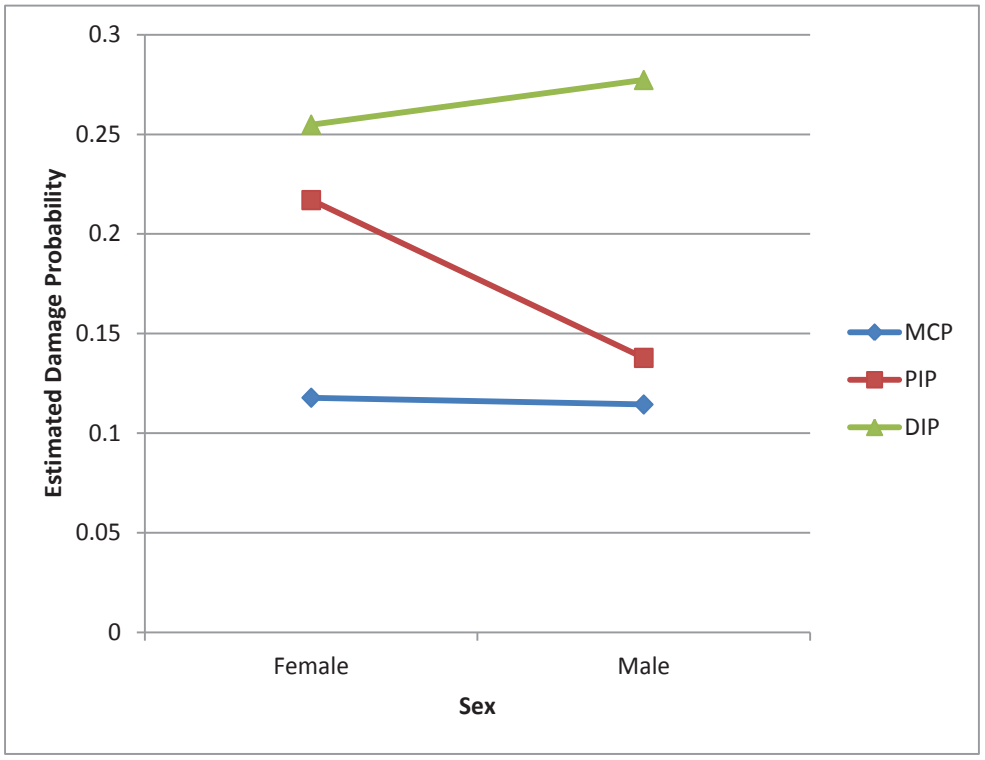

Figure 3.4: Sex by joint type interaction: at least one joint damaged

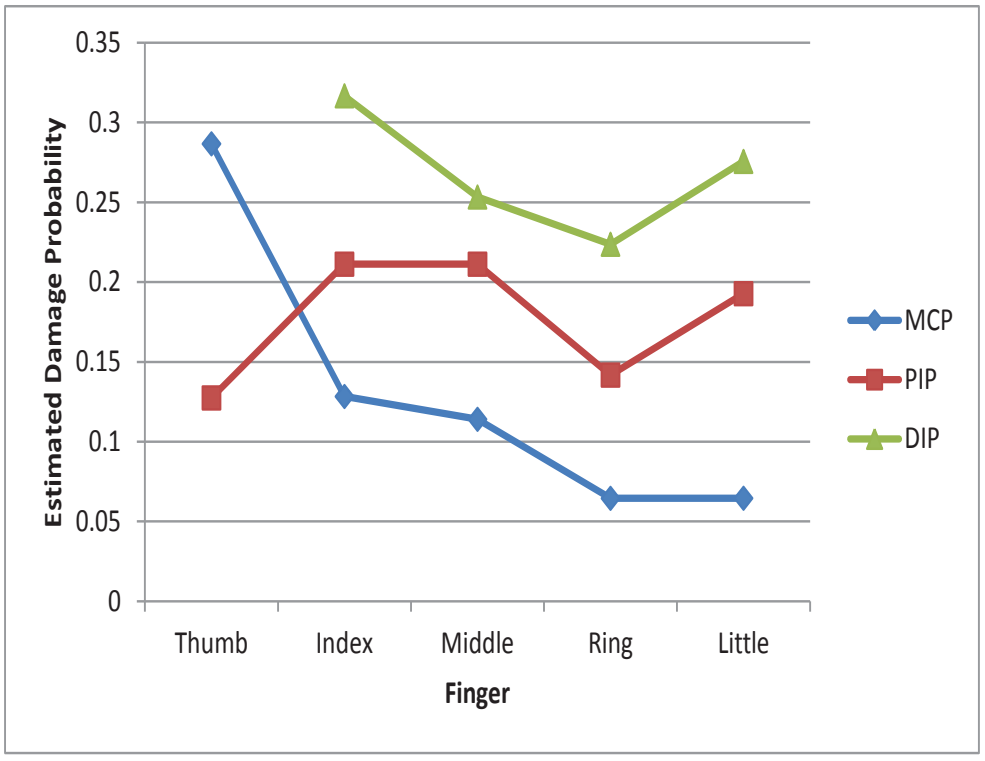

Figure 3.5: Finger by joint type interaction: at least one joint damaged

of the interactions means that interpretations regarding the sex, joint type and finger variables should be in terms of the interactions.

Figure 3.4 shows similar patterns as in Figure 3.2 when all patients were considered. However, the estimated probabilities of damage are unsurprisingly much greater in Figure 3.4 since only those with at least one joint damaged are considered. Figure 3.4 shows the highest probabilities of damage are in the DIP joints (for both males and females) with estimated probabilities of damage of approximately 0.26 (although males are now slightly higher). The lowest probabilities of damage are in the MCP joints with both males and females having an estimated damage probability of 0.12. The significant interaction seems to be due to the differing patterns in males and females with regards to the PIP joints. Females have an estimated damage probability of 0.22 whereas males have a much lower estimate of 0.14. Post hoc tests (controlling for multiple comparisons) confirmed the conclusions since males have significantly higher probabilities of damage in their DIP joints than both the PIP and MCP joints. However, females had no significant difference between their DIP and PIP joints but both of which were significantly larger than the MCP joints.

Figure 3.5 shows the same patterns as when all patients were included in Figure 3.3 although the probabilities are much greater in Figure 3.5 as expected. The significant interaction appears to be due to the differing patterns in the thumb with regards to the MCP and PIP joints. Post hoc tests confirmed this conclusion. The estimated probability of damage in the MCP joint of the thumb (0.29) is significantly larger than the PIP thumb joint (0.13). In contrast, the other fingers have larger probabilities in the PIP joint than the MCP joint although only the ring and little fingers were significantly larger. Excluding the thumb which has no DIP joint, the other four fingers have significantly higher probabilities in their DIP joint than the corresponding MCP joint (but not the corresponding PIP joint).

3.4.2.1 Conclusions for the GEE analyses

The GEE analyses in this section have allowed for a detailed analysis of the marginal patterns of clinical damage, irrespective of the association structure. The first analysis considered all 386 patients, whereas the second analysis focused on only those patients who had at least one of their 28 joints damaged. The reason for the second analysis was that patients with no damage in any of their 28 joints is a strong justification for believing that these patients may not be susceptible to damage and should be excluded from the GEE analysis.

However, both approaches led to the same conclusions in terms of which variables significantly predict clinical damage and the same final model was obtained in each case

(3.12). Although sex was found to be non-significant when only main effects were considered, it was kept in the model for theoretical reasons. The sex by joint type interaction showed the differing paterns with regards to the PIP joints in males and females. In addition, the finger by joint type interaction showed the differing patterns of the thumb in comparison to other fingers. Excluding the thumb (which has no DIP joint), the DIP joints have the highest probability of damage for each finger (significantly greater than the corresponding MCP joint but not the corresponding PIP joint). The DIP joints are well known to be affected in psoriatic arthritis so it not surprising that the same patterns emerge when clinical damage is considered as opposed to simply whether the joint is affected. Time since diagnosis was found to significantly predict clinical damage with increases in this covariate associated with higher probabilities of clinical damage. Hand and age of diagnosis were found to not significantly predict the probability of clinical damage.

Although both GEE analyses led to the same final model, the probabilities of damage were clearly much higher in the second analysis since individuals with no damage in any of their joints were excluded. Recall that only patients who had no clinical damage in any of their 28 joints upon entry to the clinic were included in the dataset with the response of interest being whether they have clinical damage at their last clinic visit (for each joint). Consequently, the GEE analysis which includes all individuals is therefore more relevant to assessing clinical damage in those who are affected with psoriatic arthritis, but may not necessarily be susceptible to clinical damage. In contrast, the GEE analysis which only includes patients with at least one joint damaged is more relevant to assessing clinical damage in those who are susceptible to clinical damage.

The results in this section also agree with the results from Bond et al. (2007) that were discussed in Section 3.1.1 in the sense that females are associated with greater progression to damage than males (although not significantly). Bond et al. (2007) also found time in the clinic to significantly predict the progression to damage (although time since diagnosis is considered in this paper). Bond et al. (2007) found age to significantly predict the progression to damage, but this could be due to age at entry being considered as opposed to the age of diagnosis considered in this paper. It should be noted that the assessment by Bond et al. (2007) was in a longitudinal sense over time rather than considering the last clinic visit (for each patient). In addition, Bond et al. (2007) included the feet joints as well as the hand joints.

The analyses conducted so far in Sections 3.3 and 3.4 concentrated purely on the marginal patterns of clinical damage and ignore the associations. The following sections use the dependence ratio approach, which gives equal emphasis to the marginal regression and the associations within a cluster (patient).

3.5 Association patterns of clinical damage: exploratory analysis

In order to assess the association patterns of clinical damage, the observed two-way dependence ratios were assessed for pairs of joints. This is performed separately for the within and between-hand dependence ratios. The between-hand dependence ratios include the symmetric dependence ratios for an assessment of symmetry, where symmetry refers to the same joint(s) on both hands being affected. The symmetric two-way dependence ratio for a given joint is defined as the joint probability of an individual having both the given joint and the same joint on the other hand damaged divided by the corresponding probability under independence.

Some previous analyses with the dependence ratio have analysed males and females separately, e.g. Jokinen et al. (2006), because they were found to follow different association structures. In the case of the psoriatic arthritis data, males and females were found to follow similar association patterns from the observed dependence ratios. Consequently, the observed dependence ratios that follow in this section will not be separated by gender. However, given the theoretical interest in the gender variable and the higher marginal probabilities of clinical damage for females from Sections 3.3 and 3.4, gender is included in the marginal regression models of Section 3.6.

As discussed in Section 3.4, 249 of the 386 patients (64.5%) had no clinical damage in any of their 28 joints. There was a similar pattern by gender since, of the 214 males, 145 had no clinical damage in any of their joints (67.76%) and, of the 172 females, 104 had no damage in any joints (60.47%). In other words, there appears to be a group of patients that are not susceptible to clinical damage. Consequently, the necessary factor association structure that was described in Section 3.2.1 is likely to appropriate for modelling the association structure. This is discussed further in Section 3.6.

In addition to considering the observed two-way dependence ratios, a proportion of max measure is also considered for assessing the association between joints. In Section 3.2.2, it was stated that the dependence ratio has a range that is constrained by the marginal probabilities. Consequently, the proportion of max measure (denoted τ^{POM}) ensures that conclusions are not majorly influenced by the marginal probabilities. It should also be noted that research in computer science (see Section 1.4.6 of the introduction to this thesis) found the range of the two-way dependence ratio to only be an issue when the counts (for the numerator) are small (less than approximately five). Although all counts were greater than five for the psoriatic arthritis data, some counts were relatively small in comparison to others. One proportion of max measure that could be used is:

$$\tau^{POM} = \frac{\tau}{max(\tau)}. \tag{3.13}$$

In other words, the value of the dependence ratio divided by its potential maximum, where the potential maximum is given by the minimum of the reciprocals of the two marginal probabilities (as shown in (3.11)). However, although this measure has a minimum value of 0 and an maximum value of 1, there is no fixed value for independence. Consequently, a preferable measure is:

$$\tau^{POM2} = \frac{\tau - 1}{max(\tau) - 1}. \tag{3.14}$$

Since all dependence ratios are 1 under independence, zero represents independence for this measure. Although the minimum value is less than 0, the maximum is still attained at 1. So, the interpretation of the measure is based upon values from independence (0) to its maximum (1). No proportion of max values were found to be less than or equal to zero in the psoriatic arthritis data. Although the conclusions were the same regardless of whether (3.13) or (3.14) were used, (3.14) is used throughout the subsequent parts of this section due to the advantage mentioned. Proportion of max values greater than 0.80 are given in bold throughout. In fact, (3.14) is the same as the statistic proposed by Durbin (1955) for the context of social mobility tables (see Section 1.4.2).

Observed dependence ratios of higher-order than two are not considered. Firstly, the numerators in some of the higher-order dependence ratios were zero. An analysis of the three-way dependence ratios revealed similar conclusions to those obtained from inspecting the two-way dependence ratios. The within-hand two-way dependence ratios are considered first along with their proportion of max.

3.5.1 Within-hand dependence ratios

The right hand observed two-way dependence ratios are considered first in Table 3.14 with the corresponding proportion of max displayed in Table 3.15. Tables 3.16 and 3.17 show the corresponding values for the left hand. Since all the dependence ratios are much larger than 1, the independence association structure is clearly not applicable for modelling the data using the dependence ratio approach in Section 3.6. However, although all of the dependence ratios are large relative to independence, interest lies in determining the greatest within-hand association patterns of clinical damage.

The largest two-way observed dependence ratios in the right hand are highlighted in bold in Table 3.14. Of particular note is the larger values of all the two-way dependence ratios involving only the MCP joints 2, 3, 4 and 5. These joints are the larger MCP joints (excluding the thumb) and are commonly known as knuckles. This is confirmed by looking at the proportion of max values in Table 3.15. In addition, similar patterns are seen in the left hand with the MCP knuckle joints having the greatest association. Overall, the two hands seem to follow similar association patterns. In other words, these

results suggest there is no need to have different association structures for the two hands in Section 3.6.

Table 3.14: Right hand observed two-way dependence ratios

Joint	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	-	5.106	4.902	5.744	6.127	6.487	3.370	3.854	4.866	3.558	4.158	4.085	4.595	3.760
2	5.106	-	**20.015**	**21.440**	**19.060**	8.830	7.148	6.918	12.61	6.226	4.595	6.552	6.433	4.386
3	4.902	**20.015**	-	**25.733**	**20.010**	7.569	7.720	7.471	**15.140**	5.811	4.902	7.148	6.862	4.679
4	5.744	**21.440**	**25.733**	-	**37.528**	11.350	9.650	9.339	**19.870**	7.782	5.744	8.042	9.650	6.580
5	6.127	**19.060**	**20.010**	**37.528**	-	12.614	10.010	8.301	**17.660**	9.685	5.106	7.148	10.010	6.823
6	6.487	8.830	7.569	11.350	12.614	-	4.541	5.860	9.349	7.324	7.028	6.307	9.839	6.709
7	3.370	7.148	7.720	9.650	10.010	4.541	-	8.301	10.596	5.396	5.208	5.719	6.004	4.679
8	3.854	6.918	7.471	9.339	8.301	5.860	8.301	-	11.719	6.025	5.929	7.609	7.471	5.094
9	4.866	12.610	**15.140**	**19.870**	**17.660**	9.349	10.596	11.719	-	8.789	8.109	9.461	11.350	7.225
10	3.558	6.226	5.811	7.782	9.685	7.324	5.396	6.025	8.789	-	4.447	5.188	6.226	5.377
11	4.158	4.595	4.902	5.744	5.106	7.028	5.208	5.929	8.109	4.447	-	7.148	7.659	5.013
12	4.085	6.552	7.148	8.042	7.148	6.307	5.719	7.609	9.461	5.188	7.148	-	8.22	5.361
13	4.595	6.433	6.862	9.650	10.010	9.839	6.004	7.471	11.350	6.226	7.659	8.220	-	6.433
14	3.760	4.386	4.679	6.580	6.823	6.709	4.679	5.094	7.225	5.377	5.013	5.361	6.433	-

Table 3.15: Right hand proportion of max

Joint	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	NA	0.501	0.476	0.579	0.626	0.670	0.289	0.348	0.472	0.312	0.386	0.377	0.439	0.355
2	0.501	NA	**0.930**	**1**	**0.883**	0.383	0.518	0.517	0.568	0.456	0.439	0.571	0.458	0.436
3	0.476	**0.930**	NA	**1**	0.769	0.303	0.566	0.565	0.651	0.420	0.476	0.632	0.494	0.473
4	0.579	**1**	**1**	NA	**0.872**	0.477	0.729	0.728	**0.869**	0.592	0.579	0.724	0.729	0.718
5	0.626	**0.883**	0.769	**0.872**	NA	0.535	0.759	0.638	0.768	0.758	0.501	0.632	0.759	0.749
6	0.670	0.383	0.303	0.477	0.535	NA	0.298	0.424	0.385	0.552	0.736	0.546	0.745	0.734
7	0.289	0.518	0.566	0.729	0.759	0.298	NA	0.638	**0.809**	0.384	0.514	0.485	0.422	0.473
8	0.348	0.517	0.565	0.728	0.638	0.424	0.638	NA	0.936	0.439	0.602	0.680	0.565	0.527
9	0.472	0.568	0.651	**0.869**	0.768	0.385	**0.809**	**0.936**	NA	0.680	**0.868**	**0.870**	**0.872**	**0.801**
10	0.312	0.456	0.420	0.592	0.758	0.552	0.384	0.439	0.680	NA	0.421	0.431	0.456	0.563
11	0.386	0.439	0.476	0.579	0.501	0.736	0.514	0.602	**0.868**	0.421	NA	0.751	**0.813**	0.516
12	0.377	0.571	0.632	0.724	0.632	0.546	0.485	0.680	**0.870**	0.431	0.751	NA	0.743	0.561
13	0.439	0.458	0.494	0.729	0.759	0.745	0.422	0.565	**0.872**	0.456	**0.813**	0.743	NA	0.699
14	0.355	0.436	0.473	0.718	0.749	0.734	0.473	0.527	**0.801**	0.563	0.516	0.561	0.699	NA

Table 3.16: Left hand observed two-way dependence ratios

Joint	15	16	17	18	19	20	21	22	23	24	25	26	27	28
15	-	5.881	5.643	6.464	8.126	7.256	4.587	4.402	3.657	3.251	4.416	4.063	4.002	3.078
16	5.881	-	**19.187**	**16.62**	**18.28**	7.739	9.83	8.804	6.501	5.688	6.625	6.385	6.156	3.694
17	5.643	**19.187**	-	**19.495**	**19.3**	7.148	10.38	9.293	6.862	6.862	6.527	6.74	6.498	3.899
18	6.464	**16.62**	**19.495**	-	**28.073**	10.03	10.19	9.358	8.422	9.825	6.103	6.016	6.38	5.317
19	8.126	**18.28**	**19.3**	**28.073**	-	12.867	11.21	11.58	10.81	10.81	6.713	6.617	8.188	8.188
20	7.256	7.739	7.148	10.03	12.867	-	5.929	6.74	7.352	5.882	5.994	5.777	7.798	5.57
21	4.587	9.83	10.38	10.19	11.21	5.929	-	9.546	7.471	6.973	5.955	6.759	6.037	3.773
22	4.402	8.804	9.293	9.358	11.58	6.74	9.546	-	7.72	7.72	5.594	6.985	6.238	5.069
23	3.657	6.501	6.862	8.422	10.81	7.352	7.471	7.72	-	9.882	4.699	5.294	7.486	4.679
24	3.251	5.688	6.862	9.825	10.81	5.882	6.973	7.72	9.882	-	4.363	3.97	6.082	6.082
25	4.416	6.625	6.527	6.103	6.713	5.994	5.955	5.594	4.699	4.363	-	7.432	6.611	5.086
26	4.063	6.385	6.74	6.016	6.617	5.777	6.759	6.985	5.294	3.97	7.432	-	8.021	5.681
27	4.002	6.156	6.498	6.38	8.188	7.798	6.037	6.238	7.486	6.082	6.611	8.021	-	6.38
28	3.078	3.694	3.899	5.317	8.188	5.57	3.773	5.069	4.679	6.082	5.086	5.681	6.38	-

Table 3.17: Left hand proportion of max

Joint	15	16	17	18	19	20	21	22	23	24	25	26	27	28
15	NA	0.533	0.507	0.597	0.778	0.683	0.392	0.371	0.290	0.246	0.462	0.334	0.328	0.227
16	0.533	NA	**0.942**	**0.809**	**0.895**	0.388	0.771	0.658	0.381	0.325	0.761	0.537	0.482	0.252
17	0.507	**0.942**	NA	**0.905**	**0.895**	0.354	**0.819**	0.699	0.406	0.406	0.748	0.572	0.514	0.271
18	0.597	**0.809**	**0.905**	NA	0.794	0.519	**0.802**	0.704	0.514	0.611	0.690	0.500	0.503	0.404
19	0.778	**0.895**	**0.895**	0.794	NA	0.683	**0.891**	**0.892**	0.679	0.679	0.773	0.560	0.672	0.672
20	0.683	0.388	0.354	0.519	0.683	NA	0.430	0.484	0.440	0.338	0.676	0.476	0.636	0.427
21	0.392	0.771	**0.819**	**0.802**	**0.891**	0.430	NA	0.746	0.565	0.522	0.670	0.574	0.471	0.259
22	0.371	0.658	0.699	0.704	**0.892**	0.484	0.746	NA	0.566	0.566	0.622	0.597	0.490	0.380
23	0.290	0.381	0.406	0.514	0.679	0.440	0.565	0.566	NA	0.615	0.500	0.428	0.606	0.344
24	0.246	0.325	0.406	0.611	0.679	0.338	0.522	0.566	0.615	NA	0.455	0.296	0.475	0.475
25	0.462	0.761	0.748	0.690	0.773	0.676	0.670	0.622	0.500	0.455	NA	**0.870**	0.759	0.553
26	0.334	0.537	0.572	0.500	0.560	0.476	0.574	0.597	0.428	0.296	**0.870**	NA	0.700	0.467
27	0.328	0.482	0.514	0.503	0.672	0.636	0.471	0.490	0.606	0.475	0.759	0.700	NA	0.503
28	0.227	0.252	0.271	0.404	0.672	0.427	0.259	0.380	0.344	0.475	0.553	0.467	0.503	NA

3.5.2 Between-hand dependence ratios

This section considers the between-hand dependence ratios. These are displayed in Table 3.18 with the diagonal elements of the table representing the symmetric dependence ratios. Although all of the dependence ratios in Table 3.18 are large relative to independence, the greatest associations again largely involve the larger MCP knuckle joints. However, these associations appear to not be as strong as the within-hand dependence ratios involving these joints. This is shown by the smaller proportion of max values in Table 3.19.

Table 3.18: Between-hand dependence ratios

Joint	1	2	3	4	5	6	7	8	9	10	11	12	13	14
15	7.497	4.515	4.740	6.349	5.643	5.378	3.725	3.604	4.780	3.604	4.353	3.950	4.063	3.694
16	4.837	**12.415**	**12.189**	**15.237**	**15.801**	7.170	8.804	6.553	**11.950**	5.243	5.805	6.208	6.772	4.617
17	5.106	**13.105**	**12.867**	**16.083**	**16.679**	7.569	9.293	7.609	**12.614**	6.226	6.127	6.552	7.148	4.386
18	5.013	**13.646**	**16.376**	**21.932**	**19.495**	8.257	9.358	9.056	**14.449**	7.924	5.848	7.798	7.018	4.785
19	6.433	**15.011**	**15.440**	**24.125**	**25.733**	11.353	11.580	8.716	**15.894**	9.961	6.433	8.578	7.720	6.141
20	6.127	6.127	6.127	9.190	8.169	11.894	4.902	5.336	7.569	4.743	5.689	5.616	6.127	5.431
21	4.447	7.609	8.301	9.339	8.301	5.860	7.886	7.632	10.254	4.820	5.336	5.880	6.226	3.962
22	4.289	7.148	7.720	9.650	10.007	6.055	8.578	9.131	10.596	6.226	5.821	6.076	6.862	4.971
23	3.676	5.147	5.147	9.650	8.578	8.174	6.691	6.475	9.991	6.475	5.882	4.718	6.691	4.913
24	3.676	5.147	5.147	7.720	6.862	7.266	6.691	6.475	8.174	5.977	4.411	3.860	5.147	4.913
25	3.996	4.662	5.035	6.293	5.594	5.923	4.755	5.143	7.404	4.602	6.393	6.527	6.433	5.531
26	4.464	6.127	5.882	8.271	7.352	7.136	5.147	6.759	9.082	4.981	7.090	7.965	8.823	6.016
27	4.456	5.199	5.459	8.773	7.798	8.945	5.459	6.414	8.945	5.660	6.684	7.148	8.578	6.380
28	2.785	3.249	3.119	4.386	5.199	6.193	3.899	5.283	6.881	4.905	5.291	5.524	5.848	5.051

The results in Sections 3.5.1 and 3.5.2 suggest that the greatest association patterns of clinically damaged joints involve the 8 larger MCP knuckle joints and in particular within a hand rather than between hands. Although the symmetric dependence ratios (the diagonal elements of Table 3.18) are all large relative to independence, the results in these sections suggest that these symmetric association patterns of damaged joints are not a strong as the within hand and between hand association patterns involving the MCP knuckle joints. The results suggest that the MCP knuckle joints at least appear to follow a different association pattern to the other joints. O'Keeffe et al. (2011) mention that strong association patterns involving the MCP knuckle joints (the larger joints) is common in psoriatic arthritis.

Although Sections 3.3 and 3.4 showed that the MCP knuckle joints had lower marginal probabilities compared to the other joints, this section has demonstrated that stronger association patterns exist with regards to these joints. To illustrate this point, consider the two-way dependence ratio for the MCP joints in the index and middle fingers (joint numbers 2 and 3) in the right hand. The marginal probabilities for these joints are 0.0466

Table 3.19: Between-hand proportion of max

Joint	1	2	3	4	5	6	7	8	9	10	11	12	13	14
15	0.7933	0.3838	0.4084	0.5841	0.5070	0.4780	0.2975	0.2844	0.4128	0.2844	0.4094	0.3222	0.3345	0.3466
16	0.4685	0.5910	0.5793	0.7371	0.7663	0.3194	0.6576	0.4850	0.5669	0.3705	0.5866	0.5356	0.4864	0.4654
17	0.5013	0.5921	0.5804	0.7378	0.7669	0.3213	0.6988	0.5772	0.5681	0.4563	0.6260	0.5711	0.5181	0.4357
18	0.4900	0.6186	0.6217	0.6140	0.5425	0.3343	0.7043	0.7035	0.6196	0.6046	0.5920	0.6992	0.5072	0.4870
19	0.6634	0.6853	0.5838	0.6150	0.6578	0.4770	**0.8916**	0.6738	0.6862	0.7825	0.6634	0.7794	0.5663	0.6614
20	0.6260	0.2950	0.2950	0.4712	0.4125	0.6268	0.3288	0.3787	0.3779	0.3269	0.5725	0.4748	0.4320	0.5700
21	0.4209	0.5772	0.6376	0.7282	0.6376	0.4244	0.6013	0.5791	**0.8081**	0.3336	0.5294	0.5019	0.4563	0.3811
22	0.4016	0.5181	0.5663	0.7289	0.7591	0.4260	0.6386	0.7100	**0.8087**	0.4563	0.5886	0.5221	0.4940	0.5109
23	0.3267	0.2872	0.2872	0.5990	0.5248	0.4968	0.4796	0.4781	0.6226	0.4781	0.5960	0.3824	0.4796	0.5034
24	0.3267	0.2872	0.2872	0.4654	0.4060	0.4339	0.4796	0.4781	0.4968	0.4346	0.4165	0.2942	0.3494	0.5034
25	0.4053	0.4954	0.5459	0.7162	0.6216	0.6661	0.5080	0.5605	**0.8664**	0.4873	0.7297	0.7477	0.7351	0.6130
26	0.4229	0.5112	0.4868	0.7251	0.6334	0.6119	0.4135	0.5743	**0.8059**	0.3969	0.7435	0.7164	0.7801	0.6453
27	0.4220	0.3925	0.4168	0.7266	0.6355	0.7427	0.4168	0.5062	0.7427	0.4356	0.6940	0.6324	0.7084	0.6922
28	0.2179	0.2103	0.1981	0.3166	0.3925	0.4854	0.2710	0.4003	0.5497	0.3651	0.5240	0.4653	0.4533	0.5212

(18 out of 386) and 0.0389 (15 out of 386) respectively, as shown in Table 3.4. Of the 15 who had clinical damage in the MCP joint in the middle finger, 14 of these individuals also had the MCP joint in the index finger damaged. This led to a dependence ratio of 20.015 (Table 3.14) and a proportion of max of 0.930 (Table 3.15). As a comparison, consider the corresponding two-way dependence ratio for the DIP joints in the right hand. The marginal probabilities for these joints are 0.1088 (42 out of 386) and 0.0933 (36 out of 386) respectively. However, out of the 36 who had the DIP joint in the middle finger clinically damaged, only 27 also had the DIP joint in the index finger damaged. This led to a lower dependence ratio of 7.148 (Table 3.14) and a lower proportion of max value of 0.751 (Table 3.15).

3.6 Modelling the data: The dependence ratio approach

This section presents modelling of the data using the dependence ratio approach that was discussed in detail in Section 3.2. Recall that the approach puts equal emphasis on the marginal regression (discussed in Sections 3.3 and 3.4) and the associations (discussed in Section 3.5). The exploratory analysis of the associations in Section 3.5 showed the differing association patterns of clinical damage for the MCP knuckle joints in comparison to other joints. In addition, the GEE analysis of the marginal patterns of clinical damage in Section 3.4 revealed significant interactions involving the joint type variable and consequently made interpretations with regards to the marginal patterns of clinical damage more difficult. One of the main aims of the dependence ratio approach is to provide a model that is easy to understand with a relatively small number of parameters for maximum likelihood estimation so consequently having a large number of

parameters for the interactions as in Section 3.4 is not sought after. It is noted that conclusions relating to the inclusion of a large number of interaction parameters was based upon GEE methodology. However, this may not be the case for the dependence ratio approach.

In Section 3.5, the association patterns of clinical damage involving only the 8 MCP knuckle joints were shown to be the strongest. Consequently, this section treats these 8 joints separately to the other joints. Although Section 3.5 revealed no clear association patterns among the other 20 joints, the PIP, DIP and thumb joints will be considered separately in order to see how their association structures differ from the MCP knuckle joints. In other words, the association patterns in Section 3.5 suggest that the MCP thumb joint should be treated separately from the other MCP knuckle joints so the thumb is treated separately. It should be noted that treating the joint type variable as a four level variable (thumb, MCP, PIP and DIP) in the GEE approach led to the same conclusions as when it was considered as a three level variable (MCP, PIP and DIP) in Section 3.4. Having these four separate models allows for a detailed assessment of the DIP joints, since these joints are often of particular interest in psoriatic arthritis. In addition, the DIP joints are the closest to the nails and known to be the most commonly affected (as shown in Sections 3.3 and 3.4). In addition, the thumb is of particular interest given the results in Section 3.4 and the fact it has no DIP joint. It should be noted that these models do not allow for an assessment of the associations between joint type, but these were assessed in an exploratory nature in Section 3.5.

In Section 3.4, the high proportion of patients with no damage in any of their 28 joints was advocated as a strong justification for the use of the necessary factor association structure that was described in Section 3.2.1. This is further shown by Table 3.20 which shows the proportion of individuals with no damaged joints, for each of the four joint types. Although they all have large proportions, the MCP knuckle joints have the largest proportion with 92% of patients having no damage in any of these 8 joints.

Table 3.20: Proportion of individuals with no damaged joints for each model

Joint type	Proportion no damaged joints (sample size)	Number of joints
MCP knuckle	0.920 (3088)	8
thumb	0.852 (1544)	4
PIP	0.806 (3088)	8
DIP	0.769 (3088)	8

Separate models for each of the four joint types in Table 3.20 are now presented, using the dependence ratio approach. Clinical damage is the response of interest. The following explanatory variables were considered for the marginal regression model (as specified in detail in Section 3.4) in each case: sex, age (in years) of diagnosis with psoriatic arthritis, hand, finger and time (in years) since diagnosis. Since the thumb has both MCP and PIP joints, an additional variable was included in this model to account for this. The

same procedure was followed as in the GEE approach in the sense that only variables that significantly predict the probability of clinical damage were included in each case with the exception of sex which was included regardless of its significance. Due to issues with the convergence of the parameter estimates and negative profile probabilities, the time since diagnosis covariate was treated as a categorical variable with 3 levels (0-10 years (baseline), 10-20 years, over 20 years).

Given the high proportions of individuals with no joints damaged in Table 3.20, the necessary factor association structure is considered for the association structure in each model as well as other exchangeable structures described in Section 3.2.1 (L and NL). In the case of the necessary factor association structure, the marginal regression model will be estimated conditional on the patient being susceptible to damage. This is therefore more similar to the second GEE analysis of Section 3.4 although this approach uses all patients (for a given joint type) without excluding those with no damaged joints.

Since the exchangeable association structures are not nested, the AIC is used to compare the fit of the combined regression and association models in the subsequent sections. In addition to the AIC, the selection of the final model is also based upon the interpretability of the association model as well as the fit of the profile probabilities. For example, even though a model may give a superior fit in terms of the AIC, it may produce negative profile probabilities for the unobserved profiles, which is unsatisfactory. The likelihood ratio test was used to confirm any conclusions obtained using the Wald test with regards to the regression coefficients (using a 5% significance level).

3.6.1 Model for the MCP knuckle joints

An analysis of the explanatory variables assuming independence between and within patients (irrespective of the association structure) found that finger and time since diagnosis significantly predict the probability of damage in the MCP knuckle joints (joint numbers 2-5 and 16-19). These two variables along with sex were included as explanatory variables for the regression model. Hand and age of diagnosis were not significant so were not included in the final model. There were no significant interactions.

Since 92% of individuals had no damage in any of these joints (as shown in Table 3.20), the necessary factor structure described in Section 3.2.1 was fitted as a starting point for modelling the association structure. The same conclusions were obtained from assessing the regression model under this structure as under independence. This model gave a much superior fit compared to independence (taking into account the regression model). The AIC's of independence and the necessary factor were 860.288 and 526.812 respectively. Although the latent binary association structure gave a slightly better fit, the necessary factor structure was chosen due to its more convenient interpretation. The

combined necessary and latent binary structure (NL) produced negative profile probabilities for the unobserved profiles so was not considered. Having the necessary factor varying by gender was considered but this model did not give an improvement (AIC = 528.223). In other words, there is negligible difference between males and females being susceptible to clinical damage. This is not surprising given that males and females did not follow different association patterns in terms of the observed dependence ratios in Section 3.5.

The parameter estimates for the chosen necessary factor association structure combined with the chosen regression model are shown in Table 3.21. Recall that the regression coefficients are estimated conditional on patients having the necessary factor for damage in the MCP knuckle joints. Table 3.22 shows a comparison of the AIC's for the association structures considered (taking into account the chosen regression model).

Table 3.21: Parameter estimates for the MCP knuckle joints model

Sub model	Parameter	Estimate	Standard error
Regression model	Intercept	−2.950	0.749
	Sex (female)	0.460	0.305
	Finger (middle)	−0.273	0.370
	Finger (ring)	−1.273	0.390
	Finger (little finger)	−1.273	0.390
	Time since diagnosis (10-20 years)	2.750	0.745
	Time since diagnosis (over 20 years)	3.505	0.737
Association model	$v = pr(N = 1)$	0.103	0.018

Table 3.22: Comparison of model fit for association structures of the MCP knuckle joints using AIC

Association Structure	AIC
Independence (I)	860.288
Necessary factor (N) by gender	528.223
Necessary Factor (N)	526.812

The regression model shows that females are associated with a greater probability of clinical damage than males in the MCP knuckle joints (although not significantly). The odds of damage for females are 1.584 times greater than for males (holding other variables constant). The finger variable has only 4 levels since the thumb is not present in this model. The ring and little fingers are (significantly) associated with lower probabilities of damage than the index finger (the baseline). The middle finger is associated with a lower probability of damage than the index finger (although not significantly). Despite the categorisation of the time since diagnosis variable, the same conclusion is found as the GEE analysis in the sense that increased time since diagnosis is (significantly) associated with a higher probability of damage.

In terms of interpreting the association model, the probability of a patient having the factor necessary for clinical damage in the MCP knuckle joints is given by the parameter v and is estimated to be 0.103. In other words, the proportion of patients that are not susceptible to clinical damage in their MCP knuckle joints is given by $1 - v$ which is approximately 0.90.

Despite the MCP knuckle joints having low marginal probabilities compared to other joints as well as the largest proportion of patients with no damage in any of their joints, they still showed the strongest association pattern of all joints in Section 3.5. In other words, if a patient has one of these MCP knuckle joints damaged, there is a strong chance they will have other MCP knuckle joints damaged also. This was shown in the analysis of the observed two-way dependence ratios in Section 3.5. For the necessary factor association model, all w-way dependence ratios are equal. Using (3.8), the 2-way dependence ratio along with its standard error in parentheses (calculated using the delta method, see Appendix) is estimated to be 9.679 (1.732) for the MCP knuckle joints. This estimate is larger than the estimates for the models later in this section so this model therefore takes into account the higher dependence of clinical damage between these MCP knuckle joints. However, the larger standard error shows there is more uncertainty with this estimate. The 95% confidence interval is (6.28,13.07).

3.6.2 Model for the thumb joints

Since the MCP thumb joints followed a different association pattern to the MCP knuckle joints, it was decided that the four thumb joints (two MCP joints: joint numbers 1 and 15 and two PIP joints: joint numbers 6 and 20) would be modelled separately.

The effect of the explanatory variables on clinical damage was assessed assuming independence within and between patients. An additional joint type variable was included to take into account the fact that the thumb has both MCP and PIP joints present. Time since diagnosis and joint type significantly predicted the probability of clinical damage. These two variables along with sex were included as explanatory variables. Hand and age were non-significant and there were no significant interactions.

An assessment of the association structure of the thumb joints showed that there was limited difference between the necessary factor and latent binary factor association structures in terms of AIC (taking into account the chosen regression model) but the necessary factor association structure was chosen due to its more straightforward interpretation. In addition, Table 3.20 showed that approximately 85% of patients had no damage in any of their thumb joints, thus advocating the use of the necessary factor association structure. The combined necessary and latent structure (NL) produced negative profile probabilities so was not considered. The necessary factor varying by gender was also considered but this did not yield a significant improvement. Table 3.23 shows the

parameter estimates for the combined regression and association model and Table 3.24 shows a comparison of the AIC's for the association structures considered (taking into account the chosen regression model).

Table 3.23: Parameter estimates for the thumb joints model

Sub model	Parameter	Estimate	Standard error
Regression model	Intercept	−3.208	0.559
	Sex (female)	0.011	0.329
	Joint type (MCP)	1.438	0.282
	Time since diagnosis (10-20 years)	1.787	0.566
	Time since diagnosis (over 20 years)	2.844	0.563
Association model	$v = pr(N = 1)$	0.2009	0.0274

Females are associated with a greater probability of clinical damage than males in the thumb joints but the effect is negligible. The MCP thumb joint is (significantly) associated with a greater probability of clinical damage than the PIP joint (the baseline). The odds of clinical damage in the MCP thumb joint is 4.21 times greater than in the PIP thumb joint (holding other variables constant). This is not surprising given the results from the GEE analysis in Section 3.4. Time since diagnosis is again significant with an increase associated with a greater probability of clinical damage.

The association model shows that the probability of an individual having the necessary factor for clinical damage in the thumb joints is 0.2009. This shows that patients are more susceptible to damage in the thumb joints than the MCP knuckle joints considered previously. Consequently, the probability of an individual not being susceptible to damage in the thumb joints is approximately 0.8. Using (3.8), the estimated two-way dependence ratio with associated standard error (in parentheses) for the thumb joints model is 4.977 (0.679). The 95% confidence interval is (3.65, 6.31).

Table 3.24: Comparison of model fit for association structures of the thumb joints using AIC

Association structure	AIC
Independence (I)	762.586
Necessary factor (N) by gender	595.531
Necessary factor (N)	**593.638**
Latent binary factor (L)	592.289

3.6.3 Model for the PIP joints

The 8 PIP joints (joint numbers 7-10 and 21-24) are now considered. A regression analysis (independently of the association structure) revealed significant effects of time since diagnosis and finger. However, in contrast to previous models, sex was found to be significant with females again associated with higher probabilities of damage than males. This is not surprising given the results for the GEE analysis in Section 3.4. Hand and age of diagnosis were again non-significant and there were no significant interactions.

Table 3.20 showed that approximately 81% of individuals had no damage in any of these 8 joints (less than both the MCP knuckle and thumb joints). The necessary factor association structure gave a superior fit compared to independence (taking into account the chosen regression model). However, both the latent binary association structure (L) and the combined necessary/latent binary association structure (NL) both produced negative profile probabilities. Following the exclusion of the time since diagnosis variable, a suitable model was obtained using the latent binary association structure. Since interest lies more in the finger and sex variables and the fact this model gave a much superior fit over the necessary factor with all regression variables present, this was consequently chosen as the final model. Table 3.25 shows the parameter estimates for the chosen combined regression and association model and Table 3.26 shows a comparison of the AIC for the different association structures (taking into account the chosen regression model).

Table 3.25: Parameter estimates for the PIP joints model

Sub model	Parameter	Estimate	Standard error
Regression model	Intercept	-2.700	0.187
	Sex (female)	0.414	0.154
	Finger (middle)	0.043	0.118
	Finger (ring)	-0.384	0.152
	Finger (little finger)	-0.177	0.131
Association model	$v2 = pr(L = 1)$	0.103	0.017
	κ	0.027	0.005

Females are (significantly) associated with a higher probability of clinical damage than males in the PIP joints. The odds of clinical damage are 1.51 times greater for females than males (holding other variables constant). The ring finger is (significantly) associated with a lower probability of clinical damage than the index finger (the baseline).

In terms of interpreting the association model, it can be concluded that there are two groups of individuals in the population that each have different probabilities of clinical damage (given the same covariate values). The probability of being in the latent group 0 is given by 0.897 ($= 1 - 0.103$). The κ parameter shows that the probability of clinical

damage for those in latent group 0 is only 2.7% of the probability of those in latent group 1. In other words, there appears to be two distinct group with one having a much higher probability of clinical damage than the other. Applying the results to (3.10) gave an estimated two-way dependence ratio (with standard error in parentheses, see Appendix) for this model of 6.46 (0.72). The 95% confidence interval is (5.05, 7.87).

Table 3.26: Comparison of model fit for association structures of the PIP joints using AIC

Association structure	AIC
Independence (I)	1547.532
Necessary factor (N)	1155.542
Latent binary factor (L)	**1067.298**

3.6.4 Model for the DIP joints

The final model considered is for the 8 DIP joints that are located closest to the fingertip. These joints are commonly affected in psoriatic arthritis, as discussed in previous sections. However, results in Section 3.5 showed that although the DIP joints had large marginal probabilities at each individual joint in comparison to other joints, their associations were not as strong as the MCP knuckle joints.

A regression analysis under independence revealed similar conclusions to those observed in previous models with time since diagnosis and finger being significantly associated with the probability of clinical damage. These variables were included in the regression model along with sex. Hand and age of diagnosis were non-significant and there were no significant interactions.

An assessment of the association structures found the latent binary structure to be the most appropriate. It gave a much superior fit compared to the necessary factor (taking the regression model into account). The NL structure was not chosen since it consistently produced negative profile probabilities. Table 3.27 shows the parameter estimates for this model and Table 3.28 shows the comparison of model fit.

The regression model shows that females are again associated with a greater probability of damage, but not significantly. The odds of clinical damage are 1.20 times greater for females than males (holding other variables constant). Both the ring and little fingers were associated with significantly lower probabilities of clinical damage than the index finger. The conclusions for the time since diagnosis variable remain the same.

The association model can be interpreted in the same way as for the PIP joints. There are two different groups in the population with different probabilities of clinical damage (given the same regression values). The latent group 0 accounts for 89.2% of the population and of those in this group, the probability of clinical damage is 4.3% of the

Table 3.27: Parameter estimates for the DIP joints model

Sub model	Effect	Estimate	Standard error
Regression model	Intercept	−2.833	0.398
	Sex (female)	0.185	0.102
	Finger (middle)	−0.100	0.103
	Finger (ring)	−0.251	0.114
	Finger (little finger)	−0.211	0.097
	Time since diagnosis (10-20 years)	0.502	0.417
	Time since diagnosis (over 20 years)	0.834	0.411
Association model	$v2 = pr(L = 1)$	0.108	0.019
	κ	0.043	0.007

probability of damage in the latent group 1. Applying (3.10) to this model shows that the estimated two-way dependence ratio (with standard error in parentheses) is 5.09 (0.54). The 95% confidence interval is given by (4.04, 6.15).

Table 3.28: Comparison of model fit for association structures of the DIP joints using AIC

Association structure	AIC
Independence (I)	1793.226
Necessary factor (N)	1375.683
Latent binary factor (L)	**1247.991**

3.7 Discussion

The dependence ratio approach for (population-averaged) maximum likelihood estimation of multivariate categorical data provides an intuitive way of modelling the presence of clinical damage in patients with psoriatic arthritis. The approach puts equal emphasis on the marginal regression and the associations within a patient. In addition, it can cope computationally with larger cluster sizes than the odds ratio (Lessafre et al. 2000 states that the odds ratio struggles with cluster sizes that are larger than approximately five). The regression model uses population-averaged coefficients for the the explanatory variables and the approach also allows for relevant association structures to be devised that adequately summarise the association structure using only a small number of parameters.

The focus of the GEE approach is on the marginal regression with the associations of secondary interest. Given that the associations are of strong interest for analysing clinical damage in patients with psoriatic arthritis, the dependence ratio approach therefore provides a more comprehensive analysis than the GEE approach. The GEE approach is arguably the most commonly used method for analysing multivariate categorical data. This is largely due to the fact that it is implemented in many statistical software such as SPSS but also because it can also cope with much larger cluster size than the odds ratio.

The GEE approach was used in Section 3.4 in order to assess the marginal patterns of clinical damage across all 28 joints, irrespective of the associations. This was undertaken for all patients as well as for those with at least one joint damaged with the same conclusions obtained in each case. An analysis of the main effects showed that time since diagnosis (in years), finger and joint type were the explanatory variables that significantly predict clinical damage. The DIP joints had significantly greater probabilities of clinical damage than the MCP and PIP joints. This coincides with previous research since the DIP joints are closest to the nail and are known to be commonly affected in psoriatic arthritis. Hand and age at diagnosis were consistently non-significant. Although the main effect of sex was non-significant, females were associated with greater probabilities of damage than males and it was retained in the model for theoretical reasons.

An analysis of the interactions revealed a significant interaction between each of finger and sex with joint type. The interaction between sex and joint type was due to the differing patterns of males and females with regards to the PIP joints. Previous research has acknowledged that females tend to progress quicker to clinical damage than males (Gladman et al. 2005) even though males and females are equally likely to be affected (diagnosed) with psoriatic arthritis. However, previous research has not pinpointed the PIP joints. It would be of interest to see if this holds true in other studies.

Although the GEE approach provided some useful conclusions in terms of the marginal patterns of clinical damage, it gave no insights in to the association structure and consequently missed some notable conclusions. Previous research on clinical damage in the hand joints of patients with psoriatic arthritis has largely focused on symmetrical associations (e.g. Cresswell and Farewell 2010). Symmetry is generally regarded as less common in psoriatic arthritis than in rheumatoid arthritis (Gladman et al. 2006). In contrast, the dependence ratio approach presented in Sections 3.5 and 3.6 considered both symmetrical and non-symmetrical patterns of clinical damage.

The exploratory analysis of the association structure that was conducted in Section 3.5 showed that all the dependence ratios were much larger than one, therefore indicating that independence was clearly not applicable for modelling the association structure. The strongest association patterns of clinical damage involved the 8 MCP knuckle joints and in particular within a hand rather than between hands. These joints are the largest joints and previous research has stated that these joints are associated with clinical damage in psoriatic arthritis (O'Keeffe et al. 2011). However, the results of Section 3.5 found the within-hand association patterns of clinical damage involving these joints to be stronger than the asymmetrical and symmetrical patterns. Despite the symmetric dependence ratios all being much larger than 1, the symmetric patterns were not as strong as the within-hand patterns concerning the MCP knuckle joints. No clear association patterns emerged amongst the other 20 joints in Section 3.5. Although the DIP and PIP joints had larger marginal probabilities at each joint, the MCP knuckle joints had the stronger associations. In other words, given that a patient has a joint of each type damaged, they are most likely to have another MCP knuckle joint damaged. In addition, the two hands were found to follow similar association patterns. In other words, the two hands were found to follow similar patterns in terms of both the marginal probabilities at each joint (as demonstrated in the GEE approach) and the association patterns.

In terms of modelling the dataset using the dependence ratio approach, it was decided that four separate models would be considered. Firstly, the MCP knuckle joints were analysed separately due to the strong association patterns they conveyed in Section 3.5. Although no clear patterns were found amongst the DIP, PIP and thumb joints, they were analysed separately for theoretical reasons. In addition, the DIP joints are often of particular interest in psoriatic arthritis since they are commonly affected. Hence, an individual analysis of these joints was of interest. It is acknowledged that these models do not allow for an assessment of the associations between joint type. However, these were assessed in an exploratory nature in Section 3.5.

The conclusions for the four regression models were similar. Time since diagnosis (in years) and finger were consistently significant for predicting clinical damage. However, sex was only significant in the PIP joints model with females associated with having a

significantly higher probability of clinical damage than males. Hand and age at diagnosis were consistently non-significant. In other words, the conclusions obtained for the marginal regression models coincided with the GEE analyses of Section 3.4. However, the conclusions for the four association models differed as expected. The necessary factor association structure was found to be most appropriate for the MCP knuckle and thumb models. This is not surprising given the fact that 92% and 85% of patients respectively had no damage in any of these joints. However, a latent binary association structure was found to be the most appropriate association structure for modelling the PIP and DIP joints. A potential explanation for this is that these structures had lower proportions of patients with no damaged joints (81% and 77% respectively), consequently causing there to be two groups with different susceptibilities to damage (as opposed to the necessary factor structure which only has one group susceptible).

Appendix

Proof of the delta method used for calculating standard errors

In section 3.6, the delta method was used to calculate the standard errors for the maximum likelihood estimates of the (two-way) dependence ratio parameter ($\hat{\tau}$) in each of the four models.

Maximum likelihood estimates are known to be asymptotically normal. For a parameter η:

$$\hat{\eta} \sim N(\eta, I(\eta)^{-1}),$$

where $I(\eta)^{-1}$ is the variance of $\hat{\eta}$ which is also the inverse of the Fisher information. The delta method states that for any function of η, denoted $g(\eta)$, whose derivative $g^{'}(\eta)$ exists:

$$g(\hat{\eta}) \sim N(g(\eta), I(\eta)^{-1}[g^{'}(\eta)]^2).$$

The delta method estimator of the variance of $g(\hat{\eta})$ is found by substituting in the maximum likelihood estimator of η in to the above.

For the necessary factor association structure that was used for the MCP knuckle and thumb joints models in Section 3.6, τ is a function of a single parameter (v), given by $\tau = g(v) = \frac{1}{v}$. Although the standard error of $\hat{v}$ was naturally available from drm, the standard error of $\hat{\tau}$ was not. Applying the delta method yields the following estimate for the variance of $\hat{\tau}$:

$$I(\hat{\eta})^{-1}[g^{'}(\hat{v})]^2 = \frac{I(\hat{\eta})^{-1}}{\hat{v}^4}.$$

The standard error of $\hat{\tau}$ is therefore found by square rooting the above expression.

For the latent binary association structure that was used for the PIP and DIP joint models in Section 3.6, τ is a function of two parameters (v_2, κ). It is discussed in Jokinen (2006: PhD thesis) and is given by:

$$\tau = g(v_2, \kappa) = \frac{v_2 + (1 - v_2)\kappa^2}{(v_2 + (1 - v_2)\kappa)^2}.$$

Although the standard errors of $\hat{v}_2$ and $\hat{\kappa}$ were naturally available from drm, the multivariate delta method was required to calculate the standard error of $\hat{\tau}$.

The multivariate delta method states that $\hat{\tau}$ is asymptotically normal with mean τ and variance given by:

$$\nabla(v_2, \kappa) I(v_2, \kappa)^{-1} \nabla(v_2, \kappa)^T,$$

The delta method estimator is found by substituting in the maximum likelihood estimators in to the above equation.

Note: $\nabla(v_2, \kappa) = (\frac{\partial g}{\partial v_2}, \frac{\partial g}{\partial \kappa})$ and $I(v_2, \kappa)^{-1}$ is given by:

$$\begin{pmatrix} var(v_2) & cov(v_2, \kappa) \\ cov(v_2, \kappa) & var(\kappa) \end{pmatrix}.$$

In addition:

$$g^{'}(v_2) = \frac{[v_2 + (1 - v_2)\kappa]^2[1 - \kappa^2] - 2[v_2 + (1 - v_2)\kappa^2][1 - \kappa][v_2 + (1 - v_2)\kappa]}{[v_2 + (1 - v_2)\kappa]^4},$$

$$g^{'}(\kappa) = \frac{[v_2 + (1 - v_2)\kappa]^2[2\kappa(1 - v_2)] - 2[v_2 + (1 - v_2)\kappa^2][1 - v_2][v_2 + (1 - v_2)\kappa]}{[v_2 + (1 - v_2)\kappa]^4}.$$

Chapter 4

Paper 3: Prior and posterior distributions for dependence ratios

4.1 Introduction

Multivariate categorical data are found in a range of disciplines such as clinical trials and epidemiology. In order to follow good statistical practice, traditional univariate regression techniques cannot be used for analysis of such data since they assume independence. However, multivariate techniques can be used to deal with this problem by treating each unit's (cluster's) response profile as the response. The analysis of a multivariate categorical response is more complicated than the traditional multivariate normal response since the first two moments do not completely specify the joint distribution (unless there are only two units in a cluster).

Papers 1 and 2 of this thesis focused on analysing multivariate categorical data with the dependence ratio association measure (Ekholm, Smith and McDonald, 1995). The dependence ratio was proposed as an advantageous alternative to the odds ratio. The work in papers 1 and 2 was performed using traditional frequentist approaches, with maximum likelihood estimation. However, this paper focuses on using the dependence ratio for likelihood-based methods in a Bayesian context. As in papers 1 and 2, population-averaged models are used as opposed to random-effects or conditional models. The dependence ratio is discussed in detail in Section 4.1.1.

A requirement of all likelihood approaches for multivariate categorical data is that the joint distribution must be specified. For population-averaged models, the first-order moments are treated in the same way (regardless of which association measure is considered for the second-order and higher moments) since they are regressed on any relevant explanatory variables using an appropriate link function. However, the approaches differ in their parameterisations of the second-order and higher moments.

4.1.1 Association measures for likelihood-based approaches

In addition to the dependence ratio, this section also discusses the correlation coefficient and the commonly used odds ratio in order to give a comparison to the dependence ratio and because of their relevance in later sections.

To illustrate the three association measures, consider the bivariate binary response $\boldsymbol{Y} = (Y_1, Y_2)$. The first-order moments (marginal probabilities) are specified as follows (assuming no explanatory variables):

$$\mu_k = pr(Y_k = 1), \tag{4.1}$$

where $k = 1, 2$.

Although the second-order moment of the bivariate binary case represents the association between the two random variables (Y_1 and Y_2), it gives no link to the baseline of independence whereas the three association measures discussed in this section do relate to independence. The second-order moment is given by:

$$\mu_{12} = pr(Y_1 = 1, Y_2 = 1). \tag{4.2}$$

There are four possible response profiles: (1,1), (0,0), (1,0) and (0,1). Consider a sample of size n from $\boldsymbol{Y}$=(Y_1, Y_2). This can be classified in a 2 by 2 contingency table with cell counts $\boldsymbol{n} = (n_{00}, n_{01}, n_{10}, n_{11})$ and corresponding cell probabilities $\boldsymbol{\pi} = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$. Under the multinomial sampling scheme, the likelihood function for $\boldsymbol{\pi}$ is given by:

$$L(\boldsymbol{\pi}; \boldsymbol{n}) \propto \pi_{00}^{n_{00}}\ \pi_{01}^{n_{01}}\ \pi_{10}^{n_{10}}\ \pi_{11}^{n_{11}}. \tag{4.3}$$

The odds ratio is the most commonly used association measure for analysing multivariate categorical data. It has a number of different formulations such as the conditional log odds ratio by Fitzmaurice and Laird (1993). However, for the bivariate binary case, the same result is obtained regardless of the chosen formulation. This is given by:

$$\chi = \chi_{12} = \frac{\mu_{12}(1 - \mu_1 - \mu_2 + \mu_{12})}{(\mu_2 - \mu_{12})(\mu_1 - \mu_{12})} = \frac{pr(Y_1 = 1, Y_2 = 1)pr(Y_1 = 0, Y_2 = 0)}{pr(Y_1 = 0, Y_2 = 1)pr(Y_1 = 1, Y_2 = 0)}. \tag{4.4}$$

The odds ratio has a range from 0 to infinity with 1 representing independence. Positive association is indicated by values greater than 1 and negative association by values less than 1. Typically iterative methods are required for specifying the joint distribution in

terms of the marginal probabilities and the odds ratio parameter(s). In fact, even for the simplest bivariate binary case, specification of the joint distribution in terms of μ_1, μ_2 and χ requires the solving of a quadratic. From (4.4), it can be shown (for $\chi \neq 1$) that:

$$\mu_{12} = \frac{\chi(\mu_1 + \mu_2) + (1 - \mu_1 - \mu_2) - \sqrt{[\chi(\mu_1 + \mu_2) + (1 - \mu_1 - \mu_2)]^2 - 4(\chi - 1)\mu_1\mu_2\chi}}{2(\chi - 1)}. \tag{4.5}$$

The correlation coefficient has also been proposed as an association measure for multivariate categorical data by Bahadur (1961). For the bivariate binary response, it is given by (as specified in Albert and Gupta, 1983a):

$$\rho = \rho_{12} = \frac{\mu_{12} - \mu_1\mu_2}{R}, \tag{4.6}$$

where $R = \sqrt{\mu_1(1 - \mu_1)\mu_2(1 - \mu_2)}$.

The range of the correlation coefficient for multivariate categorical data is not between -1 and 1 since it is constrained by the marginal probabilities. However, under marginal homogeneity with the marginal probabilities set equal to 0.5 ($\mu_1 = \mu_2 = 0.5$), the range is between -1 and 1. This point is discussed again in Section 4.2 when discussing the work of Albert and Gupta (1983a). The joint distribution can be specified in terms of the marginal probabilities and the correlation coefficient with the following transformation:

$$\mu_{12} = \rho R + \mu_1\mu_2. \tag{4.7}$$

The final association measure to discuss is what forms the basis of this paper. The dependence ratio was introduced by Ekholm, Smith and McDonald (1995). It was originally proposed for the purpose of maximum likelihood estimation and as a solution to the fact that the commonly used odds ratio could not deal with large cluster sizes. For the bivariate binary case, the two-way dependence ratio is given by:

$$\tau = \tau_{12} = \frac{\mu_{12}}{\mu_1\mu_2} = \frac{pr(Y_1 = 1, Y_2 = 1)}{pr(Y_1 = 1)\, pr(Y_2 = 1)}. \tag{4.8}$$

It is similar to the odds ratio in the sense that 1 represents independence with values greater than 1 representing positive association and values less than 1 representing negative association. If $Y_k = 1$ represents a success, the dependence ratio is interpreted as the joint probability of success divided by the joint probability of success under independence. Dependence ratios for higher-order moments can be expressed similarly. Lesaffre et al. (2000) state that likelihood-based estimation of population-averaged models with odds ratios is typically not feasible for cluster sizes that are greater than five. In contrast, in the dependence ratio approach, the likelihood functions can be expressed in closed form in terms of the marginal probabilities and dependence ratios of all orders which

gives the dependence ratio a key advantage over the odds ratio for maximum likelihood estimation (see Sections 1.2.4 and 1.3 of the introduction to this thesis for more details).

The cell probabilities from (4.3) can therefore be reparameterised in terms of (μ_1, μ_2, τ). The likelihood function for this case is given by:

$$\begin{aligned} L(\mu_1, \mu_2, \tau; \boldsymbol{n}) &\propto (1 - \mu_1 - \mu_2 + \mu_1\mu_2\tau)^{n_{00}} \ (\mu_2 - \mu_1\mu_2\tau)^{n_{01}} \ (\mu_1 - \mu_1\mu_2\tau)^{n_{10}} \ (\mu_1\mu_2\tau)^{n_{11}} \\ &= (1 - \mu_1 - \mu_2 + \mu_1\mu_2\tau)^{n_{00}} \ (1 - \mu_1\tau)^{n_{01}} (1 - \mu_2\tau)^{n_{10}} \mu_1^{n_{10}+n_{11}} \mu_2^{n_{01}+n_{11}} \tau^{n_{11}}. \quad (4.9) \end{aligned}$$

Consequently, models are relatively easy to fit using maximum likelihood estimation. Datasets with cluster sizes larger than five can be analysed with dependence ratios. For example, paper 1 of this thesis analysed a rater agreement study with seven subunits (pathologists) in a cluster (Landis and Koch, 1977). In addition, see papers 1 and 2 as well as Ekholm (2003) for details of the advantages and disadvantages of the dependence ratio relative to the odds ratio.

The range of the dependence ratio is of particular relevance in the current paper. It is similar to the correlation coefficient in the sense that it is constrained by the marginal probabilities. In other words, τ is not variation independent of μ_1 and μ_2. The correlations of the parameter estimates also need to be monitored since the marginal probabilities and dependence ratios are not orthogonal. The range of τ is given by:

$$\max \left\{0, \frac{1}{\mu_1} + \frac{1}{\mu_2} - \frac{1}{\mu_1\mu_2}\right\} \leq \tau \leq \min \left\{\frac{1}{\mu_1}, \frac{1}{\mu_2}\right\}. \quad (4.10)$$

Although the dependence ratio was not proposed for the purpose of modelling multivariate categorical data until Ekholm et al. (1995), the dependence ratio formulation was used prior to this (and since) in different application areas with different names for the dependence ratio (see the introduction to this thesis for a history of the dependence ratio). It should be noted that with the the exception of Good (1956) and Du Mouchel and Pregibon (2001), other uses of the dependence ratio were in a exploratory nature rather than for modelling purposes. In addition, it is believed that all other uses of the dependence ratio were in a frequentist setting other than Good (1956) and Du Mouchel and Pregibon (2001). The work of Ekholm et al. (1995) was extended by Ekholm, Smith and McDonald (2000) amongst others (see Section 1 of paper 1 for more details). The literature review in Section 2 of the current paper discusses Good (1956) further. It should be noted that the approaches by Good (1956) and Du Mouchel and Pregibon (2001) were empirical Bayes approaches, something which is not of interest in this work. This is because interest lies in developing suitable prior distributions for the parameters rather than estimating them from the data. Consequently, Bayesian approaches to multivariate categorical data using dependence ratios and traditional Bayesian approaches

as opposed to empirical Bayes approaches are not thought to have been undertaken before.

The current paper makes the first steps towards Bayesian estimation of the models proposed by Ekholm et al. (1995). The outline of the paper is as follows. In Section 4.2, a literature review of Bayesian methods for categorical data is given. Section 4.3 considers the simplest case of a bivariate binary response with relevant prior distributions considered for the marginal probabilities and the two-way dependence ratio. The resulting posterior distributions for datasets are then discussed. In addition, prior and posterior distributions for the case of marginal homogeneity (the two marginal probabilities are constrained to be equal) are also considered. Section 4.4 discusses possible extensions to include covariate information. Section 4.5 draws appropriate conclusions and discusses extensions to larger contingency tables. For all the prior distributions considered, emphasis is on non-informative priors for both the marginal probabilities and dependence ratio(s) to reflect ignorance about the location of these parameters. By non-informative, it is meant that the prior distributions should at least have little impact on the posterior inference. However, a discussion is given in Section 4.3.4 as to how informative priors could be incorporated.

4.2 Literature Review

This section gives a review of Bayesian approaches to analysing categorical data with particular emphasis on papers that are relevant to the work of this paper. The review is done largely chronologically. Similar to the review paper by Agresti and Hitchcock (2005), emphasis of this discussion is mainly on contingency table analysis. The focus of the current paper is on estimating cell probabilities in contingency tables under the multinomial sampling scheme. Loglinear approaches and Dirichlet priors are discussed but are not of interest for this paper since emphasis is on specifying prior information directly for the marginal probabilities and dependence ratio(s). This point is discussed further in due course.

4.2.1 Early Bayesian approaches to categorical data

The starting point is the work by Bayes (1763) and Laplace (1774) who were both interested in estimating a single binomial parameter p. For this case, they used the conjugate beta prior. If Y denotes a binomial random variable with n trials and parameter p, the conjugate beta prior for p, denoted $beta(a, b)$, is proportional to:

$$p^{a-1}(1-p)^{b-1}, \tag{4.11}$$

where a and $b > 0$.

Both Bayes and Laplace used a uniform prior distribution for estimating p. In this case a and b are both set equal to 1 and this prior is often referred to as the Bayes-Laplace prior. Perks (1947) and Fisher (1956) were critical of this prior. For example, Fisher noted that uniform priors on different scales would give different results. To combat this difficulty, Jeffreys (1946) proposed a prior which is invariant to reparameterisations of the parameter(s). This prior was consequently known as the Jeffreys prior (Jeffreys, 1946) and in the binomial case it corresponds to $a = b = 0.5$. Haldane (1948) considered the improper beta prior in which $a = b = 0$. The posterior mean for this case is equivalent to the maximum likelihood estimator and Haldane argued in favour of this prior for work in genetics where $\log(p)$ is thought to be roughly uniform for p near 0. The Bayes-Laplace, Jeffreys and Haldane priors can all be naturally extended to the multinomial case using a Dirichlet prior distribution.

The focus of Good's work in the 1950s and 1960s was largely in using Bayesian estimation to smooth cell counts in contingency tables where the data are sparse. Good (1956) smoothed the cell counts in large sparse two-way contingency tables by using log-normal and gamma priors in order to estimate association factors (dependence ratios) for each cell in the table. As discussed in Section 4.1, the approach was an empirical Bayes approach since the prior parameters are iteratively estimated from the data with the marginal probabilities being set equal to their observed values. The iterative process for determining the prior parameters was based upon finding the sample mean and variance of the non-zero sample dependence ratios. Good (1956) mentions that this method may well be natural to those who are familiar with the logarithm of the association factor, also known as the pointwise mutual information (PMI). In addition, Good (1956) discusses that the association factor was called the "coefficient of dependence" (COD) by Keynes and Johnson (Keynes, 1921). Both the PMI and the COD are discussed in the history of the dependence ratio section in the introduction to this thesis. Another point of particular interest is that Good (1956) allowed the prior distribution of the association factor to take values outside the range (4.10) and argued that this did not greatly affect the final posterior inference. In the current paper, emphasis is given to constraining the prior distribution of the dependence ratio to be between its upper and lower bounds from (4.10) with empirical Bayes approaches not considered. However, log-normal and gamma priors are considered for the dependence ratio in Section 4.3 with relevant constraints to the range applied. Good (1956) also states the lower bound of the two-way dependence ratio to be zero as opposed to the lower bound given in (4.10).

Other work by Good includes his 1967 paper in which he considered priors that were suitable for estimating multinomial probabilities in significance tests. Good (1967) considered the natural extension of the binomial distribution to the multinomial distribution with c categories as opposed to contingency tables. Suppose there are counts n_i which

follow a multinomial distribution with corresponding probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_c)$; $i = 1, \ldots, c$. For this case, the conjugate Dirichlet prior for $\boldsymbol{\pi}$ is given by:

$$p(\boldsymbol{\pi}) \propto \prod_{i=1}^{c} \pi_i^{\alpha_i - 1}, \tag{4.12}$$

where $\alpha_i > 0$.

Good focused on Dirichlet priors in which $\alpha_1 = \alpha_2 = \cdots = \alpha_c = \alpha$ and he referred to these as symmetric Dirichlet priors. The Bayes-Laplace ($\alpha = 1$), Jeffreys ($\alpha = 0.5$) and Haldane priors ($\alpha = 0$) described previously are all examples of symmetric Dirichlet priors. Perks (1947) also suggested a symmetric prior in which α is $\frac{1}{c}$. Good did not regard one-stage Dirichlet priors as sufficiently flexible for use in significance tests so he instead considered hierarchical priors in which α is assigned a second stage distribution. He considered an improper prior for α, given by $\frac{1}{\alpha}$ and often referred to as the Haldane-Jeffreys density (Good, 1967). Good argued that approximating this improper prior by a proper distribution was more appropriate, with the log-Cauchy distribution being one such example that was considered. Hierarchical Dirichlet priors are discussed further in the literature review of Albert and Gupta's work in the 1980s. Although hierarchical Dirichlet priors are no longer of conjugate form, they allow for greater flexibility (Agresti and Hitchcock, 2005).

The notable work of Lindley (1964) focused on Bayesian inference using odds ratios. Consider the case of a two-way contingency table with cell probabilities π_{ij}. In addition, suppose the corresponding cell counts n_{ij} have a single multinomial distribution. Lindley considered the posterior distribution of the log contrast $\lambda = \sum_{i,j} c_{ij} \log(\pi_{ij})$, where $\sum_{i,j} c_{ij} = 0$. Consider the simplest case of a 2 by 2 contingency table with cell probabilities $\boldsymbol{\pi} = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$. One possible log contrast is the log-odds ratio, which is given by:

$$\log(\pi_{11}) + \log(\pi_{00}) - \log(\pi_{10}) - \log(\pi_{01}). \tag{4.13}$$

The log-odds ratio is commonly used for measuring the association between categorical variables, one reason being that it is symmetric with regards to success and failure. Lindley used the conjugate Dirichlet prior density for $\boldsymbol{\pi}$, which is given by:

$$p(\boldsymbol{\pi}) \propto \prod_{i,j} \pi_{ij}^{\alpha_{ij} - 1}, \tag{4.14}$$

where $\alpha_{ij} > 0$.

Lindley consequently showed that the posterior distribution of λ is approximately normal with mean $\sum_{i,j} c_{ij} \log(\alpha_{ij} + n_{ij})$ and variance $\sum_{i,j} c_{ij}^2 (\alpha_{ij} + n_{ij})^{-1}$. He then used this

approximation to determine an approximate posterior distribution for the log-odds ratio in 2 by 2 tables. Lindley used the improper Dirichlet prior in which all the α_{ij} are set equal to zero. He argued in favour of this prior by noting that the smaller the values of these prior parameters, the less prior information there is about π_{ij}. Despite the improper nature of the prior, this was not an issue in this context since approximations were being used (with sufficiently large sample sizes) and the posterior distributions were proper. Novick (1969) also argued in favour of improper priors. The posterior distribution of the log-odds ratio in a 2 by 2 table is therefore approximately normal with mean $\log(\frac{n_{11}n_{00}}{n_{10}n_{01}})$ and variance $\sum_{i,j} n_{ij}^{-1}$. Improper priors are not considered in Sections 4.3 and 4.4. Notable reasons for this are to avoid the potential for improper posterior distributions and because the BUGS software used for posterior inference in this paper does not allow for the use of improper priors.

The asymptotic behaviour of the log-odds ratio has also been examined by a number of other authors in recent years. For example, Bloch and Watson (1967), Altham (1969), and Fredette and Angers (2002). In contrast to Lindley (1964), the advances in computational methods in recent years (as discussed at the end of this literature review) allow for posterior inference to be based on simulating samples from the posterior distribution rather than relying on approximations for large samples.

Leonard (1972) considered an alternative prior distribution for estimating several independent binomial parameters (p_i). Instead of the conjugate beta prior, he considered a two-stage logistic-normal distribution on p_i. Although this approach was used briefly by other researchers in the 1960s, such as Cornfield (1966), it was was not strongly promoted until the work by Leonard (1972). At stage 1, each $\text{logit}(p_i)$ is assumed to follow a $N(\mu, \sigma^2)$ distribution. At stage 2, the mean μ is assumed to follow an improper uniform prior over the real line and $\frac{v\lambda}{\sigma^2}$ has a chi-squared distribution on v degrees of freedom (independently of μ). This corresponds to an inverse chi-squared distribution on σ^2. In addition, v and λ are specified by the user, where λ is a prior estimate of σ^2 and v represents the sureness of this. Since the current paper does not focus on estimating several independent binomial parameters, this approach will not be discussed further. However, Agresti and Hitchcock (2005) discussed the more simple case of estimating a single binomial parameter (p) using a logistic-normal distribution, which is of more relevance to this work. They discuss the use of a $N(0, \sigma^2)$ prior distribution for $\text{logit}(p)$ and the implications of the choice of σ^2 on the prior for p. For example, when σ^2 is approximately equal to 2.5, the prior distribution for p is approximately uniform except near the boundaries and is therefore reasonably non-informative. The user therefore specifies the value of σ^2 based on their prior beliefs. Sections 4.3 and 4.4 of this paper consider a logistic-normal prior distribution for the marginal probabilities in contingency tables with $\mu = 0$ and $\sigma^2 = 2.5$. The logistic-normal distribution for a given marginal probability (p) is proportional to:

$$e^{-\frac{(\text{logit}(p)-\mu)^2}{2\sigma^2}}\frac{1}{p(1-p)}, \tag{4.15}$$

where μ and σ^2 are the mean and variance of logit(p).

When $\mu = 0$ and $\sigma^2 = 2.5$, (4.15) therefore simplifies to:

$$e^{-\frac{1}{5}(\text{logit}(p))^2}\frac{1}{p(1-p)}. \tag{4.16}$$

Leonard (1973) considered a multivariate extension of the logistic-normal distribution from his 1972 paper. The approach uses a multivariate normal prior distribution for multinomial logits and consequently gives a multivariate logistic-normal distribution for the multinomial parameters. Leonard proposed this prior distribution as an advantageous alternative to the Dirichlet prior for the multinomial parameters. He stated that despite the fact that the Dirichlet prior allows for the specification of the prior means and variances, it does not allow for correlations to be altered whereas the multivariate logistic-normal prior does allow such flexibility. This prior distribution is particularly relevant to multivariate categorical data since the user may have a prior belief that the marginal probabilities are correlated, making this prior more appropriate than assuming prior independence. When there are only two dimensions, this prior can be referred to as the bivariate logistic-normal distribution.

All of the approaches discussed so far concentrated on specifying prior distributions in terms of probabilities. At this point, a brief discussion is given for the alternative loglinear approaches for I by J contingency tables despite the fact they are not considered in this work. Leonard (1975) focused on the saturated loglinear model for two-way tables, which is given by:

$$\log[E(n_{ij})] = \lambda + \lambda_i^1 + \lambda_j^2 + \lambda_{ij}^{12}, \tag{4.17}$$

where $E(n_{ij})$ represents the expected cell count for a particular cell in row i and column j; $i = 1, \ldots, I; j = 1, \ldots, J$.

Leonard (1975) assumed prior independence and therefore separate priors for each of the row, column and interaction effects. He followed a similar approach to Leonard (1972) in the sense that he considered a two-stage approach in which each of the row, column and interaction effects were given their own $N(\mu, \sigma^2)$ distribution at stage 1. Similarly, at stage 2 each mean was assumed to have an improper uniform distribution over the real line and each σ^2 was assumed to have an inverse chi-squared distribution. However, prior independence of the loglinear parameters has not always been assumed. For example, Knuiman and Speed (1988) generalised Leonard (1975) by considering a multivariate normal prior for all parameters collectively, in multi-way tables. Loglinear

approaches will not be considered in this paper since specifying prior information about the marginal probabilities and association parameters appears difficult with the loglinear parameterisation (Albert and Gupta, 1983a).

Albert and Gupta wrote a number of papers in the 1980s. Their 1982 paper focused on hierarchical Dirichlet priors for estimating multinomial cell probabilities in 2 by 2 contingency tables. For this case, the conjugate Dirichlet prior density for $\boldsymbol{\pi} = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$ is given by:

$$p(\boldsymbol{\pi}) \propto {\pi_{00}}^{\alpha_{00}-1}\ \pi_{01}^{\alpha_{01}-1}\ \pi_{10}^{\alpha_{10}-1}\ \pi_{11}^{\alpha_{11}-1}. \tag{4.18}$$

This can be expressed equivalently in terms of the parameters $\gamma_{11}, \gamma_{10}, \gamma_{01}$ and K, where $K = \alpha_{00} + \alpha_{10} + \alpha_{01} + \alpha_{11}$. The γ parameters represent the prior means for the corresponding cell probabilities and are found by dividing the corresponding α parameter by K, where K reflects the sureness of these prior means. Albert and Gupta (1982) consider hierarchical Dirichlet priors using this form as opposed to (4.18). They state that hierarchical Dirichlet priors (first introduced by Good, 1967) provide a natural way of enlarging the class of Dirichlet distributions. At stage 1, the user selects a formulation that they believe for the prior means along with a value for K. Larger values of K represent more belief in the particular formulation. Albert and Gupta (1982) considered the prior beliefs of symmetry and independence for the prior means. For the symmetry formulation, $\gamma_{10} = \gamma_{01}$ and for independence, $\gamma_{11} = \gamma_1\gamma_2$, where $\gamma_1 = \gamma_{11} + \gamma_{10}$ and $\gamma_2 = \gamma_{11} + \gamma_{01}$. At stage 2, a non-informative uniform distribution was placed on the prior means to reflect prior ignorance.

Albert and Gupta (1983a) again consider 2 by 2 tables and note that a Dirichlet prior distribution does not contain enough parameters to enable separate prior information to be specified for the marginal probabilities and associations. Consequently, the Dirichlet priors from Good (1967) and Albert and Gupta (1982) will not be considered in Sections 4.3 and 4.4. Albert and Gupta (1983a) adopted a similar approach to the current paper in the sense that the joint distribution is specified in terms of the marginal probabilities and an association measure of interest. However, they considered the odds ratio and the correlation coefficient for the association as opposed to the dependence ratio. In this discussion, focus is on their use of the correlation coefficient (ρ) since it is of more relevance to the work of the current paper.

The cell probabilities from (4.3) can also be reparameterised in terms of (μ_1, μ_2, ρ), as shown in Albert and Gupta (1983a), where ρ is as specified in (4.6). The likelihood function for this case is given by:

$$L(\mu_1, \mu_2, \rho) \propto (1 - \mu_1 - \mu_2 + \rho R + \mu_1\mu_2)^{n_{00}}\ (\mu_2 - \mu_1\mu_2 - \rho R)^{n_{01}}\ (\mu_1 - \mu_1\mu_2 - \rho R)^{n_{10}}\ (\mu_1\mu_2 + \rho R)^{n_{11}}. \tag{4.19}$$

Albert and Gupta (1983a) suggested a prior for ρ that is conditional on μ_1 and μ_2. This conditional prior was based on a convenient parameterisation of the likelihood function from (4.19) and is given by:

$$P(\rho \mid \mu_1, \mu_2) \propto (1-\mu_1-\mu_2+\rho R+\mu_1\mu_2)^{v_{00}}\ (\mu_2-\mu_1\mu_2-\rho R)^{v_{01}}\ (\mu_1-\mu_1\mu_2-\rho R)^{v_{10}}\ (\mu_1\mu_2+\rho R)^{v_{11}}. \quad (4.20)$$

Although Albert and Gupta (1983a) considered beta prior distributions for μ_1 and μ_2, they instead focused on the case where μ_1 and μ_2 were set equal to 0.5. Under this case, $\frac{\rho+1}{2}$ has a beta $(v_{00}+v_{11}+1, v_{01}+v_{10}+1)$ distribution. A similar result is obtained when the dependence ratio is considered instead of ρ. It can be shown that $\frac{\tau}{2}$ has a beta $(v_{00}+v_{11}+1, v_{01}+v_{10}+1)$ distribution when $\mu_1=\mu_2=0.5$. This consequently allows prior information about ρ to be inputted through a standard distribution. In addition, by setting the marginal probabilities equal to 0.5, ρ has the same favourable range as it does for normally distributed responses since it is between -1 and 1. However, constraining the marginal probabilities in this way may not be sensible in some applications since it is often preferred to allow the marginal probabilities to take values uniformly over the entire $(0,1)$ range when no prior information is available. In addition, it can be seen from (4.10) that constraining $\mu_1=\mu_2=0.5$ leads to τ being constrained between only 0 and 2, which is not ideal. Consequently, this approach will not be considered as a conditional prior distribution for τ in Sections 4.3 and 4.4.

4.2.2 Computational advances in Bayesian approaches to categorical data

The Bayesian approaches discussed in Section 4.2.1 were often hindered by not being able to calculate the posterior distribution in situations where the prior is not conjugate. In addition, these approaches typically had to rely on large sample approximations for the posterior distribution (such as Lindley, 1964) or calculating relevant posterior summaries such as the posterior mode rather than being able to sample from the posterior distribution. However, the computational advances of the last 30 years have helped greatly to rectify these problems, some of which are now discussed.

The first computational breakthrough to mention is the importance sampling Monte Carlo simulation that was proposed by Zellner and Rossi (1984). With the focus on binary regression models, Zellner and Rossi aimed to determine suitable posterior densities for the regression parameters β. They consequently proposed their importance sampling method as an alternative to existing methods which typically ran into difficulties. For example, numerical integration was sometimes difficult for high dimensional integrals (Agresti and Hitchcock, 2005). The importance sampling method was shown to be more efficient for obtaining a sufficient approximation to the posterior distribution

of β. Zellner and Rossi (1984) considered both improper uniform distributions on β as well as some informative priors. Gelfand and Smith (1990) proposed Gibbs sampling for sampling from complex multiparameter distributions. Gibbs sampling is a Markov chain Monte Carlo (MCMC) method which samples iteratively from the conditional (usually single parameter) distributions. Some or all of these conditional distributions may not follow a standard distribution. In which case, an algorithm such as Metropolis-Hastings may be used, which is discussed in Tierney (1994). Other methods have been proposed such as adaptive rejection sampling by Gilks and Wild (1992).

Recall that the focus of this paper is on estimating parameters in contingency tables as opposed to developing regression models such as in Zellner and Rossi (1984). This will be performed using the BUGS and R software. In order to relate this discussion on computational advances to these software, consider the process that WinBUGS uses. For each parameter, if the full conditional (continuous) distribution is not of a known distribution, direct simulation cannot be used and alternative methods are sought. Adaptive rejection sampling is used provided the conditional distribution is log-concave, otherwise slice sampling (Neal, 2003) is used provided there is a restricted range, with Metropolis-Hastings being used if the range is unrestricted. Both WinBUGS and OpenBUGS are used for posterior inference in this paper, for reasons discussed in Section 4.3.2.

Model selection methods have also seen some notable contributions for Bayesian approaches. Raftery (1986) proposed using Bayes factors instead of p-values and noted that -2 times the log of the Bayes factor is approximately equal to the Bayesian Information Criterion (BIC) proposed by Schwarz (1978). Spiegelhalter et al. (2002) suggested the deviance information criterion (DIC). An advantage of the DIC over Akaike's Information Criterion (AIC), proposed by Akaike (1973), and the BIC is that it is easily obtained from MCMC samples. The DIC is therefore useful in Bayesian model selection, is readily available in the BUGS software and will be used for choosing between potential models in the examples of Sections 4.3 and 4.4.

4.2.3 Other recent developments

Square tables refer to contingency tables with the same row and column categories so are consequently relevant to the analysis of multivariate categorical data with two units in a cluster. They were discussed in detail (from a frequentist point of view) in paper 1 of this thesis. Paper 1 discussed models for square tables such as marginal homogeneity and symmetry. Some of the main Bayesian contributions to square tables are now briefly discussed. Vounatsou and Smith (1996) considered square contingency tables and focused on intraclass tables, something which was not discussed in Paper 1 of this thesis. Intraclass tables refer to square tables where the cells for the same pair of categories are not distinguishable. Forster (2004a) discussed how to construct invariant prior distributions for the parameters of a square table, using a log-linear approach.

Although intraclass tables and log-linear approaches are not considered in the work of the current paper, traditional square tables are considered for the example datasets in Section 4.3 since they are a natural occurrence in multivariate categorical data with only two observations in a cluster. As in paper 1 of this thesis, models such as marginal homogeneity and symmetry are considered with suitable non-informative priors for the model parameters (marginal probabilities and dependence ratios) proposed. Marginal homogeneity refers to the equality of the marginal distributions in terms of probability whereas symmetry refers to the joint probabilities of a particular set of two categories being equal regardless of their order. As discussed in paper 1, marginal homogeneity and symmetry are equivalent in 2 by 2 contingency tables.

4.3 Bayesian inference for the two-way dependence ratio

This section considers the bivariate binary response $\boldsymbol{Y} = (Y_1, Y_2)$ that was discussed in Section 4.1 and the corresponding 2 by 2 contingency table for the cell probabilities ($\boldsymbol{\pi} = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$) with emphasis on obtaining non-informative prior distributions for the marginal probabilities (μ_1 and μ_2), and the two-way dependence ratio (τ). Ideally, the prior distributions will therefore assign values uniformly over all possible values of the parameter in question. However, a minimum requirement is that the prior distributions chosen have little or no impact on the posterior inference. Section 4.3.1 focuses on the prior distributions that will be considered for μ_1, μ_2 and τ, taking into account the literature review of Section 4.2. Section 4.3.2 discusses the methods and software used for posterior inference as well as considering datasets in order to estimate the resulting posterior distributions. Section 4.3.3 considers the prior distributions for the special case of marginal homogeneity, where $\mu_1 = \mu_2$, as well as posterior inference for this case. It is of interest to determine whether non-informative priors differ under marginal homogeneity to under the saturated model. Section 4.3.4 discusses how informative priors could be incorporated.

4.3.1 Prior Distributions

To enable Bayesian estimates of μ_1, μ_2 and τ, a prior distribution for a parameterisation of the cell probabilities ($\boldsymbol{\pi}$) must be chosen. As discussed in the literature review of Section 4.2, the natural Dirichlet prior (4.18) will not be considered since it only enables prior information to be specified for the cell probabilities as opposed to the marginal probabilities and the two-way dependence ratio. This paper focuses on specifying prior distributions directly for μ_1, μ_2 and τ. In contrast to Good (1956), emphasis is on constraining the dependence ratio to be between its upper and lower bounds from (4.10). In other words, the prior distributions considered for τ are conditional on μ_1 and μ_2.

Recall from (4.9) that the likelihood function for $\boldsymbol{\pi}$ can be specified in terms of μ_1, μ_2 and τ. The 2 by 2 contingency table of the cell probabilities for this case is shown in Table 4.1.

Table 4.1: Cell probabilities for a bivariate binary response

	Y_2		
Y_1	1	0	Total
1	$\tau\mu_1\mu_2$	$\mu_1 - \tau\mu_1\mu_2$	μ_1
0	$\mu_2 - \tau\mu_1\mu_2$	$1 - \mu_1 - \mu_2 + \tau\mu_1\mu_2$	$1 - \mu_1$
Total	μ_2	$1 - \mu_2$	1

Prior distributions can then be specified directly on these parameters. Prior independence is assumed for the marginal probabilities throughout although in situations where substantive prior information is available, the informative priors that are discussed in Section 4.3.4 are preferred. Given the marginal probabilities range from 0 to 1, uniform$(0, 1)$ priors are an obvious choice to consider as non-informative priors for μ_1 and μ_2. In addition, taking into account the discussion of Agresti and Hitchcock (2005) from Section 4.2.1, independent logistic-normal priors are also considered for each of the marginal probabilities. For this case, Normal$(0, \sigma^2)$ priors are assigned to the logit of each marginal probability, with $\sigma^2 = 2.5$ used for this case since this leads to acceptable non-informative priors for μ_1 and μ_2.

A series of different prior distributions are considered for τ, each of which is constrained to be within the range of τ given in (4.10). A few key points were noted when considering potential prior distributions for τ. Firstly, it was important to consider prior distributions that at least account for a large proportion of the possible values for τ (pre-truncation) since otherwise the simulation algorithms will be inefficient. Secondly, it is noted that some of the common non-informative distributions considered from the statistical literature may turn out to be overly informative after truncation. In such cases, it was decided that these distributions would be excluded from the posterior inference. A useful feature of WinBUGS and OpenBUGS is that constraints can easily be put on the ranges of distributions by using the in built 'I(lower,upper)' constraint, where lower and upper represent the lower and upper bounds that are specified by the user. Since τ cannot take negative values, only conditional priors that can take non-negative values will be considered for τ. For example, the log-normal and gamma priors used by Good (1956) will be considered since they satisfy this requirement. Prior distributions of the form given in (4.21) are therefore considered:

$$p(\tau, \mu_1, \mu_2) \propto p(\tau|\mu_1, \mu_2)p(\mu_1)p(\mu_2). \tag{4.21}$$

Table 4.2 outlines the 10 prior formulations that were considered for μ_1, μ_2 and τ. The software R was used for simulating from these prior formulations as opposed to BUGS

(used for posterior inference). The reason for this is that BUGS is based upon the MCMC method Gibbs sampling (discussed in Section 4.2.2) which does not sample from the prior distributions exactly. In contrast, the method of factorisation can be used in R to sample from a combination of marginal and conditional distributions in order to sample from the prior distributions exactly, as shown in (4.21). In other words, after drawing μ_1 and μ_2 from their separate prior distributions, τ is then drawn from its conditional distribution given μ_1 and μ_2. For each of the uniform and logistic-normal priors assumed for the marginal probabilities, 5 conditional prior distributions were considered for τ. Firstly, prior formulations 1 and 2 involve τ being assigned a uniform distribution between the lower and upper bounds in (4.10). Prior formulations 3 and 4 relate to τ being given a gamma distribution, denoted $\Gamma(r, \omega)$, between its lower and upper limits from (4.10), as noted by 'I(lower, upper)' in Table 4.2. This distribution has a shape parameter (r) and an inverse scale parameter (ω), as given below:

$$p(\tau) \propto \tau^{r-1} e^{-\tau\omega}, \tag{4.22}$$

where $\tau > 0$.

When there is no prior knowledge, the gamma parameters are often assigned values close to zero. The $\Gamma(0.01, 0.01)$ was considered since values much smaller than this led to simulated estimates of τ consistently being outside the range of possible values from (4.10). Prior formulations 5 and 6 are the same as 3 and 4 except a $\Gamma(1, 1)$ was considered instead for τ, for reasons which will be discussed in due course.

Prior formulations 7 and 8 relate to τ following a log-normal(μ, σ) distribution. On the logarithmic scale, this prior distribution has a location parameter (μ), a scale parameter (σ) and is given by:

$$p(\tau) \propto \frac{1}{\tau} e^{-\frac{1}{2\sigma^2}(\log(\tau)-\mu)^2}, \tag{4.23}$$

where $\tau > 0$.

The parameters μ and σ relate to the mean and standard deviation of the logarithm of τ respectively. This prior for τ is also constrained between the range in (4.10). In terms of the values to assign to the log-normal parameters, μ was set at zero given that it can take any negative or positive values (on the log scale). The scale parameter σ was given a second-stage uniform(0,100) distribution (a commonly used non-informative prior). Prior formulations 9 and 10 are the same as 7 and 8 except both μ and σ were set equal to 1, the reasons for which are discussed below.

Table 4.2: Prior formulations for μ_1, μ_2, τ

Prior Formulation	Marginal Probabilities (μ_1, μ_2)	Dependence ratio (τ)
1	Both $\sim$ uniform(0,1)	$\tau \sim$ uniform(lower,upper)
2	Both $\sim$ logistic-normal(0,2.5)	$\tau \sim$ uniform(lower,upper)
3	Both $\sim$ uniform(0,1)	$\tau \sim \Gamma(0.01, 0.01)$ I(lower,upper)
4	Both $\sim$ logistic-normal(0,2.5)	$\tau \sim \Gamma(0.01, 0.01)$ I(lower,upper)
5	Both $\sim$ uniform(0,1)	$\tau \sim \Gamma(1, 1)$ I(lower,upper)
6	Both $\sim$ logistic-normal(0,2.5)	$\tau \sim \Gamma(1, 1)$ I(lower,upper)
7	Both $\sim$ uniform(0,1)	$\tau \sim$ log-normal(0, uniform(0,100)) I(lower,upper)
8	Both $\sim$ logistic-normal(0,2.5)	$\tau \sim$ log-normal(0, uniform(0,100)) I(lower,upper)
9	Both $\sim$ uniform(0,1)	$\tau \sim$ log-normal(1, 1) I(lower,upper)
10	Both $\sim$ logistic-normal(0,2.5)	$\tau \sim$ log-normal(1, 1) I(lower,upper)

For each prior formulation, $10,000$ simulations were run in R. For each simulated pair of marginal probabilities, a value is simulated from the relevant conditional distribution for τ. In cases where a simulated τ lies outside the range of (4.10), the value of τ is rejected and then resampled until it lies within (4.10). In other words, more than $10,000$ simulations of τ may be required in order to ensure that each τ consistently lies between the range in (4.10). This is only relevant to priors 3-10 since no resampling is necessary for priors 1 and 2.

The estimated density plots obtained for the marginal probabilities from the simulations were consistently similar to the theoretical density plots. In other words, they approximately favoured all values between 0 and 1 equally.

Figure 4.1 shows the prior densities for τ, for each of the prior formulations considered. The conditional prior distributions for τ consistently favoured values near the independence value of 1. However, this is reasonable when there is no prior knowledge. Consider the formula for τ in (4.8). If there is no prior knowledge, the four cell probabilities from Table 4.1 will often be assumed to be equally likely with $\mu_{12} = \tau\mu_1\mu_2 = 0.25$. In addition, if the marginal probabilities are set at their mean value of 0.5 (assuming a uniform distribution), then the denominator in (4.8) is also 0.25 and τ is consequently given by the independence value of 1. The conditional $\Gamma(0.01, 0.01)$ prior for τ in prior formulations 3 and 4 led to values near zero being favoured as well as independence, which is less desirable. This underlines the fact that prior distributions that are typically non-informative may not be once truncation has been applied. Priors 3 and 4 are consequently not considered for posterior inference in Section 4.3.2. Prior formulations 5 and 6 were also considered since the conditional $\Gamma(1, 1)$ prior led to only independence being favoured (largely). However, there still appears to be too much prior probability given to values near zero. Priors 5 and 6 are retained for posterior inference in order to assess the consequences of this. The conditional log-normal priors for τ in prior formulations 7 and 8 also favour zero as well as independence so will consequently not be considered for posterior inference in Section 4.3.2. Prior formulations 9 and 10 were considered

as an alternative log-normal conditional distribution for τ since only independence was favoured. Prior formulations $1, 2, 5, 6, 9$ and 10 will therefore be taken forward for posterior inference in Section 4.3.2. The 'From=0' command in the density function of R was used to ensure all values of τ were non-negative in Figure 4.1 since the smoother used in R gave negative values on occasions (see the R code in the Appendix).

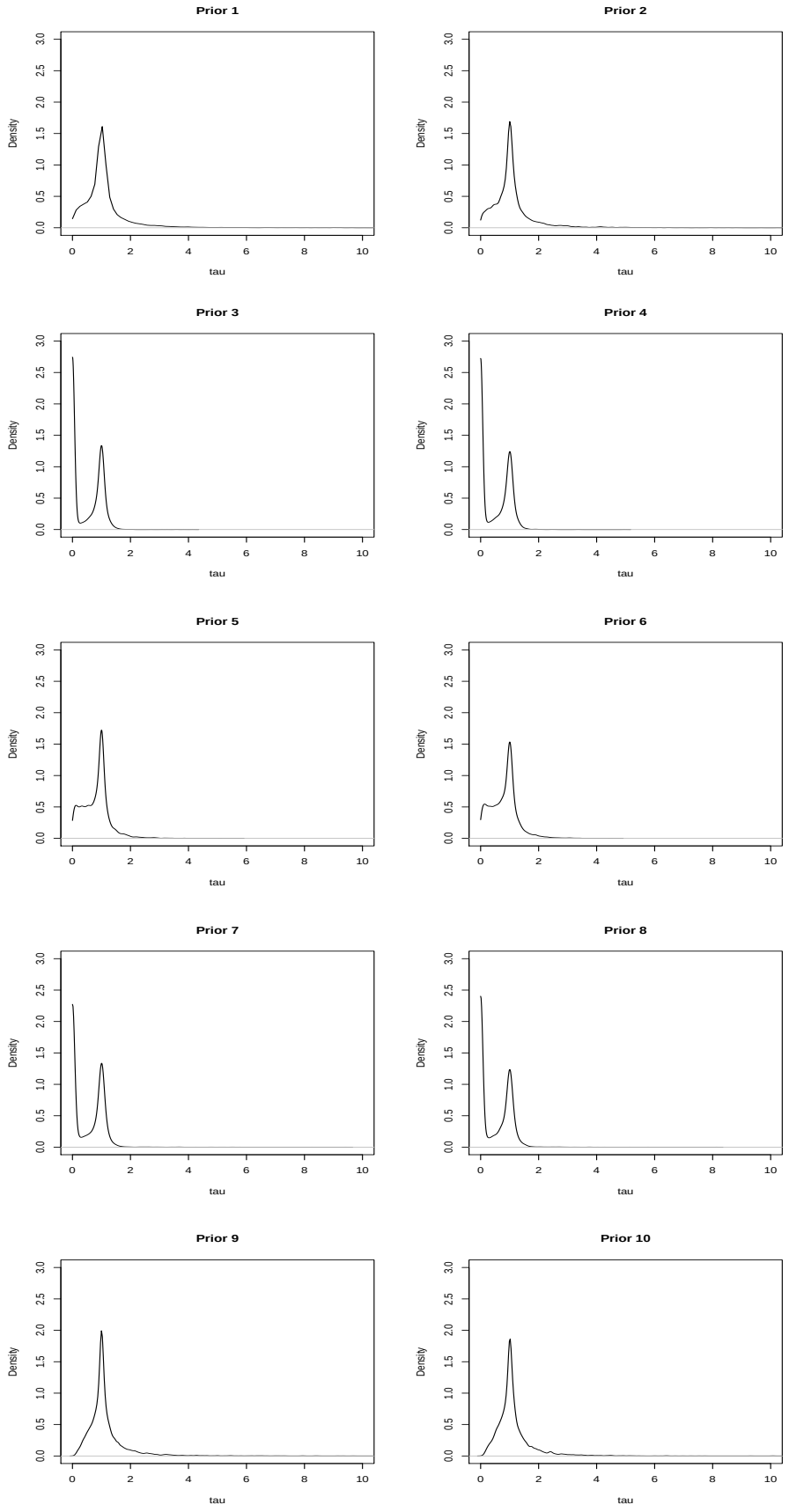

Figure 4.1: Prior distributions for τ

4.3.2 Posterior inference

The following section considers posterior inference for the 6 prior formulations taken forward from Section 4.3.1. Desirable non-informative priors are those with posterior estimates that are dominated by the data in the examples considered in Section 4.3.2.1.

The prior formulations in Section 4.3.1 did not require MCMC methods since they could be simulated from directly by simulating from a combination of marginal and conditional distributions (4.21) in R. However, once they are combined with the likelihood from (4.9), denoted $L(\tau, \mu_1, \mu_2)$, they do not give a known distributional form for the joint posterior distribution. In addition, direct sampling is not possible for posterior inference. The same is also true for the conditional distributions of each parameter. In addition, the fact τ is constrained by (4.10) adds a further complication. This contrasts with the conjugate Dirichlet distribution which allows for direct simulation from the posterior, although as discussed previously, the Dirichlet distribution is not considered due to its inability to include prior information for the marginal probabilities and the association measure of interest (dependence ratio). Table 4.3 shows the joint posterior distributions for the 6 prior formulations taken forward from Section 4.3.1.

Table 4.3: Prior formulations and corresponding joint posterior

Prior	Joint Posterior: $p(\tau, \mu_1, \mu_2 \mid \mathbf{n})$
1	$L(\tau, \mu_1, \mu_2) \times \frac{1}{upper-lower} = L(\tau, \mu_1, \mu_2) \times \frac{1}{f(\mu_1, \mu_2)}$
2	$L(\tau, \mu_1, \mu_2) \times \frac{1}{f(\mu_1, \mu_2)} \times e^{-\frac{1}{5}(\text{logit}(\mu_1))^2} \times e^{-\frac{1}{5}(\text{logit}(\mu_2))^2} \times \frac{1}{\mu_1(1-\mu_1)} \times \frac{1}{\mu_2(1-\mu_2)}$
5	$L(\tau, \mu_1, \mu_2) \times e^{-\tau}$
6	$L(\tau, \mu_1, \mu_2) \times e^{-\tau} \times e^{-\frac{1}{5}(\text{logit}(\mu_1))^2} \times e^{-\frac{1}{5}(\text{logit}(\mu_2))^2} \times \frac{1}{\mu_1(1-\mu_1)} \times \frac{1}{\mu_2(1-\mu_2)}$
9	$L(\tau, \mu_1, \mu_2) \times \frac{1}{\tau} \times e^{-\frac{1}{2}(\log(\tau)-1)^2}$
10	$L(\tau, \mu_1, \mu_2) \times \frac{1}{\tau} \times e^{-\frac{1}{2}(\log(\tau)-1)^2} \times e^{-\frac{1}{5}(\text{logit}(\mu_1))^2} \times e^{-\frac{1}{5}(\text{logit}(\mu_2))^2} \times \frac{1}{\mu_1(1-\mu_1)} \times \frac{1}{\mu_2(1-\mu_2)}$

Note: upper and lower in prior formulations 1 and 2 represent the upper and lower bounds of τ from (4.10). Both the upper and lower bounds of (4.10) are functions of both μ_1 and μ_2. Hence the uniform prior for τ in prior formulations 1 and 2 can be denoted by $\frac{1}{f(\mu_1, \mu_2)}$, where $f(\mu_1, \mu_2)$ is a function of μ_1 and μ_2.

Given the above points, posterior inference will be conducted using the MCMC methods in the BUGS software. In Section 4.2.2, the sampling methods that WinBUGS uses for sampling from the posterior was discussed. Both WinBUGS and OpenBUGS use the principal of Gibbs sampling in the sense that the full conditional distribution for each parameter is obtained before an appropriate sampling method can be chosen. In addition to the sampling methods available in WinBUGS (direct sampling, adaptive rejection sampling, slice sampling and Metropolis-Hastings), OpenBUGS also allows for various other sampling methods such as a multivariate extension of Metropolis. OpenBUGS

generally proved more useful than WinBUGS for sampling from the posterior distributions in the current paper. For example, posterior inference was sometimes found to be difficult for certain prior formulations in WinBUGS. However, WinBUGS was preferred for certain prior formulations such as in Section 4.4. Hence, both WinBUGS and OpenBUGS are used for posterior inference in Sections 4.3 and 4.4.

Slice sampling is of particular relevance to the current paper since this was the sampling method that was consistently chosen by BUGS for posterior inference in Sections 4.3 and 4.4. The key points are now discussed with the reader referred to Neal (2003) for more details. The principal of slice sampling is as follows. Suppose interest lies in sampling from the density $f(X)$ of a random variable X (which cannot be directly sampled from). Slice sampling introduces an auxiliary variable Y, such that the joint distribution of X and Y is uniform over a region $U = (x, y) : 0 < y < f(x)$. The final stage is to sample from the joint distribution of X and Y and then discard Y to obtain simulated values of X. However, it can be difficult to obtain independent draws from U. Neal (2003) proposed a univariate slice sampler for obtaining draws from $f(X)$, using the following three step procedure. Firstly, for the current value x_0, a value of y is drawn from the uniform region $(0, f(x_0))$. This consequently gives a horizontal slice, denoted $S = (x : y < f(x))$ and shown in bold in Figure 4.2.

Secondly, an interval around x_0 is constructed such that it contains all (or at least most) of the slice. Neal (2003) describes two approaches for selecting this interval. The stepping out procedure constructs an interval of size w at random around x_0, where w is an estimate of the typical width of the slice. This interval is then expanded by size w until both ends of the interval are outside the slice. Although the alternative doubling procedure also starts with an interval of size w, it is instead doubled in size until both ends of the interval are outside the slice. The final stage of the slice sampler is to sample a new value x_1 from the part of the slice that lies within the chosen interval. The simplest approach to selecting a new value x_1 is to sample values uniformly from within the given interval until a value within the slice is obtained. However, this can be inefficient if the slice does not account for a particularly large proportion of the interval (Neal, 2003). An alternative approach proposed by Neal (2003) initially samples from the selected interval as before but each time a value is selected that does not lie within the slice, this interval is shrunk until a point within the slice is found. The following subsection now applies the 6 prior formulations to datasets.

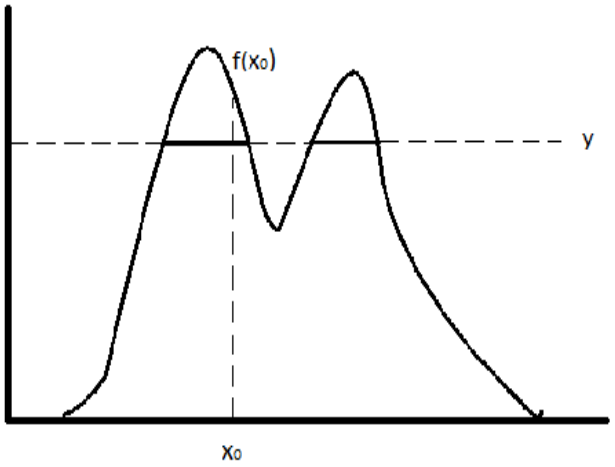

Figure 4.2: Slice Sampling

4.3.2.1 Application to datasets

The first dataset considered for posterior inference is the dataset from Braungart (1971) that was also analysed by Albert and Gupta (1983a). The dataset assesses the association between political affiliation and family structure by cross classifying 271 college students, as shown in Table 4.4.

Table 4.4: Parental decision making and political affiliation

	Political affiliation	
Parental Decision Making	Students for a democratic society (SDS)	Young Americans for freedom (YAF)
Authoritarian	29	33
Democratic	131	78

The cell probabilities corresponding to the cell counts in Table 4.4 are the same as in Table 4.1. In other words, μ_1 represents the probability of a student being an authoritarian, μ_2 represents the probability of a student belonging to SDS and τ represents the probability of a student being both authoritarian and a member of SDS divided by the corresponding probability under independence.

For this dataset, no clear prior information was available, thus justifying the use of the priors in Table 4.2. In order for the priors to be considered non-informative, posterior estimates for each of the 6 chosen prior formulations should be dominated by the data. In other words, they should be similar to a frequentist analysis of the saturated model using maximum likelihood estimation. Table 4.5 shows the estimates of the saturated model (obtained using the R package drm that is described in the introduction to this thesis). The standard errors of the marginal probabilities were not naturally available from drm since the marginal regression model parameterises in terms of the logits of probabilities. They were obtained by inverting the Fisher information matrix (see the

Appendix of paper 1 for more details). The probability of a student being authoritarian is therefore 0.2288 and the probability of a student belonging to SDS is 0.5904. The probability of a student being both authoritarian and a member of the SDS is 0.7922 times the probability under independence, thus indicating negative association.

Table 4.5: Saturated estimates from a frequentist analysis

Parameter	Value	Standard error
μ_1	0.2288	0.0255
μ_2	0.5904	0.0299
τ	0.7922	0.0946

For each of the 6 prior formulations from Table 4.3, $100,000$ simulations were run in BUGS, for each of two chains of initial values (excluding a suitable burn-in period). Convergence was assessed using relevant diagnostics that are available in BUGS such as making sure the Monte Carlo error is sufficiently low for each parameter compared to its standard deviation (5% is the typical rule of thumb). Table 4.6 shows posterior summaries for the 6 prior formulations. The 2.5, 50th and 97.5 percentiles are given along with the posterior mean and also the sampling method that was used in BUGS.

All of the prior formulations gave similar results to the saturated model from Table 4.5. However, prior formulations 1 and 2 that assume a uniform distribution for τ between the range in (4.10) gave the closest results to Table 4.5 and therefore appear the most appropriate non-informative priors to consider for this dataset at least. According to these prior formulations, the probability of a student being both authoritarian and a member of the SDS is 0.795 times the probability under independence, thus indicating negative association. Prior formulations 5 and 6 appear to be the least appropriate since the posterior estimate for τ was the furthest from the value in Table 4.5. This is perhaps not surprising given the fact too much prior weight appears to be given to values of τ near zero. Figure 4.3 shows the corresponding posterior densities. The densities appear similar for each prior formulation, which is desirable given emphasis is on obtaining non-informative priors that have little impact on the posterior inference.

Table 4.6: Posterior summaries for the 6 prior formulations

		Posterior mean	2.5	50th	97.5	Sampling method
Prior 1	μ_1	0.2299	0.1821	0.2293	0.2815	Slice sampling
	μ_2	0.5895	0.5312	0.5896	0.6468	Slice sampling
	τ	0.7947	0.6126	0.7948	0.9758	Slice sampling
Prior 2	μ_1	0.2298	0.1818	0.2291	0.2815	Slice sampling
	μ_2	0.5897	0.5310	0.5900	0.6472	Slice sampling
	τ	0.7948	0.6120	0.7952	0.9754	Slice sampling
Prior 5	μ_1	0.2300	0.1823	0.2294	0.2815	Slice sampling
	μ_2	0.5877	0.5289	0.5879	0.6449	Slice sampling
	τ	0.7853	0.6032	0.7854	0.9669	Slice sampling
Prior 6	μ_1	0.2298	0.1819	0.2291	0.2815	Slice sampling
	μ_2	0.5878	0.5287	0.5880	0.6454	Slice sampling
	τ	0.7856	0.6030	0.7857	0.9670	Slice sampling
Prior 9	μ_1	0.2302	0.1823	0.2295	0.2819	Slice sampling
	μ_2	0.5882	0.5295	0.5884	0.6458	Slice sampling
	τ	0.7969	0.6155	0.7970	0.9767	Slice sampling
Prior 10	μ_1	0.2299	0.1822	0.2291	0.2812	Slice sampling
	μ_2	0.5883	0.5297	0.5884	0.6460	Slice sampling
	τ	0.7968	0.6161	0.7969	0.9777	Slice sampling

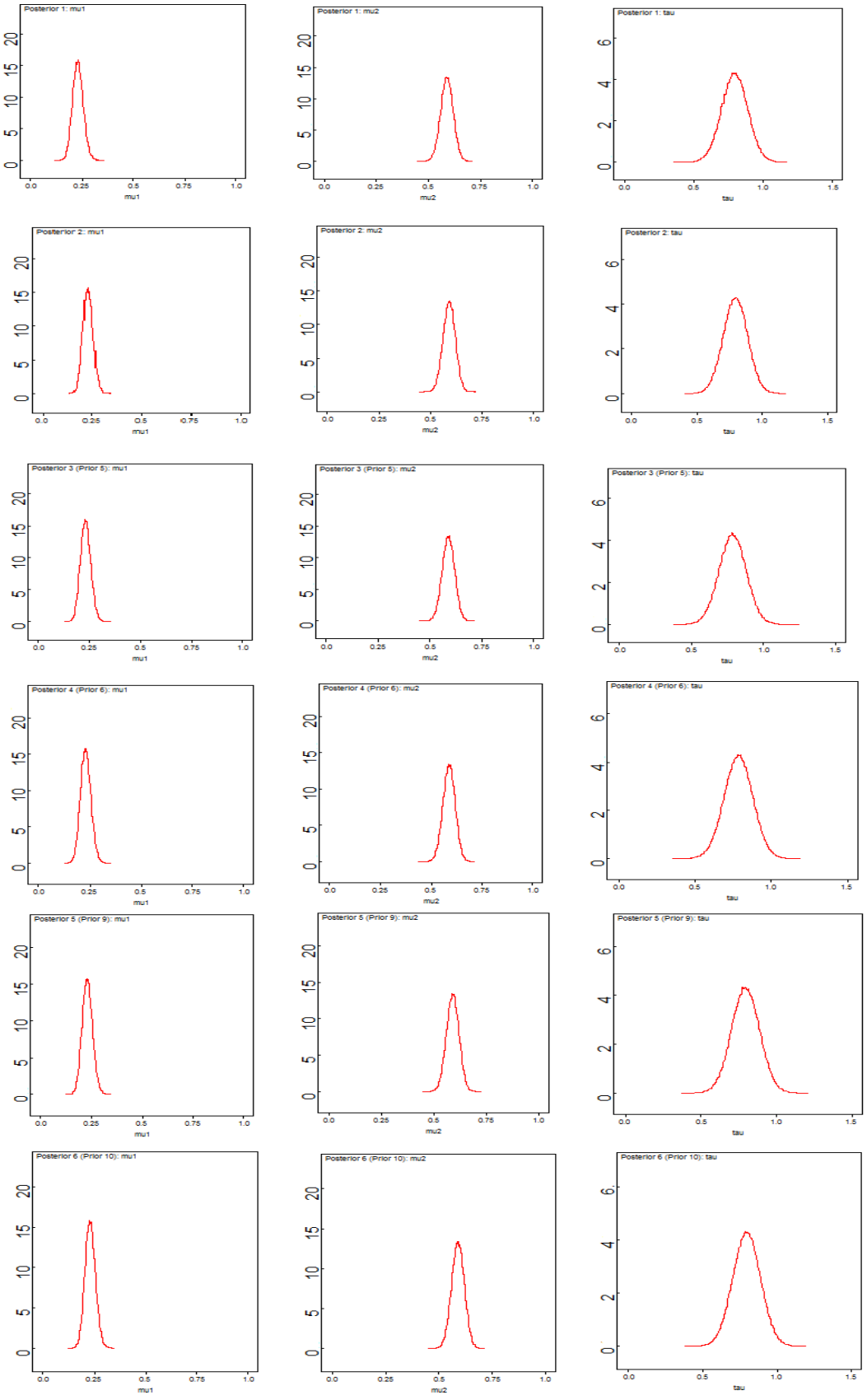

Figure 4.3: Posterior distributions for μ_1, μ_2, τ

Prior distributions are known to have a stronger influence on posterior inference as the sample size is reduced. It is of interest to determine if the 6 prior distributions taken forward for posterior inference remain approximately non-informative in smaller samples. In order to assess the robustness of these priors, the same dataset is considered but with the hypothetical situation of half the sample size (approximately). The cell counts of $29, 33, 131$ and 78 are replaced by $14, 17, 66$ and 39 respectively. The resulting estimates for a frequentist analysis of this reduced sample are $0.2279, 0.5882$ and 0.7677 for μ_1, μ_2 and τ respectively. The same process was followed as for the complete sample in the sense that $100,000$ simulations were run in BUGS, for each of two chains of initial values (excluding burn-in). Table 4.7 shows the posterior summaries. Posterior densities are not given to avoid too much repetition.

Table 4.7: Posterior summaries for the 6 prior formulations of the reduced sample

		Posterior mean	2.5	50th	97.5	Sampling method
Prior 1	μ_1	0.2302	0.1641	0.2290	0.3036	Slice sampling
	μ_2	0.5867	0.5044	0.5871	0.6665	Slice sampling
	τ	0.7737	0.5214	0.7737	1.0270	Slice sampling
Prior 2	μ_1	0.2299	0.1641	0.2285	0.3034	Slice sampling
	μ_2	0.5870	0.5043	0.5874	0.6673	Slice sampling
	τ	0.7738	0.5213	0.7740	1.0250	Slice sampling
Prior 5	μ_1	0.2304	0.1645	0.2291	0.3038	Slice sampling
	μ_2	0.5830	0.5000	0.5834	0.6634	Slice sampling
	τ	0.7559	0.5050	0.7557	1.0090	Slice sampling
Prior 6	μ_1	0.2299	0.1640	0.2285	0.3033	Slice sampling
	μ_2	0.5833	0.5066	0.5837	0.6636	Slice sampling
	τ	0.7560	0.5033	0.7556	1.0110	Slice sampling
Prior 9	μ_1	0.2306	0.1646	0.2293	0.3037	Slice sampling
	μ_2	0.5884	0.5013	0.5884	0.6646	Slice sampling
	τ	0.7790	0.5297	0.7783	1.0310	Slice sampling
Prior 10	μ_1	0.2303	0.1645	0.2290	0.3033	Slice sampling
	μ_2	0.5844	0.5015	0.5849	0.6647	Slice sampling
	τ	0.7791	0.5307	0.7785	1.0300	Slice sampling

As expected, the prior distributions have a greater effect on posterior inference in this reduced sample. Although all of the posterior estimates remain similar to the frequentist estimates, it is clear that prior formulations 1 and 2 still give posterior estimates that are closest to these and therefore appear to be the most appropriate non-informative priors of those considered. In other words, constraining τ to be uniform between the lower and upper bounds from (4.10) appears to be the most appropriate non-informative prior for τ. In contrast to the other priors for τ, this prior also guarantees a simulated τ in the range from (4.10), for every simulated pair of marginal probabilities. It also generally requires less computational time.

4.3.2.2 Simulation study

In order to give more definitive conclusions on the most appropriate non-informative priors for the marginal probabilities and dependence ratio, a simulation study was conducted. The marginal probabilities (μ_1, μ_2), dependence ratio (τ) and sample size (n) were varied to give a number of different scenarios. The chosen marginal probabilities were 0.2 and 0.4 since the dependence ratio approach has been shown to be most applicable when the marginal probabilities are less than approximately 0.5 (see Section 4.3.4 or the introduction to the thesis for more details). The chosen values for the dependence ratio were 1 (independence) and 2 with the sample sizes selected being 50, 100 and 200. This consequently gives 18 different scenarios, as shown in Table 4.8.

Following a similar approach to the work of Lambert et al. (2005), one hundred datasets were simulated for each scenario with each dataset being analysed using the six prior distributions that were chosen for posterior inference. In addition, ten thousand iterations were run for each dataset (after a burn-in of 4000 iterations). For each scenario, the prior distributions were compared in terms of how closely the mean value of the posterior median (of each parameter) matched the true value. In addition, the prior distributions were also compared on the frequentist property of bias, where bias is defined as the difference between the mean of the posterior median and the true value. Table 4.8 shows the mean values of the posterior medians and Table 4.9 shows the bias.

The results of the simulation study indicate that the estimates of the marginal probabilities are similar (and relatively unbiased) for each of the prior distributions considered in each of the simulation scenarios. However, the choice of prior distribution appears more crucial for the dependence ratio. Prior distributions 1 and 2 (which assume a uniform distribution for the dependence ratio between its upper and lower bounds) generally produce the most unbiased estimates across the different simulation scenarios, most of which being close to the true value. This coincides with the conclusions from the datasets analysed previously. However, even priors 1 and 2 can yield somewhat biased estimates when two or more of the following are apparent: small marginal probabilities, small dependence ratio or small sample size. Some entries in the above tables had to be left blank due to the fact that the gamma and log-normal priors for the dependence ratio (priors 5,6,9 and 10) were unable to produce adequate estimates. This appears to be due to these priors being unable to cope with zero counts in the simulated datasets. Given all these points, prior distributions 1 and 2 appear to be the most appropriate non-informative priors for the dependence ratio. However, the fact they have the potential to produce more biased estimates in certain extreme scenarios underlines the importance of conducting a sensitivity analysis in any Bayesian analysis in order to assess how changing the prior distributions influences the posterior estimates obtained.

Table 4.8: Simulation study: Mean of the Posterior medians for each parameter

True Parameter values	n	Parameter	Prior 1	Prior 2	Prior 5	Prior 6	Prior 9	Prior 10
$\mu1 = 0.2, \mu2 = 0.2, \tau = 1$	50	$\mu1$	0.2017	0.2014	0.1964	0.1961	0.1959	0.1954
		$\mu2$	0.2110	0.2106	0.2053	0.2049	0.2047	0.2042
		τ	1.0807	1.0789	0.8377	0.8376	1.1179	1.1182
$\mu1 = 0.2, \mu2 = 0.2, \tau = 1$	100	$\mu1$	0.2063	0.2060	0.2030	0.2028	0.2027	0.2025
		$\mu2$	0.2031	0.2026	0.2000	0.1999	0.1999	0.1997
		τ	1.0705	1.0686	0.9320	0.9318	1.0854	1.0857
$\mu1 = 0.2, \mu2 = 0.2, \tau = 1$	200	$\mu1$	0.2031	0.2030	0.2014	0.2013	0.2013	0.2012
		$\mu2$	0.1996	0.1995	0.1980	0.1980	0.1979	0.1978
		τ	1.0918	1.0906	1.0148	1.0149	1.0936	1.0927
$\mu1 = 0.2, \mu2 = 0.2, \tau = 2$	50	$\mu1$	0.2090	0.2088	0.2078	0.2073	0.2038	0.2032
		$\mu2$	0.2057	0.2054	0.2047	0.2042	0.2007	0.2003
		τ	1.9686	1.9692	1.6265	1.6278	1.8992	1.9017
$\mu1 = 0.2, \mu2 = 0.2, \tau = 2$	100	$\mu1$	0.2046	0.2042	0.2038	0.2034	0.2015	0.2009
		$\mu2$	0.2022	0.2019	0.2017	0.2011	0.1994	0.1988
		τ	1.9749	1.9772	1.7871	1.7880	1.9302	1.9344
$\mu1 = 0.2, \mu2 = 0.2, \tau = 2$	200	$\mu1$	0.2023	0.2023	0.2018	0.2017	0.2006	0.2004
		$\mu2$	0.2001	0.1999	0.1997	0.1994	0.1986	0.1983
		τ	1.9877	1.9883	1.8900	1.8910	1.9621	1.9640
$\mu1 = 0.4, \mu2 = 0.4, \tau = 1$	50	$\mu1$	0.4020	0.4018	0.3966	0.3960	0.3964	0.3960
		$\mu2$	0.3955	0.3952	0.3904	0.3902	0.3905	0.3900
		τ	1.0027	1.0027	0.9603	0.9601	1.0097	1.0103
$\mu1 = 0.4, \mu2 = 0.4, \tau = 1$	100	$\mu1$	0.3949	0.3948	0.3924	0.3923	0.3924	0.3923
		$\mu2$	0.4003	0.3999	0.3975	0.3972	0.3975	0.3973
		τ	0.9951	0.9950	0.9728	0.9725	0.9975	0.9972
$\mu1 = 0.4, \mu2 = 0.4, \tau = 1$	200	$\mu1$	0.4024	0.4024	0.4009	0.4009	0.4009	0.4009
		$\mu2$	0.3956	0.3953	0.3943	0.3943	0.3942	0.3942
		τ	1.0167	1.0166	1.0054	1.0056	1.0171	1.0173
$\mu1 = 0.4, \mu2 = 0.4, \tau = 2$	50	$\mu1$	0.4044	0.4037	0.4076	0.4068	0.3989	0.3983
		$\mu2$	0.4058	0.4056	0.4092	0.4083	0.4005	0.3998
		τ	1.9356	1.9376	1.8910	1.8946	1.9451	1.9498
$\mu1 = 0.4, \mu2 = 0.4, \tau = 2$	100	$\mu1$	0.3968	0.3963	0.3994	0.3989	0.3944	0.3939
		$\mu2$	0.3977	0.3974	0.4003	0.3997	0.3953	0.3947
		τ	1.9802	1.9816	1.9545	1.9572	1.9847	1.9880
$\mu1 = 0.4, \mu2 = 0.4, \tau = 2$	200	$\mu1$	0.4012	0.4010	0.4027	0.4025	0.4001	0.4001
		$\mu2$	0.4018	0.4014	0.4031	0.4030	0.4007	0.4006
		τ	1.9792	1.9802	1.9669	1.9678	1.9810	1.9818
$\mu1 = 0.2, \mu2 = 0.4, \tau = 1$	50	$\mu1$	0.2052	0.2051	0.2054	0.2050	0.2050	0.2046
		$\mu2$	0.4011	0.4007	0.3918	0.3916	0.3914	0.3910
		τ	1.0266	1.0265	0.9244	0.9245	1.0461	1.0460
$\mu1 = 0.2, \mu2 = 0.4, \tau = 1$	100	$\mu1$	0.1975	0.1973	0.1976	0.1972	0.1975	0.1971
		$\mu2$	0.4079	0.4076	0.4027	0.4025	0.4025	0.4025
		τ	1.0366	1.0359	0.9820	0.9816	1.0443	1.0441
$\mu1 = 0.2, \mu2 = 0.4, \tau = 1$	200	$\mu1$	0.1976	0.1976	0.1977	0.1976	0.1977	0.1976
		$\mu2$	0.4033	0.4029	0.4005	0.4004	0.4005	0.4004
		τ	0.9932	0.9929	0.9643	0.9632	0.9960	0.9952
$\mu1 = 0.2, \mu2 = 0.4, \tau = 2$	50	$\mu1$	0.2202	0.2195				
		$\mu2$	0.4133	0.4130				
		τ	1.8644	1.8644				
$\mu1 = 0.2, \mu2 = 0.4, \tau = 2$	100	$\mu1$	0.2020	0.2017	0.2039	0.2034	0.2014	0.2010
		$\mu2$	0.3965	0.3964	0.3979	0.3975	0.3932	0.3927
		τ	1.9773	1.9776	1.9220	1.9236	1.9708	1.9736
$\mu1 = 0.2, \mu2 = 0.4, \tau = 2$	200	$\mu1$	0.2000	0.1997	0.2009	0.2006	0.1997	0.1993
		$\mu2$	0.3940	0.3937	0.3949	0.3945	0.3924	0.3921
		τ	1.9941	1.9950	1.9670	1.9683	1.9908	1.9922

Table 4.9: Simulation study: Bias of each parameter

True Parameter values	n	Parameter	Prior 1	Prior 2	Prior 5	Prior 6	Prior 9	Prior 10
$\mu1 = 0.2, \mu2 = 0.2, \tau = 1$	50	$\mu1$	0.0017	0.0014	-0.0036	-0.0039	-0.0041	-0.0046
		$\mu2$	0.0110	0.0106	0.0053	0.0049	0.0047	0.0042
		τ	0.0807	0.0789	-0.1623	-0.1624	0.1179	0.1182
$\mu1 = 0.2, \mu2 = 0.2, \tau = 1$	100	$\mu1$	0.0063	0.0060	0.0030	0.0028	0.0027	0.0025
		$\mu2$	0.0031	0.0026	0.0000	-0.0001	-0.0001	-0.0003
		τ	0.0705	0.0686	-0.0680	-0.0682	0.0854	0.0857
$\mu1 = 0.2, \mu2 = 0.2, \tau = 1$	200	$\mu1$	0.0031	0.0030	0.0014	0.0013	0.0013	0.0012
		$\mu2$	-0.0004	-0.0005	-0.0020	-0.0020	-0.0021	-0.0022
		τ	0.0918	0.0906	0.0148	0.0149	0.0936	0.0927
$\mu1 = 0.2, \mu2 = 0.2, \tau = 2$	50	$\mu1$	0.0090	0.0088	0.0078	0.0073	0.0038	0.0032
		$\mu2$	0.0057	0.0054	0.0047	0.0042	0.0007	0.0003
		τ	-0.0314	-0.0308	-0.3735	-0.3722	-0.1008	-0.0983
$\mu1 = 0.2, \mu2 = 0.2, \tau = 2$	100	$\mu1$	0.0046	0.0042	0.0038	0.0034	0.0015	0.0009
		$\mu2$	0.0022	0.0019	0.0017	0.0011	-0.0006	-0.0012
		τ	-0.0251	-0.0228	-0.2129	-0.2120	-0.0698	-0.0666
$\mu1 = 0.2, \mu2 = 0.2, \tau = 2$	200	$\mu1$	0.0023	0.0023	-0.0018	-0.0017	0.0006	0.0004
		$\mu2$	0.0001	0.0001	-0.0003	-0.0006	-0.0014	-0.0017
		τ	-0.0123	-0.0117	-0.1100	-0.1090	-0.0379	-0.0360
$\mu1 = 0.4, \mu2 = 0.4, \tau = 1$	50	$\mu1$	0.0020	0.0018	-0.0034	-0.0040	-0.0036	-0.0040
		$\mu2$	-0.0045	-0.0042	-0.0096	-0.0098	-0.0095	-0.0100
		τ	0.0027	0.0027	-0.0397	-0.0399	0.0097	0.0103
$\mu1 = 0.4, \mu2 = 0.4, \tau = 1$	100	$\mu1$	-0.0051	-0.0052	-0.0076	-0.0077	-0.0076	-0.0077
		$\mu2$	0.0003	-0.0001	-0.0025	-0.0028	-0.0025	-0.0027
		τ	-0.0049	-0.0050	-0.0272	-0.0275	-0.0027	-0.0028
$\mu1 = 0.4, \mu2 = 0.4, \tau = 1$	200	$\mu1$	0.0024	0.0024	0.0009	0.0009	0.0009	0.0009
		$\mu2$	-0.0044	-0.0047	-0.0057	-0.0057	-0.0058	-0.0058
		τ	0.0167	0.0166	0.0054	0.0056	0.0171	0.0173
$\mu1 = 0.4, \mu2 = 0.4, \tau = 2$	50	$\mu1$	0.0044	0.0037	0.0076	0.0068	-0.0011	-0.0017
		$\mu2$	0.0058	0.0056	0.0092	0.0083	0.0005	-0.0002
		τ	-0.0644	-0.0624	-0.1090	-0.1054	-0.0549	-0.0502
$\mu1 = 0.4, \mu2 = 0.4, \tau = 2$	100	$\mu1$	-0.0032	-0.0037	-0.0006	-0.0011	-0.0056	-0.0061
		$\mu2$	-0.0023	-0.0026	0.0003	-0.0003	-0.0047	-0.0053
		τ	-0.0198	-0.0184	-0.0455	-0.0428	-0.0153	-0.0120
$\mu1 = 0.4, \mu2 = 0.4, \tau = 2$	200	$\mu1$	0.0012	0.0010	0.0027	0.0025	0.0001	0.0001
		$\mu2$	0.0018	0.0014	0.0031	0.0030	0.0007	0.0006
		τ	-0.0208	-0.0198	-0.0331	-0.0322	-0.0190	-0.0182
$\mu1 = 0.2, \mu2 = 0.4, \tau = 1$	50	$\mu1$	0.0052	0.0051	0.0054	0.0050	0.0050	0.0046
		$\mu2$	0.0011	0.0007	-0.0082	-0.0084	0.0086	0.0090
		τ	0.0266	0.0265	-0.0756	-0.0755	0.0461	0.0460
$\mu1 = 0.2, \mu2 = 0.4, \tau = 1$	100	$\mu1$	-0.0025	-0.0027	-0.0024	-0.0028	-0.0025	-0.0029
		$\mu2$	0.0079	0.0076	0.0027	0.0025	0.0025	0.0025
		τ	0.0366	0.0359	-0.0180	-0.0184	0.0443	0.0441
$\mu1 = 0.2, \mu2 = 0.4, \tau = 1$	200	$\mu1$	-0.0024	-0.0024	-0.0023	-0.0025	-0.0023	-0.0024
		$\mu2$	0.0033	0.0029	0.0005	0.0004	0.0005	0.0004
		τ	-0.0068	-0.0071	-0.0357	-0.0368	-0.0040	-0.0048
$\mu1 = 0.2, \mu2 = 0.4, \tau = 2$	50	$\mu1$	0.0202	0.0195				
		$\mu2$	0.0133	0.0130				
		τ	-0.1356	-0.1356				
$\mu1 = 0.2, \mu2 = 0.4, \tau = 2$	100	$\mu1$	0.0020	0.0017	0.0039	0.0034	0.0014	0.0010
		$\mu2$	-0.0035	-0.0036	-0.0021	-0.0025	-0.0068	-0.0073
		τ	-0.0227	-0.0240	-0.0780	-0.0764	-0.0292	-0.0264
$\mu1 = 0.2, \mu2 = 0.4, \tau = 2$	200	$\mu1$	0.0000	-0.0003	0.0009	0.0006	-0.0003	-0.0007
		$\mu2$	-0.0060	-0.0063	-0.0051	-0.0055	-0.0076	-0.0079
		τ	-0.0059	-0.0050	-0.0330	-0.0317	-0.0092	-0.0078

4.3.3 Marginal Homogeneity

So far, the prior distributions that have been deemed non-informative for 2 by 2 contingency tables were for the saturated model with 3 parameters (μ_1, μ_2, τ). Another model which is often also considered is marginal homogeneity in which $\mu_1 = \mu_2 = \mu$. It is of interest to determine if similar conclusions are obtained with regards to which prior distributions are deemed non-informative for the marginal homogeneity case. The cell probabilities for marginal homogeneity are given in Table 4.10.

Table 4.10: Cell probabilities for a bivariate binary response under marginal homogeneity

	Y_2		
Y_1	1	0	Total
1	$\tau\mu^2$	$\mu - \tau\mu^2$	μ
0	$\mu - \tau\mu^2$	$1 - 2\mu + \tau\mu^2$	$1 - \mu$
Total	μ	$1 - \mu$	1

The likelihood function for the marginal homogeneity model (assuming a multinomial sampling scheme) is therefore given by:

$$\mathrm{L}(\mu, \tau; \mathbf{n}) \propto (1 - 2\mu + \tau\mu^2)^{n_{00}} \, (\mu - \tau\mu^2)^{n_{01}} \, (\mu - \tau\mu^2)^{n_{10}} \, (\tau\mu^2)^{n_{11}}. \tag{4.24}$$

Following the same process as for the saturated model, prior distributions are assessed first, using the R software. The same 10 prior formulations are considered as before but this time, there is only one parameter for the marginal probabilities (μ). In addition, τ is constrained by μ with the range given in (4.25).

$$\max \left\{0, \frac{2}{\mu} - \frac{1}{\mu^2}\right\} \leq \tau \leq \left\{\frac{1}{\mu}\right\}. \tag{4.25}$$

Prior distributions of the form given in (4.26) are therefore considered. Table 4.11 shows the prior formulations.

$$p(\tau, \mu) \propto p(\tau|\mu)p(\mu). \tag{4.26}$$

Table 4.11: Prior formulations for μ, τ (marginal homogeneity)

Prior Formulation	Marginal Probability (μ)	Dependence ratio (τ)
1	$\mu \sim$ uniform(0,1)	$\tau \sim$ uniform(lower,upper)
2	$\mu \sim$ logistic-normal(0,2.5)	$\tau \sim$ uniform(lower,upper)
3	$\mu \sim$ uniform(0,1)	$\tau \sim \Gamma(0.01, 0.01)$ I(lower,upper)
4	$\mu \sim$ logistic-normal(0,2.5)	$\tau \sim \Gamma(0.01, 0.01)$ I(lower,upper)
5	$\mu \sim$ uniform(0,1)	$\tau \sim \Gamma(1, 1)$ I(lower,upper)
6	$\mu \sim$ logistic-normal(0,2.5)	$\tau \sim \Gamma(1, 1)$ I(lower,upper)
7	$\mu \sim$ uniform(0,1)	$\tau \sim$ log-normal(0, uniform(0,100)) I(lower,upper)
8	$\mu \sim$ logistic-normal(0,2.5)	$\tau \sim$ log-normal(0, uniform(0,100)) I(lower,upper)
9	$\mu \sim$ uniform(0,1)	$\tau \sim$ log-normal(1, 1) I(lower,upper)
10	$\mu \sim$ logistic-normal(0,2.5)	$\tau \sim$ log-normal(1, 1) I(lower,upper)

The 10 prior densities for τ (conditional on μ and based on $10,000$ simulations in R) are given in Figure 4.4. The prior densities for μ roughly favour all values between 0 and 1 equally. The prior densities for τ are similar to those from Figure 4.1. Priors $3, 4, 7$ and 8 are excluded from posterior inference since they favour zero in addition to independence and do not allow for values of τ greater than approximately 2. Priors $1, 2, 5, 6, 9$ and 10 are therefore retained for posterior inference.

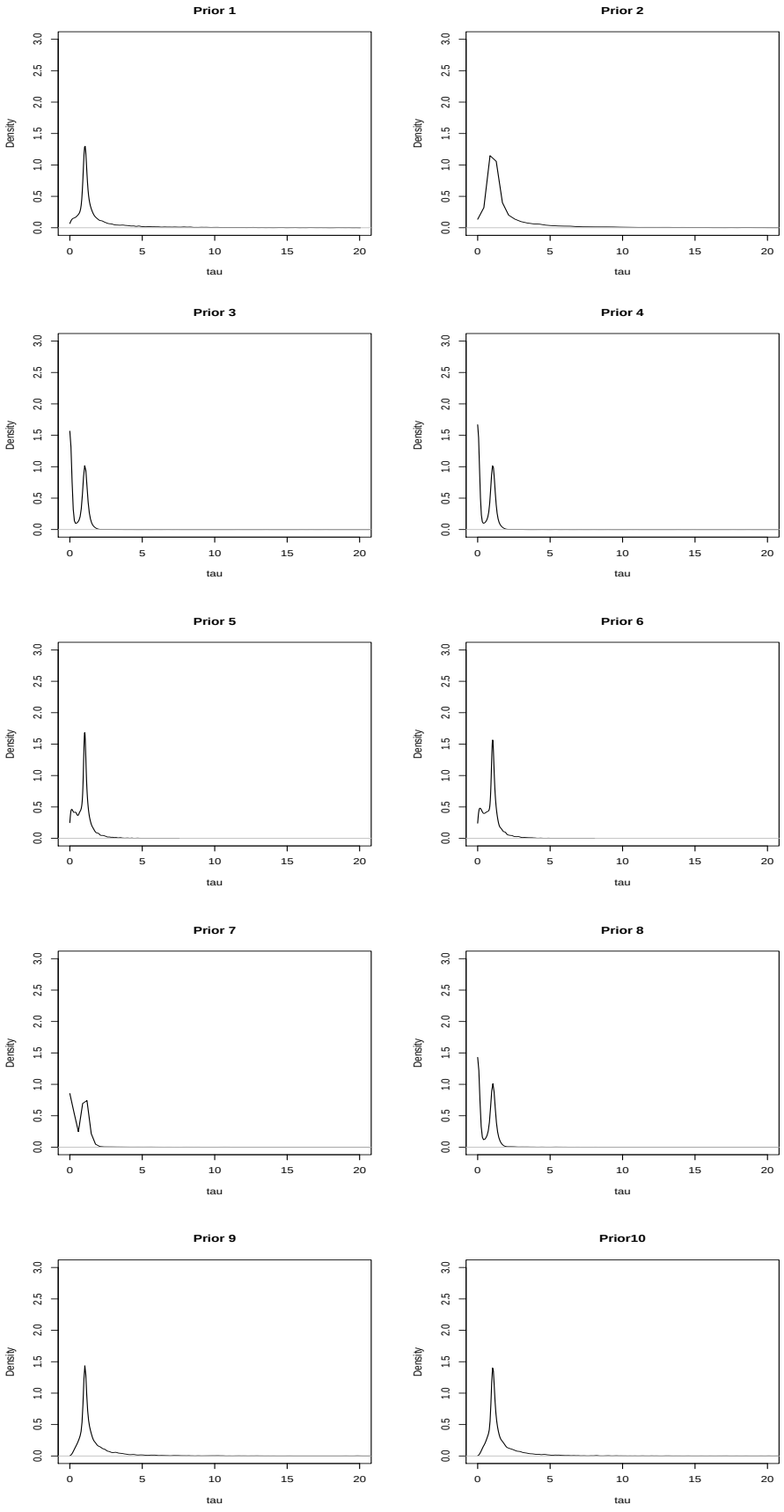

Figure 4.4: Prior distributions for τ under marginal homogeneity

4.3.3.1 Application to datasets

The dataset considered for posterior inference of the marginal homogeneity model was also considered in paper 2 of this thesis and was jointly provided by the MRC Biostatistics Unit in Cambridge and the Toronto Psoriatic Arthritis Clinic. It relates to 386 patients who enrolled in the Toronto Psoriatic Arthritis Clinic between 1978 and 2000. Each patient was followed longitudinally with regular assessments made on each of their 28 hand joints (excluding the wrist). Clinical damage was the response of interest considered in paper 2 of this thesis and in order to make the patients comparable, no patients in the dataset had any clinically damaged joints upon entry to the clinic. A joint is said to be clinically damaged if there is a limitation in its movement of more than 20% of the range not related to joint swelling (Gladman et al. 1990; Cresswell and Farewell, 2010). For a more precise definition, see paper 2 of this thesis.

Only the last clinic visit for each joint of each patient was considered in paper 2 since clinical damage is an irreversible process. In paper 2 of this thesis, the probability of clinical damage was found to be similar in both the right and left hands. Therefore, a 2 by 2 contingency table relating to clinical damage in the right and left hands seemed a sensible choice for posterior inference of the marginal homogeneity model. Table 4.12 considers the association between clinical damage in the right and left hands by cross classifying the 386 patients. A patient is classified as having clinical damage in a given hand if at least one joint of the 14 in that hand are clinically damaged. It should be noted that this is only a subset of the dataset with paper 2 of this thesis conducting a more thorough analysis.

Table 4.12: Clinical damage by hand

	Left Hand	
Right Hand	At least one joint damaged	None damaged
At least one joint damaged	75	38
None damaged	24	249

In order for the prior formulations taken forward for posterior inference to be considered non-informative for the marginal homogeneity model described in Section 4.3.3, posterior inference should be dominated by the data. A frequentist analysis using maximum likelihood estimation (using the R package drm) yielded the estimates in Table 4.13.

Table 4.13: Maximum likelihood estimates from a frequentist analysis of the marginal homogeneity model

Parameter	Value	Standard error
μ	0.2746	0.0203
τ	2.5765	0.1863

The estimates from a frequentist analysis of the saturated model were $0.2927, 0.2565$ and 2.5878 for μ_1, μ_2 and τ respectively. The marginal homogeneity model therefore appears to be a sensible choice with its model fit being assessed in a Bayesian context using the DIC in due course. Similar to the case of the saturated model, direct simulation was not possible for posterior inference when each prior formulation was combined with the likelihood from (4.24). Therefore, for each of the 6 prior formulations taken forward for posterior inference in this section, $100,000$ simulations were run in BUGS, for each of two chains of initial values (excluding burn-in). Table 4.14 gives the posterior summaries for each of the 6 prior formulations and Figure 4.5 gives the posterior densities.

Table 4.14: Posterior summaries for the 6 prior formulations for the marginal homogeneity model

		Posterior mean	2.5	50th	97.5	Sampling method
Prior 1	μ	0.2749	0.2364	0.2745	0.3152	Slice sampling
	τ	2.5760	2.2380	2.5660	2.9690	Slice sampling
Prior 2	μ	0.2746	0.2363	0.2743	0.3146	Slice sampling
	τ	2.5770	2.2390	2.5670	2.9700	Slice sampling
Prior 5	μ	0.2763	0.2377	0.2759	0.3167	Slice sampling
	τ	2.5510	2.2180	2.5420	2.9370	Slice sampling
Prior 6	μ	0.2760	0.2374	0.2757	0.3161	Slice sampling
	τ	2.5530	2.2200	2.5430	2.9400	Slice sampling
Prior 9	μ	0.2745	0.2358	0.2741	0.3149	Slice sampling
	τ	2.5730	2.2350	2.5630	2.9660	Slice sampling
Prior 10	μ	0.2743	0.2356	0.2740	0.3144	Slice sampling
	τ	2.5740	2.2360	2.5640	2.9700	Slice sampling

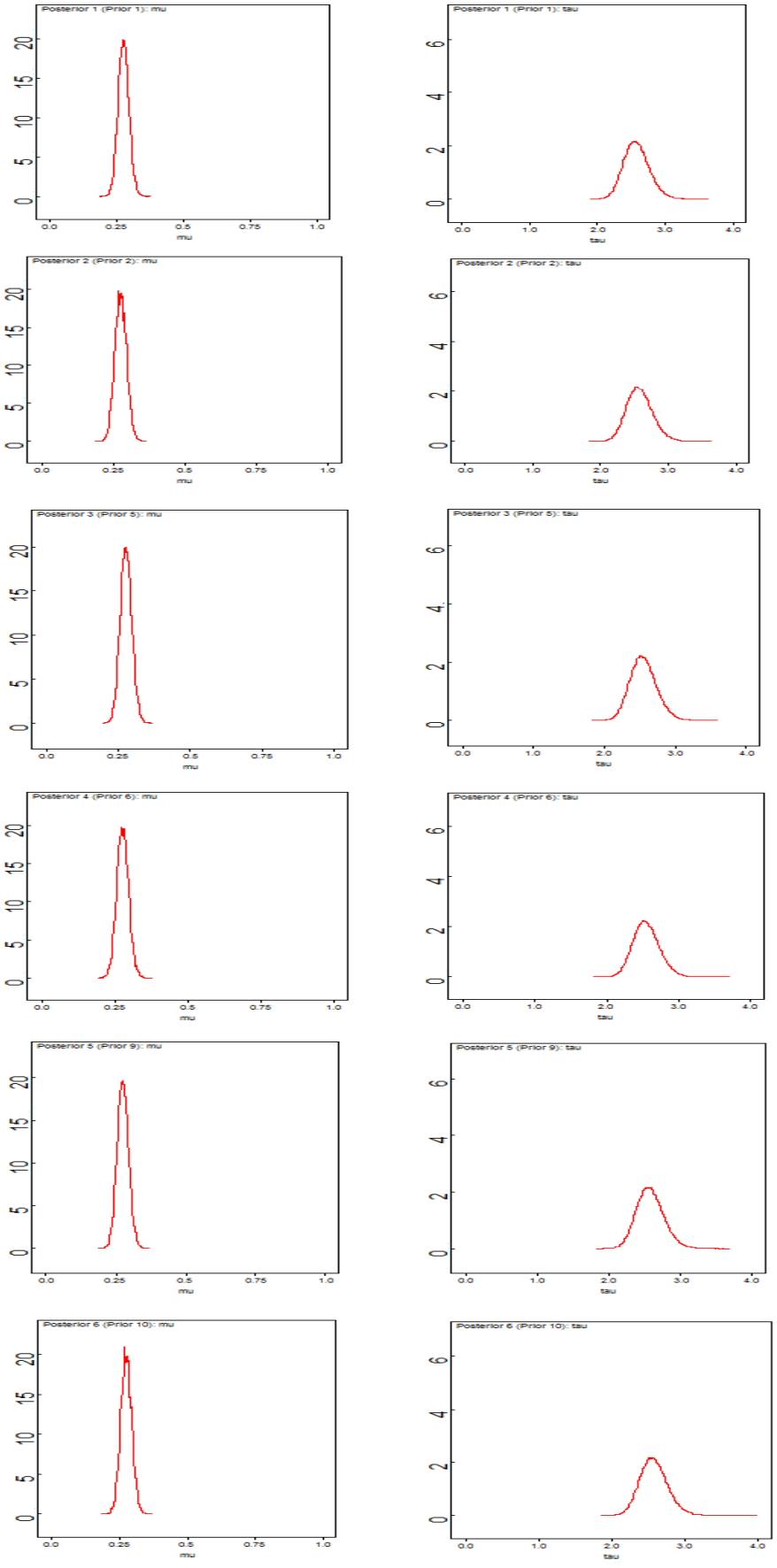

Figure 4.5: Posterior distributions for μ, τ under marginal homogeneity

Prior formulations $1, 9, 10$ and in particular 2 gave posterior mean estimates that are close to the maximum likelihood estimates from Table 4.13 and therefore appear non-informative for this dataset at least. However, prior formulations 5 and 6 yielded posterior estimates that differ more substantially from those in Table 4.13 and therefore do not seem to be the most appropriate non-informative priors for the marginal homogeneity model, most likely due to assigning too much prior probability to zero.

In order to assess the fit of the marginal homogeneity model, the DIC was used to compare the fit to the saturated model, for reasons discussed in Section 4.2.2. Prior formulation 2 was chosen for both models due to the conclusions from Figure 4.4 and Table 4.14. The DIC statistic was 23.36 and 22.13 for the marginal homogeneity and saturated models respectively. Spiegelhalter et al. (2002) suggested using the DIC like AIC for model comparison with differences of between 0 and 2 described as essentially none. The marginal homogeneity model therefore gives a satisfactory fit. From this model, the posterior mean parameters indicate that the probability of a patient having at least one joint damaged in a particular hand is 0.2746 and the probability of a patient having at least one joint damaged in both hands is 2.577 times the corresponding probability under independence.

As for the saturated case, a further analysis was conducted on the hypothetical situation of the same dataset but with (approximately) half the sample size in order to assess if conclusions changed with a reduced sample. Thus, the sample counts of $75, 38, 24$ and 249 were replaced with counts of $38, 19, 12$ and 124. The frequentist maximum likelihood estimates for this dataset were 0.2772 and 2.5623 for μ and τ respectively. Table 4.15 gives the posterior summaries for each of the 6 prior formulations.

Table 4.15: Posterior summaries for the 6 prior formulations of the reduced sample for the marginal homogeneity model

		Posterior mean	2.5	50th	97.5	Sampling method
Prior 1	μ	0.2775	0.2235	0.2769	0.3352	Slice sampling
	τ	2.5620	2.1000	2.5420	3.1320	Slice sampling
Prior 2	μ	0.2772	0.2231	0.2765	0.3353	Slice sampling
	τ	2.5640	2.1030	2.5460	3.1340	Slice sampling
Prior 5	μ	0.2804	0.2267	0.2797	0.3378	Slice sampling
	τ	2.5120	2.0680	2.4950	3.0570	Slice sampling
Prior 6	μ	0.2800	0.2263	0.2791	0.3375	Slice sampling
	τ	2.5160	2.0690	2.4980	3.0620	Slice sampling
Prior 9	μ	0.2765	0.2228	0.2759	0.3342	Slice sampling
	τ	2.5580	2.0970	2.5390	3.1290	Slice sampling
Prior 10	μ	0.2763	0.2227	0.2754	0.3341	Slice sampling
	τ	2.5590	2.0990	2.5410	3.1270	Slice sampling

Prior formulations 1 and 2 gave the closest (posterior mean) estimates to an analysis of the data alone using maximum likelihood and therefore appear to be the most appropriate non-informative priors to consider for the marginal homogeneity model. Prior

formulations 5 and 6 again yield posterior mean estimates that lie furthest from the maximum likelihood estimates.

4.3.4 Informative Priors

Section 4.2.1 discussed the use of the multivariate logistic-normal distribution for the marginal probabilities if substantive prior information is available, as opposed to assuming prior independence. This distribution assumes a multivariate normal distribution for the multinomial logits of the marginal probabilities and consequently gives a multivariate logistic-normal distribution for the multinomial parameters (marginal probabilities). For the bivariate case (with marginal probabilities μ_1 and μ_2) that has been the focus of this section, this distribution is referred to as the bivariate logistic-normal distribution. Consider two random variables $(X, Y) = (\text{logit}(\mu_1), \text{logit}(\mu_2))$ that have a bivariate normal distribution, which is specified as:

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\lambda_x)^2}{\sigma_x^2} + \frac{(y-\lambda_y)^2}{\sigma_y^2} - \frac{2\rho(x-\lambda_x)(y-\lambda_y)}{\sigma_x\sigma_y}\right]\right\}, \quad (4.27)$$

where λ_x and λ_y are the means of X and Y respectively, σ_x and σ_y are the standard deviations of X and Y respectively, $\sigma_x > 0$ and $\sigma_y > 0$, ρ is the correlation between X and $Y, -1 \leq \rho \leq 1$.

When ρ is 0, there is prior independence between X and Y, as was considered for the logistic-normal priors in this section. In terms of what values to assign to the parameters in (4.27), this will likely come from expert judgment in the relevant research area or previous studies. However, a notable disadvantage with this prior is that it likely to be difficult to specify prior information for the parameters on the logit scale, as noted in Albert and Gupta (1983a).

An alternative approach is to specify prior distributions directly for the marginal probabilities and dependence ratio, as was considered previously for the non-informative priors. For example, the researcher may have a prior belief that the marginal probabilities should not range between 0 and 1. Even if no prior information is available for τ, its prior distribution may well be informative if an informative prior is used for the marginal probabilities since τ is constrained by the marginal probabilities.

There is an advantage to assuming the marginal probabilities lie between 0 and 0.5 as a priori (if such prior information is available in advance of the data). Consider the case where both the marginal probabilities are less than 0.5. In this situation, they are reasonably variation independent of τ, as can be determined from (4.10). Ekholm (2003) discusses this in detail for the marginal homogeneity case. Table 4.16 presents the range of τ for differing values of μ_1 and μ_2.

Table 4.16: Range of τ for differing values of μ_1 and μ_2

$\mu 1$	$\mu 2$	Range of τ
0.1	0.1	$0 \leq \tau \leq 10$
0.25	0.25	$0 \leq \tau \leq 4$
0.1	0.4	$0 \leq \tau \leq 2.5$
0.4	0.4	$0 \leq \tau \leq 2.5$
0.6	0.6	$0.56 \leq \tau \leq 1.67$
0.75	0.75	$0.89 \leq \tau \leq 1.33$
0.6	0.9	$0.93 \leq \tau \leq 1.11$
0.9	0.9	$0.99 \leq \tau \leq 1.11$
1	1	1

Table 4.16 shows that as the marginal probabilities increase, τ has a range that becomes narrower around the independence value of 1. This therefore shows that the dependence ratio is most appropriate for cases where the marginal probabilities are less than approximately 0.5, such as the arthritis dataset considered in Table 4.12. In other words, modelling the rarer event of a bivariate binary response (either $Y_k = 1$ or $Y_k = 0, k = 1, 2$) is more appropriate. For example, for the arthritis dataset in Table 4.12, modelling the probability of having at least one joint damaged in a given hand ($Y_k = 1$) was more appropriate than modelling the probability of having no joints damaged in a given hand. In addition to being the rarer of the two events, the probability of having at least one joint damaged was of more interest than modelling the probability of having no joints damaged.

In order to illustrate the way in which an informative prior could be applied in this context, consider the hypothetical situation that the researcher had prior knowledge that the marginal probabilities ranged between 0.2 and 0.4 as opposed to 0 and 1 in the arthritis dataset from Table 4.12. It should be noted that this analysis is not advocating the use of the empirical Bayes approach whereby prior distributions are influenced by the data, it is purely for demonstrating how an informative prior could be applied. Prior formulation 1 is the focus of this analysis, where τ is assigned a uniform distribution between its lower and upper bounds. Figure 4.6 shows the prior density for τ.

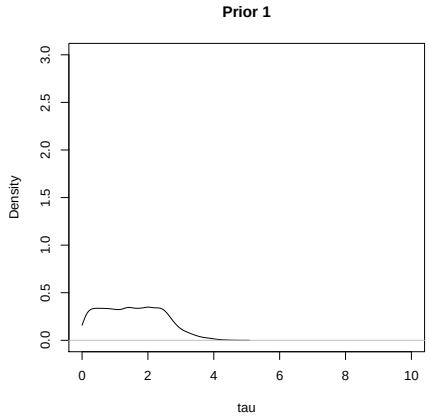

Figure 4.6: Prior density for tau with prior information for the marginal probabilities

Despite the added prior information for the marginal probabilities, values of τ between 0 and approximately 2 are favoured equally with larger values of τ being favoured less. The same conclusions were obtained when considering the same prior under the marginal homogeneity model. The marginal homogeneity model again gave a superior fit over the saturated model in terms of DIC (23.38 compared to 22.12). In this case, the posterior estimates do not differ greatly from the non-informative case considered previously however they are included in Table 4.17 for completeness. Figure 4.7 shows the posterior densities.

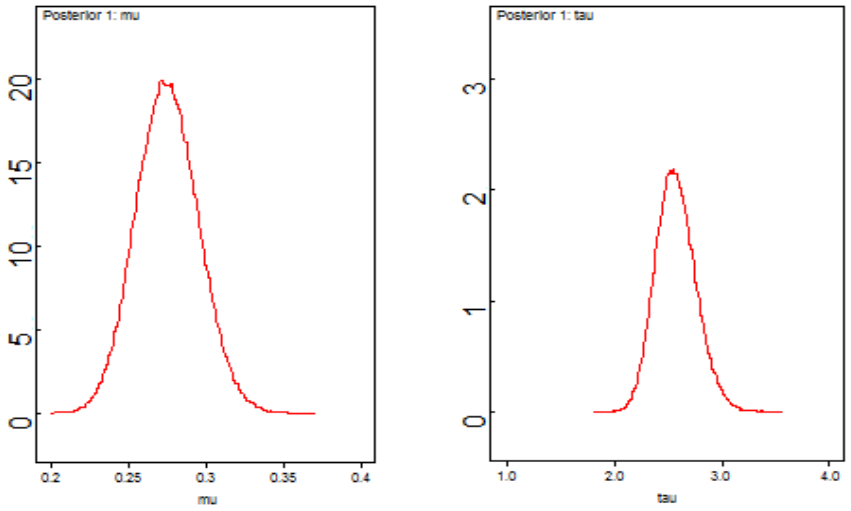

Figure 4.7: Posterior densities with prior information for the marginal probabilities

Table 4.17: Posterior summaries for the marginal homogeneity model with prior information

		Posterior mean	2.5	50th	97.5	Sampling method
Prior 1	μ	0.2748	0.2363	0.2745	0.3153	Slice sampling
	τ	2.5760	2.2370	2.5660	2.9690	Slice sampling

4.4 Extension to $2 \times 2 \times 2$ contingency tables

The following section extends the models described for 2×2 contingency tables to the inclusion of covariate information. This is undertaken for the simplest case of an additional categorical variable with two categories, which therefore yields a $2 \times 2 \times 2$ contingency table. In order to demonstrate this, the arthritis dataset from Section 4.3.3.1 is extended to account for gender. This dataset is given in Table 4.18.

Table 4.18: Clinical damage by hand and gender

Gender:	Male		Female		
Right Hand: Left Hand	At least one	None	At least one	None	Total
At least one joint damaged	38	15	37	23	113
None damaged	16	145	8	104	273
Total	54	160	45	127	386

Assuming a multinomial sampling scheme, f^q multinomial probabilities have to be estimated for each covariate combination, where f is the number of categories in the response and q is the number of subunits in a cluster. Thus, for the example in Table 4.18 $2 \times 2^2 = 8$ multinomial probabilities have to be estimated. Since the probabilities must sum to one for each covariate combination, $2 \times (f^q - 1)$ parameters are required to completely specify the joint distribution. In other words, 6 parameters are required to completely specify the joint distribution for the example in Table 4.18. The cell probabilities for each covariate combination are shown in Table 4.19, in terms of the dependence ratios and marginal probabilities.

Table 4.19: Cell probabilities for the arthritis dataset by gender

Males	Left Hand		
Right Hand	At least one	None	Total
At least one	$\tau^m \mu_1^m \mu_2^m$	$\mu_1^m - \tau^m \mu_1^m \mu_2^m$	μ_1^m
None	$\mu_2^m - \tau^m \mu_1^m \mu_2^m$	$1 - \mu_1^m - \mu_2^m + \tau^m \mu_1^m \mu_2^m$	$1 - \mu_1^m$
Total	μ_2^m	$1 - \mu_2^m$	1

Females	Left Hand		
Right Hand	At least one	None	Total
At least one	$\tau^f \mu_1^f \mu_2^f$	$\mu_1^f - \tau^f \mu_1^f \mu_2^f$	μ_1^f
None	$\mu_2^f - \tau^f \mu_1^f \mu_2^f$	$1 - \mu_1^f - \mu_2^f + \tau^f \mu_1^f \mu_2^f$	$1 - \mu_1^f$
Total	μ_2^f	$1 - \mu_2^f$	1

In other words, 2 dependence ratios (τ^m, τ^f) and 4 marginal probabilities $(\mu_1^m, \mu_1^f, \mu_2^m, \mu_2^f)$ are required to completely specify the joint distribution, where m and f refer to males and females respectively. There are consequently a number of possible models that can be considered, ranging from the saturated model (6 parameters) in Table 4.19 to a model that assumes an independent association structure ($\tau_f = \tau_m = 1$) and all marginal probabilities are equal (1 parameter=μ). In terms of the prior distributions to assign to the parameters, attention is focused on prior formulations 1 and 2 from Section 4.3 (for both males and females) since these proved to be the most non-informative of those considered. In other words, the marginal probabilities are assigned uniform(0,1) and logistic-normal(0,2.5) distributions respectively whereas the dependence ratios are given a uniform distribution between the relevant range for both formulations, using (4.10). In order for the priors to be considered non-informative, it is expected that they should yield similar estimates to a model from the data alone. The model in Table 4.19 is chosen as a starting point. Table 4.20 shows the estimates from a frequentist analysis of the saturated model.

Table 4.20: Saturated estimates from a frequentist analysis of the arthritis data by gender

Parameter	Value	Standard error
μ_{1m}	0.2477	0.0295
μ_{2m}	0.2523	0.0297
μ_{1f}	0.3488	0.0363
μ_{2f}	0.2616	0.0335
τ_m	2.8414	0.2954
τ_f	2.3570	0.2325

For each of the 2 prior formulations considered, $100,000$ simulations were run in BUGS (after burn-in), for each of two chains of initial values. Table 4.21 shows the posterior estimates for this model and Figure 4.8 shows the posterior densities for the 6 parameters of prior 1. The corresponding densities of prior 2 were similar so are therefore not included to avoid too much repetition.

Table 4.21: Posterior summaries for the 2 prior formulations for the arthritis data by gender

		Posterior mean	2.5	50th	97.5	Sampling method
Prior 1	μ_{1m}	0.2495	0.1945	0.2487	0.3092	Slice sampling
	μ_{2m}	0.2542	0.1985	0.2534	0.3138	Slice sampling
	μ_{1f}	0.3496	0.2810	0.3492	0.4215	Slice sampling
	μ_{2f}	0.2627	0.2005	0.2619	0.3300	Slice sampling
	τ_m	2.8050	2.2920	2.7840	3.4370	Slice sampling
	τ_f	2.3300	1.9270	2.3140	2.8270	Slice sampling
Prior 2	μ_{1m}	0.2494	0.1938	0.2485	0.3097	Slice sampling
	μ_{2m}	0.2541	0.1981	0.2533	0.3149	Slice sampling
	μ_{1f}	0.3492	0.2791	0.3489	0.4214	Slice sampling
	μ_{2f}	0.2623	0.1995	0.2615	0.3301	Slice sampling
	τ_m	2.8060	2.2920	2.7840	3.4470	Slice sampling
	τ_f	2.3330	1.9250	2.3150	2.8440	Slice sampling

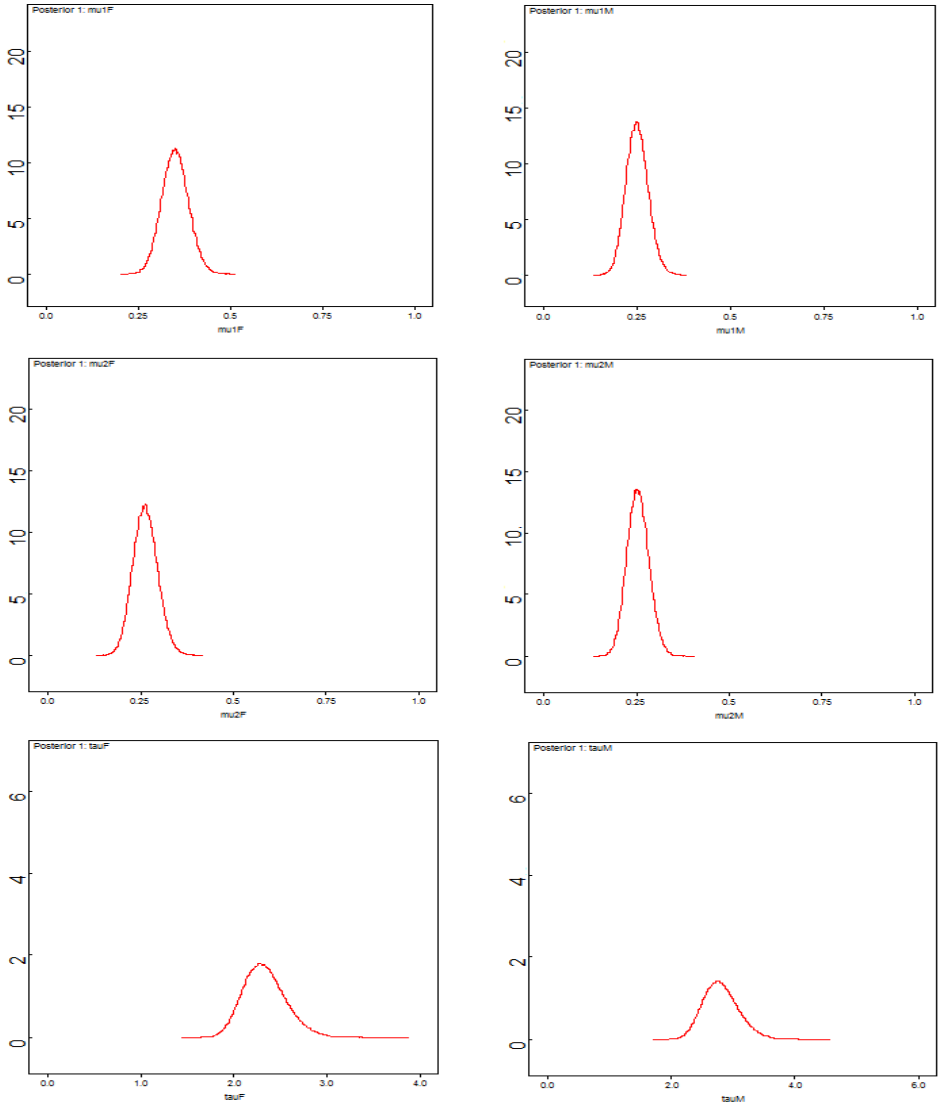

Figure 4.8: Posterior distributions for $\mu_1^f, \mu_1^m, \mu_2^f, \mu_2^m, \tau^f, \tau^m$

Both prior formulations gave similar results to the saturated model from Table 4.20. As discussed previously, there are a large number of possible models that can be considered. The prior distributions will be kept consistent between the models in the sense that for each of the uniform and logistic-normal priors for the marginal probabilities, the dependence ratio(s) are assigned a uniform distribution between the relevant lower and upper bounds from (4.10). The DIC is used to compare potential models. For the full model with 6 parameters, the DIC's were 39.98 and 40.055 for prior formulations 1 and 2 respectively. In order to assess the gender effect, consider a model with $\mu_1^f = \mu_2^f = \mu_1, \mu_1^m = \mu_2^m = \mu_2$ and separate dependence ratios for males and females. This model gave a DIC of 42.6 and 42.61 for priors 1 and 2 respectively. Following Spiegelhalter et al. (2002) who deem differences in DIC of less than 4 to be unimportant, it can be concluded that having the marginal probabilities varying by gender is not necessary. By further constraining $\mu_1 = \mu_2 = \mu$, the DIC is given by 43.46 and 43.51 for priors 1 and 2 respectively. Constraining the dependence ratios of males and females to also be equal yielded a DIC of 41.90 and 41.92 for priors 1 and 2 respectively. Table 4.22 shows the DIC's for all of these models as well as other potential models.

Table 4.22: Model comparison of DIC

Model Parameters	Prior Formulation	DIC
$\tau_f, \tau_m, \mu_{1m}, \mu_{1f}, \mu_{2m}, \mu_{2f}$	1	39.980
$\tau_f, \tau_m, \mu_{1m}, \mu_{1f}, \mu_{2m}, \mu_{2f}$	2	40.055
$\tau_f, \tau_m, \mu_1, \mu_2$	1	42.600
$\tau_f, \tau_m, \mu_1, \mu_2$	2	42.610
τ_f, τ_m, μ	1	43.460
τ_f, τ_m, μ	2	43.510
τ, μ_1, μ_2	1	40.720
τ, μ_1, μ_2	2	40.758
τ, μ	1	41.900
τ, μ	2	41.920
μ	1	170.700
μ	2	170.700

All of the models in Table 4.22 as well as other models (not shown) gave similar values for the DIC except for the model with a single parameter (μ) which yielded a significantly worse fit. In addition, prior 1 consistently gave smaller values than prior 2 (for each model). Given this, the model with prior 1 and two parameters (τ, μ) is chosen as the final model due to the interest of model parsimony. In other words, females and males show no significantly different patterns with regards to the association structure or the marginal probabilities. In addition, marginal homogeneity is also present. The final chosen model is therefore the same as prior formulation 1 in Table 4.14. The probability of an individual (male or female) having at least one joint damaged in both their right and left hands is 2.576 times the probability under independence. In addition, the probability of a patient having at least one joint damaged in a particular hand is 0.2749 (for both females and males).

The Bayesian analysis of this dataset coincides with the conclusions from the frequentist analysis of the dataset in paper 2 of this thesis in the sense that males and females followed similar patterns in terms of the association structure and the marginal probabilities. Of course, the analysis in paper 2 was performed in much greater detail since additional covariates were included and the response of interest was the probability of clinical damage at each individual joint as opposed to the probability of having at least one joint damaged.

4.5 Conclusions

The key conclusion from Section 4.3 was that prior formulations 1 and 2 (and to a lesser extent 9 and 10) are the most appropriate non-informative priors for the marginal probabilities and τ (of those considered), for both the saturated and marginal homogeneity models in 2 by 2 contingency tables. In addition to consistently giving the closest estimates to an analysis of the data alone, they also do not require resampling for the conditional prior of τ since τ takes values uniformly between its upper and lower bounds from (4.10). Although considered based on prior density plots, it is clear that prior formulations 5 and 6 were too informative and should not be considered in future analyses.

It is well known that prior distributions have more of an impact on posterior inference in smaller samples. These analyses with reduced sample sizes still favoured prior formulations 1 and 2 although clearly the smaller samples meant estimates deviated more from the corresponding maximum likelihood estimates. In addition, priors 1 and 2 generally yielded unbiased estimates in the simulation study that was conducted (and less biased than the other priors).

Informative priors may also be of interest for future analyses. The researcher may have a prior belief that the marginal probabilities (μ_1 and μ_2) are correlated rather than independent, thus advocating the use of the multivariate logistic-normal distribution. However, the researcher may find it difficult to assign values for the parameters on the logit scale. In another situation, the researcher may have prior information that the marginal probabilities should not range between 0 and 1, as discussed in Section 4.3.4.

4.5.1 Further extensions

Section 4.4 showed how the approach from Section 4.3 can be extended to include categorical variables with only two levels. There are a number of other possible extensions for the approach. The extension to larger contingency tables is the focus of this discussion (with no covariates). Consider a $f \times f$ contingency table, where f is the number of categories in the response. There are $f^q - q(f-1) - 1$ dependence ratios and $q(f-1)$

marginal probabilities to completely specify the joint distribution, where q is the number of subunits in a cluster. Consider a 3×3 contingency table with a bivariate response $\boldsymbol{Y} = (Y_1, Y_2)$. There are thus 4 marginal probabilities and 4 dependence ratios to completely specify the joint distribution, as shown in Table 4.23.

Table 4.23: Cell probabilities for a 3 by 3 contingency table

	Y_2			
Y_1	3	2	1	Marginal
3	$\tau_{12}^{(3,3)}\mu_1^{(3)}\mu_2^{(3)}$	$\tau_{12}^{(3,2)}\mu_1^{(3)}\mu_2^{(2)}$	$\mu_1^{(3)} - \tau_{12}^{(3,3)}\mu_1^{(3)}\mu_2^{(3)} - \tau_{12}^{(3,2)}\mu_1^{(3)}\mu_2^{(2)}$	$\mu_1^{(3)}$
2	$\tau_{12}^{(2,3)}\mu_1^{(2)}\mu_2^{(3)}$	$\tau_{12}^{(2,2)}\mu_1^{(2)}\mu_2^{(2)}$	$\mu_1^{(2)} - \tau_{12}^{(2,3)}\mu_1^{(2)}\mu_2^{(3)} - \tau_{12}^{(2,2)}\mu_1^{(2)}\mu_2^{(2)}$	$\mu_1^{(2)}$
1	$\mu_2^{(3)} - \tau_{12}^{(3,3)}\mu_1^{(3)}\mu_2^{(3)} - \tau_{12}^{(2,3)}\mu_1^{(2)}\mu_2^{(3)}$	$\mu_2^{(2)} - \tau_{12}^{(3,2)}\mu_1^{(3)}\mu_2^{(2)} - \tau_{12}^{(2,2)}\mu_1^{(2)}\mu_2^{(2)}$	$1 - \mu_2^{(3)} - \mu_2^{(2)} - \mu_1^{(3)} - \tau_{12}^{(3,3)}\mu_1^{(3)}\mu_2^{(3)} - \tau_{12}^{(3,2)}\mu_1^{(3)}\mu_2^{(2)} - \mu_1^{(2)} - \tau_{12}^{(2,3)}\mu_1^{(2)}\mu_2^{(3)} - \tau_{12}^{(2,2)}\mu_1^{(2)}\mu_2^{(2)}$	$1 - \mu_1^{(3)} - \mu_1^{(2)}$
Marginal	$\mu_2^{(3)}$	$\mu_2^{(2)}$	$1 - \mu_2^{(3)} - \mu_2^{(2)}$	1

The marginal probabilities are given by:

$$\mu_1^{(a_1)} = pr(Y_1 = a_1), \mu_2^{(a_2)} = pr(Y_2 = a_2), \tag{4.28}$$

for $a_1, a_2 = 2, 3$.

The dependence ratios are given by:

$$\tau_{12}^{(a_1,a_2)} = \frac{\mu_{12}^{(a_1,a_2)}}{\mu_1^{(a_1)}\mu_2^{(a_2)}} = \frac{pr(Y_1 = a_1, Y_2 = a_2)}{pr(Y_1 = a_1)\, pr(Y_2 = a_2)}, \tag{4.29}$$

for $a_1, a_2 = 2, 3$.

In contrast to the 2×2 contingency table, it is no longer appropriate to consider uniform(0,1) prior distributions for all of the marginal probabilities. For example, assigning uniform(0,1) priors to both $\mu_1^{(3)}$ and $\mu_1^{(2)}$ may yield prior estimates that exceed the fact that $\mu_1^{(3)} + \mu_1^{(2)} + \mu_1^{(1)} = 1$. The same is true for $\mu_2^{(3)}$ and $\mu_2^{(2)}$. One solution is to treat $\mu_1^{(3)}, \mu_1^{(2)}, \mu_1^{(1)}$ as Dirichlet and $\mu_2^{(3)}, \mu_2^{(2)}, \mu_2^{(1)}$ as Dirichlet. This allows for prior information to still be specified for the marginal probabilities. Similar to the bivariate binary case, the prior distributions for the dependence ratios can be constrained between their relevant range and depending on the application, some of the dependence ratios may be constrained to be equal in the likelihood.

4.5.1.1 Negative profile probabilities

Another area of further research is to determine whether the Bayesian approach can be used to fit models that encounter negative profile probabilities in a frequentist setting. Negative profile probabilities are discussed in detail in the introduction to this thesis. They can arise due to the fact that the dependence ratio approach uses the probabilities as opposed to the logits of the probabilities. Solving this issue will not be straightforward. For example, if MCMC is being used to sample iteratively from the posterior distribution, the positivity of the profiles will need to be checked after each iteration. The simplest solution is to simply ignore the draws from the posterior distribution that produce negative profile probabilities. However, if this leads to a large number of draws being rejected, alternative methods will need to be sought.

Appendix

R code for the prior formulations in Section 4.3.1

```
# Prior formulation 1

Prior1<-function(n){
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
mu1[i]<-runif(1,0,1)
mu2[i]<-runif(1,0,1)
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min((1/mu1[i]),(1/mu2[i]))
tau[i]<-runif(1,lower[i],upper[i])
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p1<-Prior1(10000)
mu1<-p1$mu1
mu2<-p1$mu2
tau1<-p1$tau
density<-plot(density(tau1,from=0),main="Prior 1",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 1",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 1",xlab="mu2"
,xlim=c(0,1))

# Prior formulation 2

Prior2<-function(n){
logitmu1<-NULL
logitmu2<-NULL
mu1<-NULL
```

```
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
logitmu1[i]<-rnorm(1, mean = 0, sd = 1.58)
logitmu2[i]<-rnorm(1, mean = 0, sd = 1.58)
mu1[i]<-exp(logitmu1[i])/(1+exp(logitmu1[i]))
mu2[i]<-exp(logitmu2[i])/(1+exp(logitmu2[i]))
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
tau[i]<-runif(1,lower[i],upper[i])
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p2<-Prior2(10000)
mu1<-p2$mu1
mu2<-p2$mu2
tau2<-p2$tau
density<-plot(density(tau2,from=0),main="Prior 2",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 2",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 2",xlab="mu2"
,xlim=c(0,1))


# Prior formulation 3

Prior3<-function(n,shape,rate){
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
mu1[i]<-runif(1,0,1)
mu2[i]<-runif(1,0,1)
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
```

```
upper[i]<-min(1/mu1[i],1/mu2[i])
repeat {
tau[i]<-rgamma(1,shape,rate)
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p3<-Prior3(10000,0.01,0.01)
mu1<-p3$mu1
mu2<-p3$mu2
tau3<-p3$tau
density<-plot(density(tau3,from=0),main="Prior 3",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 3",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 3",xlab="mu2"
,xlim=c(0,1))

# Prior formulation 4

Prior4<-function(n,shape,rate){
logitmu1<-NULL
logitmu2<-NULL
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
logitmu1[i]<-rnorm(1, mean = 0, sd = 1.58)
logitmu2[i]<-rnorm(1, mean = 0, sd = 1.58)
mu1[i]<-exp(logitmu1[i])/(1+exp(logitmu1[i]))
mu2[i]<-exp(logitmu2[i])/(1+exp(logitmu2[i]))
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
repeat {
tau[i]<-rgamma(1,shape,rate)
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
```

```
}
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p4<-Prior4(10000,0.01,0.01)
mu1<-p4$mu1
mu2<-p4$mu2
tau4<-p4$tau
density<-plot(density(tau4,from=0),main="Prior 4",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 4",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 4",xlab="mu2"
,xlim=c(0,1))

# Prior formulation 5

Prior5<-function(n,shape,rate){
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
mu1[i]<-runif(1,0,1)
mu2[i]<-runif(1,0,1)
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
repeat {
tau[i]<-rgamma(1,shape,rate)
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p5<-Prior5(10000,1,1)
mu1<-p5$mu1
mu2<-p5$mu2
```

```
tau5<-p5$tau
density<-plot(density(tau5,from=0),main="Prior 5",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 5",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 5",xlab="mu2"
,xlim=c(0,1))

# Prior formulation 6

Prior6<-function(n,shape,rate){
logitmu1<-NULL
logitmu2<-NULL
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
logitmu1[i]<-rnorm(1, mean = 0, sd = 1.58)
logitmu2[i]<-rnorm(1, mean = 0, sd = 1.58)
mu1[i]<-exp(logitmu1[i])/(1+exp(logitmu1[i]))
mu2[i]<-exp(logitmu2[i])/(1+exp(logitmu2[i]))
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
repeat {
tau[i]<-rgamma(1,shape,rate)
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p6<-Prior6(10000,1,1)
mu1<-p6$mu1
mu2<-p6$mu2
tau6<-p6$tau
density<-plot(density(tau6,from=0),main="Prior 6",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 6",xlab="mu1"
```

```
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 6",xlab="mu2"
,xlim=c(0,1))

Prior7<-function(n){
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
b<-NULL
for (i in 1:n){
mu1[i]<-runif(1,0,1)
mu2[i]<-runif(1,0,1)
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
b[i]<-runif(1,0,100)
repeat {
tau[i]<-rlnorm(1,meanlog=0,sdlog=b[i])
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p7<-Prior7(10000)
mu1<-p7$mu1
mu2<-p7$mu2
tau7<-p7$tau
density<-plot(density(tau7,from=0),main="Prior 7"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 7",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 7",xlab="mu2"
,xlim=c(0,1))

Prior8<-function(n){
logitmu1<-NULL
logitmu2<-NULL
mu1<-NULL
```

```
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
b<-NULL
for (i in 1:n){
logitmu1[i]<-rnorm(1, mean = 0, sd = 1.58)
logitmu2[i]<-rnorm(1, mean = 0, sd = 1.58)
mu1[i]<-exp(logitmu1[i])/(1+exp(logitmu1[i]))
mu2[i]<-exp(logitmu2[i])/(1+exp(logitmu2[i]))
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
b[i]<-runif(1,0,100)
repeat {
tau[i]<-rlnorm(1,meanlog=0,sdlog=b[i])
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p8<-Prior8(10000)
mu1<-p8$mu1
mu2<-p8$mu2
tau8<-p8$tau
density<-plot(density(tau8,from=0),main="Prior 8",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 8",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 8",xlab="mu2"
,xlim=c(0,1))

Prior9<-function(n,meanlog,sdlog){
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
mu1[i]<-runif(1,0,1)
```

```
mu2[i]<-runif(1,0,1)
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
repeat {
tau[i]<-rlnorm(1,meanlog,sdlog)
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
}
return(data.frame(mu1,mu2,tau,lower,upper))

p9<-Prior9(10000,1,1)
mu1<-p9$mu1
mu2<-p9$mu2
tau9<-p9$tau
density<-plot(density(tau9,from=0),main="Prior 9",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 9",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 9",xlab="mu2"
,xlim=c(0,1))

Prior10<-function(n,meanlog,sdlog){
logitmu1<-NULL
logitmu2<-NULL
mu1<-NULL
mu2<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
logitmu1[i]<-rnorm(1, mean = 0, sd = 1.58)
logitmu2[i]<-rnorm(1, mean = 0, sd = 1.58)
mu1[i]<-exp(logitmu1[i])/(1+exp(logitmu1[i]))
mu2[i]<-exp(logitmu2[i])/(1+exp(logitmu2[i]))
lower[i]<-max(0,(1/mu1[i])+(1/mu2[i])-(1/(mu1[i]*mu2[i])))
upper[i]<-min(1/mu1[i],1/mu2[i])
repeat {
tau[i]<-rlnorm(1,meanlog,sdlog)
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
```

```
}
return(data.frame(mu1,mu2,tau,lower,upper))
}

p10<-Prior10(10000,1,1)
mu1<-p10$mu1
mu2<-p10$mu2
tau10<-p10$tau
density<-plot(density(tau10,from=0),main="Prior 10",xlab="tau"
,xlim=c(0,10),ylim=c(0,3))
density<-plot(density(mu1,from=0,to=1),main="Prior 10",xlab="mu1"
,xlim=c(0,1))
density<-plot(density(mu2,from=0,to=1),main="Prior 10",xlab="mu2"
,xlim=c(0,1))

# Note: The prior formulations used for marginal homogeneity in
# Section 4.3.3 can be determined by simple adjustments to the
# above code. For example, for prior formulation 3:

Prior3MH<-function(n,shape,rate){
mu<-NULL
lower<-NULL
upper<-NULL
tau<-NULL
for (i in 1:n){
mu[i]<-runif(1,0,1)
lower[i]<-max(0,(1/mu[i])+(1/mu[i])-(1/(mu[i]*mu[i])))
upper[i]<-1/mu[i]
repeat {
tau[i]<-rgamma(1,shape,rate)
if(tau[i]>=lower[i]&tau[i]<=upper[i]) {break}
}
}
return(data.frame(mu,tau,lower,upper))
}

p3MH<-Prior3MH(10000,0.01,0.01)
muMH<-p3MH$mu
tau3MH<-p3MH$tau
density<-plot(density(tau3MH,from=0),main="Prior 3",xlab="tau"
```

```
,xlim=c(0,20),ylim=c(0,3))
density<-plot(density(muMH,from=0,to=1),main="Prior 3",xlab="mu"
,xlim=c(0,1))
```

Bugs code for the posterior formulations in Section 4.3.2

```
# Note: Although OpenBUGS was typically beneficial over WinBUGS,
# some models could only be fit in WinBUGS.

# data for the parental dataset:

list(x=c(29,33,131,78))

# Posterior 1 (Prior formulation 1):

model {
pi[1] <- tau*mu1*mu2
pi[2] <- mu1-(tau*mu1*mu2)
pi[3] <- mu2-(tau*mu1*mu2)
pi[4] <- 1+(tau*mu1*mu2)-mu1-mu2
x[1:4] ~ dmulti(pi[],n)
mu1 ~ dunif(0,1)
mu2 ~ dunif(0,1)
lower<-max(0,(1/mu1)+(1/mu2)-1/(mu1*mu2))
upper<-min(1/mu1,1/mu2)
tau ~ dunif(lower,upper)
n <- sum(x[]) }


# Initial values for the parameters:

list(mu1=0.5,mu2=0.5,tau=1)
list(mu1=0.1,mu2=0.1,tau=1)

# Posterior 2 (Prior formulation 2):

model {
pi[1] <- tau*mu1*mu2
pi[2] <- mu1-(tau*mu1*mu2)
```

```
pi[3] <- mu2-(tau*mu1*mu2)
pi[4] <- 1+(tau*mu1*mu2)-mu1-mu2
x[1:4] ~ dmulti(pi[],n)
# Normal distribution assumed for the logit of each marginal probability
# BUGS uses the precision=1/variance=1/2.5=0.4
a~dnorm(0,0.4)
b~dnorm(0,0.4)
mu1<-exp(a)/(1+exp(a))
mu2<-exp(b)/(1+exp(b))
lower<-max(0,(1/mu1)+(1/mu2)-1/(mu1*mu2))
upper<-min(1/mu1,1/mu2)
tau ~ dunif(lower,upper)
n <- sum(x[]) }

# Initial values for the parameters

list(a=0.5,b=0.5,tau=1)
list(a=0,b=0,tau=1)

# Posterior 3 (Prior formulation 5):

model {
pi[1] <- tau*mu1*mu2
pi[2] <- mu1-(tau*mu1*mu2)
pi[3] <- mu2-(tau*mu1*mu2)
pi[4] <- 1+(tau*mu1*mu2)-mu1-mu2
x[1:4] ~ dmulti(pi[],n)
mu1 ~ dunif(0,1)
mu2 ~ dunif(0,1)
lower<-max(0,(1/mu1)+(1/mu2)-1/(mu1*mu2))
upper<-min(1/mu1,1/mu2)
tau~dgamma(1,1)I(lower,upper)
n <- sum(x[]) }

# Posterior 4 (Prior formulation 6):

model {
pi[1] <- tau*mu1*mu2
pi[2] <- mu1-(tau*mu1*mu2)
pi[3] <- mu2-(tau*mu1*mu2)
```

```
pi[4] <- 1+(tau*mu1*mu2)-mu1-mu2
x[1:4] ~ dmulti(pi[],n)
# Normal distribution assumed for the logit of each marginal probability
# BUGS uses the precision=1/variance=1/2.5=0.4
a~dnorm(0,0.4)
b~dnorm(0,0.4)
mu1<-exp(a)/(1+exp(a))
mu2<-exp(b)/(1+exp(b))
lower<-max(0,(1/mu1)+(1/mu2)-1/(mu1*mu2))
upper<-min(1/mu1,1/mu2)
tau~dgamma(1,1)I(lower,upper)
n <- sum(x[]) }

# Posterior 5 (Prior formulation 9):

model {
pi[1] <- tau*mu1*mu2
pi[2] <- mu1-(tau*mu1*mu2)
pi[3] <- mu2-(tau*mu1*mu2)
pi[4] <- 1+(tau*mu1*mu2)-mu1-mu2
x[1:4] ~ dmulti(pi[],n)
mu1 ~ dunif(0,1)
mu2 ~ dunif(0,1)
lower<-max(0,(1/mu1)+(1/mu2)-1/(mu1*mu2))
upper<-min(1/mu1,1/mu2)
tau ~ dlnorm(1,1)I(lower,upper)
n <- sum(x[]) }

# Posterior 6 (Prior formulation 10):

model {
pi[1] <- tau*mu1*mu2
pi[2] <- mu1-(tau*mu1*mu2)
pi[3] <- mu2-(tau*mu1*mu2)
pi[4] <- 1+(tau*mu1*mu2)-mu1-mu2
x[1:4] ~ dmulti(pi[],n)
# Normal distribution assumed for the logit of each marginal probability
# BUGS uses the precision=1/variance=1/2.5=0.4
a~dnorm(0,0.4)
b~dnorm(0,0.4)
```

```
mu1<-exp(a)/(1+exp(a))
mu2<-exp(b)/(1+exp(b))
lower<-max(0,(1/mu1)+(1/mu2)-1/(mu1*mu2))
upper<-min(1/mu1,1/mu2)
tau ~ dlnorm(1,1)I(lower,upper)
n <- sum(x[]) }

# Note: The posterior distributions for marginal homogeneity in Section
# 4.3.3 can be determined by simple adjustments to the above code.
# For example, for posterior distribution 3 (prior formulation 5):

model {
pi[1] <- tau*pow(mu,2)
pi[2] <- mu-(tau*pow(mu,2))
pi[3] <- mu-(tau*pow(mu,2))
pi[4] <- 1+(tau*pow(mu,2))-(2*mu)
x[1:4] ~ dmulti(pi[],n)
mu ~ dunif(0,1)
lower<-max(0,(1/mu)+(1/mu)-(1/pow(mu,2)))
upper<-1/mu
tau~dgamma(1,1)I(lower,upper)
n <- sum(x[]) }

# Data for the arthritis dataset:

list(x=c(75,38,24,249))
```

Bugs code for the posterior distributions of the arthritis dataset by gender

```
# Data for the arthritis dataset by gender (xF=females, xM=males):

list(xF=c(37,23,8,104),xM=c(38,15,16,145))

# Posterior distributions for the saturated model (prior formulation 1):

model {
piF[1] <- tauF*mu1F*mu2F
piF[2] <- mu1F-(tauF*mu1F*mu2F)
```

```
piF[3] <- mu2F-(tauF*mu1F*mu2F)
piF[4] <- 1+(tauF*mu1F*mu2F)-mu1F-mu2F
piM[1] <- tauM*mu1M*mu2M
piM[2] <- mu1M-(tauM*mu1M*mu2M)
piM[3] <- mu2M-(tauM*mu1M*mu2M)
piM[4] <- 1+(tauM*mu1M*mu2M)-mu1M-mu2M
xF[1:4] ~ dmulti(piF[],nF)
xM[1:4] ~ dmulti(piM[],nM)
mu1F~ dunif(0,1)
mu2F ~ dunif(0,1)
lowerF<-max(0,(1/mu1F)+(1/mu2F)-1/(mu1F*mu2F))
upperF<-min(1/mu1F,1/mu2F)
tauF~dunif(lowerF,upperF)
mu1M~ dunif(0,1)
mu2M ~ dunif(0,1)
lowerM<-max(0,(1/mu1M)+(1/mu2M)-1/(mu1M*mu2M))
upperM<-min(1/mu1M,1/mu2M)
tauM ~ dunif(lowerM,upperM)
nF<- sum(xF[])
nM<- sum(xM[]) }
list(xF=c(37,23,8,104),xM=c(38,15,16,145))

# Initial values for the parameters

list(mu1F=0.5,mu2F=0.5,tauF=1,mu1M=0.5,mu2M=0.5,tauM=1)
list(mu1F=0.1,mu2F=0.1,tauF=1,mu1M=0.1,mu2M=0.1,tauM=1)

# Posterior distributions for the saturated model (prior formulation 2):

model {
piF[1] <- tauF*mu1F*mu2F
piF[2] <- mu1F-(tauF*mu1F*mu2F)
piF[3] <- mu2F-(tauF*mu1F*mu2F)
piF[4] <- 1+(tauF*mu1F*mu2F)-mu1F-mu2F
piM[1] <- tauM*mu1M*mu2M
piM[2] <- mu1M-(tauM*mu1M*mu2M)
piM[3] <- mu2M-(tauM*mu1M*mu2M)
piM[4] <- 1+(tauM*mu1M*mu2M)-mu1M-mu2M
xF[1:4] ~ dmulti(piF[],nF)
xM[1:4] ~ dmulti(piM[],nM)
```

```
aF~dnorm(0,0.4)
bF~dnorm(0,0.4)
mu1F<-exp(aF)/(1+exp(aF))
mu2F<-exp(bF)/(1+exp(bF))
lowerF<-max(0,(1/mu1F)+(1/mu2F)-1/(mu1F*mu2F))
upperF<-min(1/mu1F,1/mu2F)
tauF~dunif(lowerF,upperF)
aM~dnorm(0,0.4)
bM~dnorm(0,0.4)
mu1M<-exp(aM)/(1+exp(aM))
mu2M<-exp(bM)/(1+exp(bM))
lowerM<-max(0,(1/mu1M)+(1/mu2M)-1/(mu1M*mu2M))
upperM<-min(1/mu1M,1/mu2M)
tauM ~ dunif(lowerM,upperM)
nF<- sum(xF[])
nM<- sum(xM[]) }
list(xF=c(37,23,8,104),xM=c(38,15,16,145))

# Intial values for the parameters:

list(aF=0.1, bF=0.1,aM=0.1,bM=0.1,tauF=1,tauM=1)
list(aF=0.5, bF=0.5,aM=0.5,bM=0.5,tauF=1,tauM=1)

# Note: The models from Table 4.22 can be obtained
by adapting the code above.
```

R code for the simulation study conducted in Section 4.3.2.2

```
#load in required packages
library(MASS)
library(R2OpenBUGS)

# Number of datasets
B<-100
# Number of parameters
p=3

# Create matrices to store the simulation results

results_MEDIAN=matrix(0,nrow=B,ncol=p)
```

```
# R function to simulate the data

Sim<-function(mu1,mu2,tau,n){
pij = matrix(c(tau*mu1*mu2, mu2-tau*mu1*mu2,mu1-tau*mu1*mu2,
1-mu1-mu2+tau*mu1*mu2), nrow = 2, ncol = 2)
nij = rmultinom(1, size = n, prob = pij)
mat<-matrix(nij, nrow = nrow(pij))
return(c(nij[1],nij[3],nij[2],nij[4]))
}

# Set.seed() command used to ensure same datasets
# are used for each prior in each simulation scenario

set.seed(1)

for (i in 1:B) {
# simulate data for a given scenario
x <-Sim(0.2,0.2,1,50)
data<-list("x")
parameters <-c("mu1","mu2","tau")
init1 <-list(mu1=0.5,mu2=0.5,tau=1)
init2 <-list(mu1=0.1,mu2=0.1,tau=1)
#init1 <-list(a=0.5,b=0.5,tau=1)
#init2 <-list(a=0.1,b=0.1,tau=1)
print(x)
# Call OpenBUGS from R, .txt file changed for each prior
model<-bugs(data=data,inits=list(init1, init2),parameters.to.save=parameters,
model.file="C:/Users/Rob/Documents/PhD work/Paper 3/Simulations/Prior1.txt",
n.chains=2,n.iter=14000,n.burnin=4000,n.thin=1,
working.directory="C:/Users/Rob/Documents/PhD work/Paper 3/Simulations")
# Statistics of interest stored in matrices
results_MEDIAN[i,1:p]=model$summary[1:p,5]
}

# Calculate mean of the posterior medians
MEANMEDIAN = apply(results_MEDIAN,2,mean)

# True parameter values
TRUTH=c(0.2,0.2,1)

# Calculate bias
BIAS=MEANMEDIAN-TRUTH

STATS=t(rbind(TRUTH, MEANMEDIAN,BIAS))
rownames(STATS)<-c("mu1","mu2","tau")
colnames(STATS)<-c("TRUE", "MEANMEDIAN","BIAS")
summary=round(STATS,4)
```

```
print(summary)
#$

# Example OpenBUGS code for prior 1, saved as .txt file.

model {
pi[1] <- tau*mu1*mu2
pi[2] <- mu1-(tau*mu1*mu2)
pi[3] <- mu2-(tau*mu1*mu2)
pi[4] <- 1+(tau*mu1*mu2)-mu1-mu2
x[1:4] ~ dmulti(pi[],n)
mu1 ~ dunif(0,1)
mu2 ~ dunif(0,1)
lower<-max(0,(1/mu1)+(1/mu2)-1/(mu1*mu2))
upper<-min(1/mu1,1/mu2)
tau ~ dunif(lower,upper)
n <- sum(x[]) }
```

Chapter 5

Conclusions and future work

This thesis has extended the use of the dependence ratio association measure for multivariate categorical data to relevant applications, using both frequentist and Bayesian approaches. Paper 1 focused on square contingency tables as well as extensions to matched sets data (multi-way contingency tables). Existing models from the literature were discussed such as marginal homogeneity and symmetry with the symmetry model replicated directly in terms of dependence ratios for both the square tables and matched sets case.

For square tables there are only two subunits in a cluster. Although the dependence ratio approach is particularly beneficial for larger clusters sizes (as discussed in Section 1.3 of the introduction to this thesis), it still has a number of benefits even for the simpler case of square tables. For example, existing approaches to square tables that are based on odds ratios are criticised for not being as easy to interpret as association measures that are based on a ratio of probabilities such as the dependence ratio or the relative risk. The relative risk has been favoured over the odds ratio by a number of authors such as Sackett et al. (1996). Agresti (1999) states the importance of being able to communicate results to nonstatisticians. As noted in Jokinen et al. (2006), the dependence ratio is useful for achieving this goal, particularly to those familiar with the relative risk. Consequently, the dependence ratio approach to square tables may be preferred by nonstatisticians over existing approaches such as the marginal and loglinear approaches described in Sections 2.4.1 and 2.4.2.

In addition, a convenient feature of the dependence ratio approach for square tables is that constraints can easily be imposed on the dependence ratio parameters (using the R package drm) if a satisfactory fit cannot be achieved with any of the common models for square tables, such as marginal homogeneity and symmetry. The drm package (Jokinen, 2007) was a major breakthrough for the dependence ratio approach (discussed in Section 1.3 of the introduction to this thesis) since prior to this, the dependence ratio approach was not readily available for use in applied work. For example, consider the female (Stuart, 1955) and male (Stuart, 1953) eye grade datasets that were analysed in detail using the dependence ratio approach in paper 1. For the eye grade female dataset, a satisfactory fit was not achieved with the common models for square tables (marginal homogeneity, symmetry, quasi-symmetry and quasi-independence). However, a superior fit was achieved using the dependence ratio approach which combined a

proportional odds regression model with an association model that contained constraints on the dependence ratios. In contrast, the symmetry model was found to give a good fit for the eye grade males dataset.

A notable advantage of the dependence ratio approach is its ability to cope with larger cluster sizes than approaches based on a odds ratio parameterisation. Lesaffre et al. (2000) state that maximum likelihood estimation using odds ratios as the association is typically not feasible for cluster sizes greater than five. In the context of paper 1, this was particularly advantageous for extensions of the dependence ratio approach from square tables to matched sets data. Rater agreement had not previously been considered with a dependence ratio approach until the analysis of the pathologist dataset. The dependence ratio approach allowed for an assessment of both the marginal distributions of the raters (pathologists) and the associations between the pathologists ratings. In terms of the former, marginal homogeneity was not found to hold between the pathologists. In terms of the latter, a latent binary factor association structure was found to give the best fit of the association structures considered. There appeared to be two distinct groups, one that accounted for 44% of the population with a low presence rate and the other that accounts for 56% of the population with a much higher presence rate.

The models for square tables and matched sets in paper 1 of this thesis did not consider the inclusion of additional explanatory variables. Paper 2 of this thesis focused on a specific dataset that contained patients with psoriatic arthritis in their hand joints (14 in each hand). The response of interest (for each joint of each patient, at their last clinic visit) was the binary presence of clinical damage. The dependence ratio approach has been used to analyse datasets with large cluster sizes and explanatory variables before, such as the government spending dataset in Jokinen et al. (2006). However, the purpose of paper 2 was to further advocate the use of the dependence ratio approach for analysing both the marginal regression and the associations within a cluster (patient) for datasets with large cluster sizes that would otherwise be difficult to fit with alternative approaches based on a odds ratio parameterisation. Although the GEE approach can cope with larger cluster sizes (as shown in paper 2 of this thesis), the fact its emphasis is purely on the marginal regression means the dependence ratio approach provides a more thorough analysis. In addition to the dependence ratio approach, a GEE analysis was also conducted to provide a comparison. This was performed separately for both all the patients in the dataset and only those patients with at least one joint damaged, with the same final model obtained in both cases (although clearly clinical damage probabilities were higher for the latter). The conclusions obtained from the GEE approach generally coincided with previous literature. For example, time since diagnosis was found to significantly predict clinical damage with increased time since diagnosis associated with a greater probability of damage. The DIP joints, which were known to be commonly affected in psoriatic arthritis, were associated with higher probabilities of clinical damage than the MCP and PIP joints, for each finger (excluding the thumb which has no DIP

joint). However, they were only significantly higher than the MCP joints (for each finger). More detailed conclusions are given in paper 2.

Although the GEE analysis allowed for a detailed assessment of the marginal patterns of clinical damage, some conclusions were missed by not properly taking into account the associations within a patient. The exploratory analysis of the associations in paper 2 showed that the strongest association patterns of clinical damage existed between the MCP knuckle joints. In other words, although the DIP joints had the highest marginal probabilities of clinical damage, the MCP knuckle joints had the strongest associations. This underlines the importance of modelling the full joint distribution in order to obtain a complete analysis of the data. The dependence ratio approach considered separate models for each of the MCP knuckle joints, thumb joints, DIP joints and PIP joints (each with relevant explanatory variables). The model for the MCP knuckle joints was useful to consider given the strong association patterns they conveyed. Although no notable association patterns existed amongst the other joints, separate models were considered for the thumb, DIP and PIP joints. The model for the DIP joints is perhaps of particular interest given their strong marginal patterns of clinical damage and the fact they are commonly affected in psoriatic arthritis. It is noted that although these models do not provide an assessment of the associations between the different joint types, this was assessed in an exploratory sense. Paper 2 was also useful for showing the use of plausible association structures, a notable advantage of the dependence ratio approach. For example, the necessary factor association structure was found to be most appropriate for the MCP knuckle joints, which is not surprising given the fact that 92% of patients had none of their MCP knuckle joints damaged. The model also accounts for the stronger association amongst these joints.

Paper 3 of this thesis provided the first known use of the dependence ratio using traditional Bayesian approaches whereby prior distributions are assigned to the parameters as opposed to the empirical Bayes approaches by Good (1956) and Du Mouchel and Pregibon (2001) which estimate the prior parameters from the data. The primary focus of paper 3 was on identifying non-informative prior distributions for the marginal probabilities and the single dependence ratio (τ) in the simplest bivariate binary case, with emphasis placed on constraining τ to be between its lower and upper bounds. Of the 10 prior formulations considered, the 2 most appropriate non-informative formulations both had τ being assigned a uniform distribution between its bounds, with the marginal probabilities being assigned Uniform(0,1) and logistic-normal(0,2.5) distributions respectively. The same conclusions were also obtained for the marginal homogeneity case in which there is a single parameter for the marginal probabilities.

Extensions from the bivariate binary response were also considered. Firstly, the extension to include a single explanatory variable with two categories. The arthritis dataset that was previously considered for posterior inference of the marginal homogeneity model was extended to account for gender, with the chosen non-informative priors used for the

marginal probabilities and dependence ratios. Interestingly, the conclusions obtained from the Bayesian analysis of the arthritis dataset in paper 3 coincided with the conclusions from the frequentist analysis in paper 2. The Bayesian analysis found a model with a single dependence ratio and single marginal probability to be the most appropriate model to consider of those in Table 4.22. In other words, neither the association or the marginal patterns differed by gender and the marginal probabilities did not differ by hand. This coincided with paper 2 in the sense that the association structures considered for the dependence ratio approach were not found to vary by gender. In addition, gender and hand were not found to significantly predict clinical damage in paper 2. However, it should be noted that paper 2 conducted a more thorough analysis than paper 3 in the sense that more explanatory variables were included and the response of interest was clinical damage at each individual joint as opposed to the probability of having clinical damage in at least one joint of a given hand. A potential area for future work is to use the Bayesian approach to conduct a full analysis of the arthritis dataset. Other extensions for the Bayesian approach include the inclusion of more explanatory variables, larger contingency tables and considering association structures such as the necessary factor. The extension to larger contingency tables was discussed briefly at the end of paper 3. It is of interest to determine whether the same conclusions are drawn with regards to non-informative priors in this case. The Bayesian approach with dependence ratios also has the advantage over frequentist approaches of not relying on asymptotic approximations. For example, the smaller sample size of the bilirubin dataset in paper 1 made the asymptotic approximations used more questionable.

The dependence ratio approach has some limitations, none of which should be treated as serious disadvantages but should be properly understood before the approach can be adequately implemented. Firstly, the maximum likelihood estimation procedure used in the frequentist context of papers 1 and 2 can encounter negative profile probabilities for some of the unobserved profiles. Although some may argue that this is a disadvantage of the approach, it can also be seen as a tool for model validation (Ekholm, 2003). For the occasions where negative profile probabilities were encountered in papers 1 and 2, the models in question were clearly disregarded outright. As argued in Jokinen (2006: PhD Thesis), negative profile probabilities typically only occur in datasets with large cluster sizes, which cannot be fit using an odds ratio parameterisation anyway. An interesting future investigation is whether models that produced negative profile probabilities in a frequentist context can be fit using Bayesian approaches. In terms of the arthritis dataset from paper 2, the final models had a maximum of 8 units in a cluster. In order to assess the strength of the dependence ratio approach, some additional analyses found the approach was viable for cluster sizes as large as approximately 13 but was problematic for larger cluster sizes, partly due to negative profile probabilities. This is still a considerable improvement over odds ratio approaches which struggle with cluster sizes larger than five (Lesaffre et al. (2000)).

The range of the dependence ratio has received criticism due to the fact that it is constrained by the marginal probabilities. Research in computer science (discussed in the history of dependence ratio section in the introduction of this thesis) found that this was only an issue for small counts. In addition, as demonstrated in Section 4.3.4 and discussed in Ekholm (2003), the range of the dependence ratio is reasonably variation independent of the marginal probabilities when the marginal probabilities are less than approximately 0.5. This therefore supports modelling the rarer event in the dependence ratio approach. For example, in paper 2 of this thesis, the presence of clinical damage (the rarer event) was treated as the response of interest as opposed to the absence of clinical damage. The rarer event was also used in the bilirubin dataset (paper 1) as well as the parental and arthritis datasets in paper 3. The fact the dependence ratio has a finite upper bound can also be seen as an advantage over the odds ratio which has an infinite upper bound.

To conclude, the dependence ratio approach has a number of advantages for modelling multivariate categorical data using population-averaged based approaches. Paper 1 of this thesis applied the dependence ratio to square tables, which may be particularly advantageous for researchers who are more familiar with concepts based on probabilities (such as the relative risk) as opposed to odds ratios. The extension of the approaches for square tables to matched sets data also showed the benefit of the dependence ratio approach for analysing rater agreement and large cluster sizes, something which is likely to be unfeasible with approaches based on a odds ratio parameterisation. Paper 2 of this thesis considered the dependence ratio approach for a specific application and further showed the benefits the approach can offer for analysing larger cluster sizes. Although a fair amount of research has been conducted on the dependence ratio in a frequentist context using maximum likelihood estimation (see Section 1 of paper 1 for a discussion), including papers 1 and 2 of this thesis, little has been done by researchers who were not the original founders of the approach or actively involved with its development. This may in part be due to the fact that the GEE approach is the only tool some researchers have at their disposal in certain software, such as SPSS. However, the dependence ratio has clear advantages for analysing population-averaged approaches in a frequentist context, particularly in situations where the researcher has a large number of units in a cluster and interest in modelling the associations such as the arthritis dataset in paper 2. Paper 3 of this thesis provided the first use of the dependence ratio in a Bayesian context (excluding empirical Bayes approaches). Although paper 3 provided a starting point for the Bayesian context, it is clear that more research is needed in order to make the dependence ratio appropriate for a range of possible datasets.

References

Aggarwal, C. C. and Yu, P. S. (1998). A New Framework for Item Set Generation. *In: Proceedings of the ACM PODS Symposium on Principles of Database Systems, Seattle, Washington (USA)*, pages 18–24.

Agrawal, R. Imielinski, T. and Swami, A. N. (1993). Mining association rules between sets of items in large databases. *In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington (USA)*, pages 207–216.

Agresti, A. (1983). A simple diagonals-parameter symmetry and quasi-symmetry model. *Statistics and Probability Letters*, 1:313–316.

Agresti, A. (1999). Modelling Ordered Categorical Data: Recent Advances and Future Challenges. *Statistics in Medicine*, 18:2191–2207.

Agresti, A. (2002). *Categorical Data Analysis, 2nd edition.* John Wiley and Sons, Inc, New Jersey.

Agresti, A. and Hitchcock, D. B. (2005). Bayesian Inference for Categorical Data Analysis. *Statistical Methods and Applications*, 14:297–330.

Agresti, A. and Lang, J. B. (1993). Quasi-symmetric latent class models, with application to rater agreement. *Biometrics*, 49:131–139.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *B. N. Petrov, and F. Csaki (Editors), Second International Symposium on Information Theory. Akademiai Kiado, Budapest*, pages 267–281.

Albert, J. (2010). Good Smoothing. *In: Frontiers of Statistical Decision Making and Bayesian Analysis, M.-H. Chen, D.K.Dey, P.Muller, D.Sun, and K.Ye (Editors)*, New York: Springer:419–436.

Albert, J. H. and Gupta, A. K. (1982). Mixtures of Dirichlet distributions and estimation in contingency tables. *The Annals of Statistics*, 10:1261–1268.

Albert, J. H. and Gupta, A. K. (1983a). Estimation in contingency tables using prior information. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45:60–69.

Albert, J. H. and Gupta, A. K. (1983b). Bayesian estimation methods for 2 by 2 contingency tables using mixtures of Dirichlet distributions. *Journal of the American Statistical Association*, 78:708–717.

Altham, P. M. E. (1969). Exact Bayesian analysis of a 2 by 2 contingency table, and Fisher's exact significance test. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31:261–269.

Anatolyev, S. (2008). Multi-market direction-of-change modeling using dependence ratios. *Studies in Nonlinear Dynamics and Econometrics*, 13:1–24.

Bahadur, R. R. (1961). *A representation of the joint distribution of responses to n dichotomous items. In: Studies in Item Analysis and Prediction.* H. Solomon (Editor), Stanford University Press, Stanford, California.

Bayes, T. (1763). An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53:370–418.

Beh, E. J. (2004). Simple Correspondence Analysis: A Bibliographic Review. *International Statistical Review*, 72:257–284.

Beh, E. J. (2008). Simple Correspondence Analysis of Nominal-Ordinal Contingency Tables. *Journal of Applied Mathematics and Decision Sciences.*

Bergsma, W. P. (1997). *Marginal models for categorical data.* Tilburg University Press, Tilburg, Netherlands.

Bhapkar, V. P. (1966). A note on the equivalence of two test criteria for hypotheses in categorical data. *Journal of the American Statistical Association*, 61:228–235.

Bibby, J. (1975). Methods of Measuring Mobility. *Quality and Quantity*, 9:107–136.

Bishop, Y.M.M. Fienberg, S. and Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge.

Bloch, D. A. and Watson, G. S. (1967). A Bayesian study of the multinomial distribution. *The Annals of Mathematical Statistics*, 38:1423–1435.

Bond, S. J. et al. (2007). Predictors for radiological damage in psoriatic arthritis: results from a single centre. *Annals of the Rheumatic Diseases*, 66:370–376.

Bond, S. J. and Farewell, V. F. (2009). Likelihood estimation for a longitudinal negative binomial regression model with missing outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58:369–382.

Bouma, G. (2009). Normalized (Pointwise) Mutual Information in Collocation Extraction. *In: Chiarcos, Eckart de Castilho and Stede (Editors), Von der Form zur Bedeutung: Texte automatisch verarbeiten / From Form to Meaning: Processing Texts Automatically, Proceedings of the Biennial GSCL Conference 2009*, Tubingen; Gunter Narr Verlag:pages 31–40.

Braungart, R. G. (1971). Family status, socialization, and student politics: a multivariate analysis. *American Journal of Sociology*, 77:108–130.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research: Vol. 1 - The Analysis of Case-Control Studies*. Lyon, France, IARC Scientific Publications.

Brin, S. Motwani, R. and Silverstein, C. (1997). Beyond Market Baskets: Generalizing Association Rules to Dependence Rules. *In: Data Mining and Knowledge Discovery*, 2:39–68.

Carlsson, G. (1958). Social Mobility and Class Structure. *Lund, Sweden. Gleerup.*

Caussinus, H. (1966). Contribution à l'analyse statistique des tableaux de corrélation. Annales de la Faculté des Sciences de l'Université de Toulouse. *Psychometrika*, 12:153–157.

Chaganty, N. R. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society: Series B (Methodological)*, 66:851–860.

Church, K. W. and Hanks, P. (1990). Word Association Norms, Mutual Information, And Lexicography. *Computational Linguistics*, 16:22–29.

Cornfield, J. (1966). A Bayesian test of some classical hypotheses - With applications to sequential clinical trials. *Journal of the American Statistical Association*, 61:577–594.

Cresswell, L. and Farewell, V. F. (2010). Assessment of joint symmetry in arthritis. *Statistics in Medicine*, 30:973–983.

Davies, H. T. O. Crombie, I. K. and Tavakoli, M. (1998). When can odds ratios mislead? *British Medical Journal*, 316:989–991.

Diggle, P. J. et al. (2002). *The Analysis of Longitudinal Data, second edition.* Oxford University press, Oxford.

Duffy, D. L. et al. (2001). Human twinning is not linked to the region of chromosome 4 syntenic with the sheep twinning gene fecb. *American Journal of Medical Genetics*, 100:182–186.

DuMouchel, W. (1999). Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous Reporting System (with discussion). *The American Statistician*, 53:177–202.

DuMouchel, W. and Pregibon, D. (2001). Empirical Bayes Screening for Multi-Item Associations. proceedings of the 7th acm sigkdd international conference on knowledge. *Discovery and Data Mining, San Francisco (USA)*, pages 67–76.

Durbin, J. (1955). Appendix note on a statistical question raised in the preceding paper. *Population Studies*, 9:101.

Ekholm, A. (2003). Comparing the odds and the dependence ratios. *In: Statistics, Econometrics, and Society: Essays in Honour of Leif Nordberg. Statistics Finland, Helsinki*, pages 13–25.

Ekholm, A. et al. (2003). Joint regression and association modelling for longitudinal ordinal data. *Biometrics*, 59:795–803.

Ekholm, A. and Skinner, C. (1998). The Muscatine children's obesity data reanalysed using pattern mixture models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47:251–263.

Ekholm, A. Jokinen, J. and Kilpi, T. (2002). Combining regression and association modelling for longitudinal data on bacterial carriage. *Statistics in Medicine*, 21:773–791.

Ekholm, A. Smith, P. W. F. and McDonald, J. W. (1995). Marginal regression analysis of a multivariate binary response. *Biometrika*, 82:847–854.

Ekholm, A. Smith, P. W. F. and McDonald, J. W. (2000). Association models for a multivariate binary response. *Biometrics*, 56:712–718.

Fay, M. P. (2011). *Exact McNemar's Test and Matching Confidence Intervals*. http://cran.r-project.org/web/packages/exact2x2/vignettes/exactMcNemar.pdf.

Fisher, R. A. (1956). *Statistical Methods for Scientific Inference*. Edinburgh: Oliver and Boyd.

Fitzmaurice, G. M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, 80:141–151.

Forster, J. J. (2004a). Bayesian inference for square contingency tables. *Working paper*.

Frechet, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Journal of the Royal Statistical Society: Series B (Methodological)*, 14:53–77.

Fredette, M. and Angers, J.-F. (2002). A new approximation of the posterior distribution of the log-odds ratio. *Statistica Neerlandica*, 56:314–329.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

Gilks, W. R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Samping. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41:337–348.

Gladman, D. D. (2006). Clinical, radiological, and functional assessment in psoriatic arthritis: Is it different from other inflammatory joint diseases? *Annals of the Rheumatic Diseases*, 65 (Supp III):22–24.

Gladman, D. D. et al. (1990). Reliability of measurements of active and damaged joints in psoriatic arthritis. *The Journal of Rheumatology*, 17:62–64.

Gladman, D. D. et al. (2004). Assessment of Patients With Psoriatic Arthritis. A review of currently available measures. *American College of Rheumatology*, 50:24–35.

Gladman, D. D. et al. (2005). Psoriatic arthritis: epidemiology, clinical features, course, and outcome. *Annals of the Rheumatic Diseases*, 64 (Supp II):14–17.

Gladman, D. D. Farewell, V. F. and Nadeau, C. (1995). Clinical indicators of progression in psoriatic arthritis: multivariate relative risk model. *The Journal of Rheumatology*, 22:675–679.

Glass (Editor), D. V. (1954). *Social Mobility in Britain.* London: Routledge and Kegan Paul.

Good, I. J. (1956). On the estimation of small frequencies in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 18:113–124.

Good, I. J. (1967). A Bayesian significance test for multinomial distributions (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 29:399–431.

Goodman, L. A. (1979a). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74:537–552.

Goodman, L. A. (1979b). Multiplicative models for square contingency tables with ordered categories. *Biometrika*, 66:413–418.

Goodman, L. A. (1996). A single general method for the analysis of cross-classified data: Reconciliation and synthesis of some methods of Pearson, Yule, and Fisher, and also some methods of correspondence analysis and association analysis. *Journal of the American Statistical Association*, 91:408–427.

Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology*, 125:761–768.

Guo, S. W. (2000). Familial aggregation of environmental risk factors and familial aggregation of disease. *American Journal of Epidemiology*, 151:1121–1131.

Haldane, J. B. S. (1948). The precision of observed values of small frequencies. *Biometrika*, 35:297–303.

Hanley, J. A. Negassa, A. and Edwardes, M. D. (2000). GEE Analysis of negatively correlated binary responses: a caution. *Statistics in Medicine*, 19:715–722.

Helliwell, P. et al. (2000). Joint symmetry in early and late rheumatoid and psoriatic arthritis. *Arthritis Rheumatology*, 43:865–871.

Jeffreys, H. (1946). An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186:453–461.

Jokinen, J. (2006). Fast estimation algorithm for likelihood-based analysis of repeated categorical responses. *Computational Statistics and Data Analysis*, 51:1509–1522.

Jokinen, J. (2006: PhD Thesis). *Joint regression and association models for repeated categorical responses.* PhD thesis, University of Helsinki, Helsinki.

Jokinen, J. (2007). *drm: Regression and association models for repeated categorical data.* R package version 0.5-8, http://CRAN.R-project.org/package=drm.

Jokinen, J. McDonald, J. W. and Smith, P. W. F. (2006). Meaningful regression and association models for clustered ordinal data. *Sociological Methodology*, 36.

Kastner, C. Fieger, A. and Heumann, C. (1997). MAREG and WinMAREG A tool for marginal regression models. *Computational Statistics and Data Analysis*, 24:237–241.

Keynes, J. M. (1921). *A Treatise on Probability.* London: Macmillan and Co.

Knuiman, M. W. and Speed, T. P. (1988). Incorporating prior information into the analysis of contingency tables. *Biometrics*, 44:1061–1071.

Kullback, S. (1971). The Homogeneity of the Sex Ratio of Adjacent Sibs in Human Families. *Biometrics*, 27:452–457.

Lambert, P. et al. (2005). How vague is vague? a simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, 24:2401–2428.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174.

Lang, J. B. (2004). Maximum likelihood fitting of multinomial-Poisson homogeneous (MPH) models for contingency tables using MPH.FIT. *Online HTML document, http://www.cs.uiowa.edu/ jblang/mph.fitting/mph.fit.documentation.htm.*

Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les événements. *Mém. Acad. R. Sci, Paris*, 6:621–656.

Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika*, 59:581–589.

Leonard, T. (1973). A Bayesian method for histograms. *Biometrika*, 60:297–308.

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

Lesaffre, E. and Spiessens, B. (2000). Discussion of the paper by Heagerty P.J., Zeger S.L. Marginalized multilevel models and likelihood inference. *Statistical Science*, 15:1–26.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73:13–22.

Liang, K. Y. Zeger, S. L. and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54:3–40.

Lindley, D. V. (1964). The Bayesian analysis of contingency tables. *The Annals of Mathematical Statistics*, 35:1622–1643.

Lindsey, J. K. and Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, 17:447–469.

McCullagh, P. (1978). A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*, 65:413–418.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12:153–157.

Muller, H. J. (1916). The mechanism of crossing over. *The American Naturalist*, 50:193–221, 284–305, 350–366, 421–434.

Neal, R. M. (2003). Slice Sampling. *The Annals of Statistics*, 31:705–767.

Novick, M. R. (1969). Multiparameter Bayesian indifference procedure (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 31:29–64.

O'Keeffe, A. G. Tom, B. D. M. and Farewell, V. F. (2011). A case-study in the clinical epidemiology of psoriatic arthritis: multistate models and causal arguments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60:675–699.

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, 57:120–125.

Pantel, P. and Lin, D. (2002). Discovering Word Senses from Text. *In: Proceedings of ACM Conference on Knowledge Discovery and Data Mining (KDD-02)*, pages 613–619.

Perks, W. (1947). Some observations on inverse probability including a new indifference rule. *Journal of the Institute of Actuaries*, 73:285–334.

Plackett, R. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60:516–522.

Raftery, A. E. (1986). A note on Bayes factors for log-linear contingency table models with vague prior information. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48:249–250.

Risch, N. (1990a). Linkage Strategies for Genetically Complex Traits. I. Multilocus models. *The American Journal of Human Genetics*, 42:222–228.

Risch, N. (1990b). Linkage Strategies for Genetically Complex Traits. II. Multilocus models. *The American Journal of Human Genetics*, 42:229–241.

Risch, N. (1990c). Linkage Strategies for Genetically Complex Traits. III. Multilocus models. *The American Journal of Human Genetics*, 42:242–253.

Rogoff, N. (1953). Recent Trends in Occupational Mobility. *New York, Arno Glencoe Free Press.*

Sackett, D. L. Deeks, J. J. and Altman, D. G. (1996). Down with odds ratios! *Evidence-Based Medicine*, 1:164–166.

Schwarz, G. (1978). Estimating The Dimension Of A Model. *The Annals of Statistics*, 6:461–464.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423.

Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication. Urbana.* University of Illinois Press.

Siannis, F. et al. (2006). Clinical and radiological damage in psoriatic arthritis. *Annals of the Rheumatic Diseases*, 65:478–481.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Spiegelhalter, D. J. et al. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Methodological)*, 64:583–639.

Stuart, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika*, 40:105–110.

Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, 42:412–416.

Sun, X. and Yang, Z. (2008). Statistics and Data Analysis, Paper 382, Generalized McNemars Test for Homogeneity of the Marginal Distributions. *SAS Global Forum.*

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 76:103–154.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics,*, 22:1701–1728.

Tomizawa, S. (1989). Decompositions for conditional symmetry model into palindronic symmetry and modified marginal homogeneity models. *Australian Journal of Statistics*, 31:287–296.

Tyree, A. (1973). Mobility Ratios and Association in Mobility Tables. *Population Studies*, 27:577–588.

Van de Cruys, T. (2011). Two Multivariate Generalizations of Pointwise Mutual Information. *Proceedings of the ACL workshop. Distributional Semantic and Compositionality (DISCO), Portland, Oregon, USA*, pages 16–20.

Vounatsou, P. and Smith, A. (1996). Bayesian analysis of contingency tables: A Simulation and graphics-based approach. *Statistics and Computing*, 6:277–287.

Wacholder, S. (1996). The Case-Control Study as Data Missing by Design: Estimating Risk Differences. *Epidemiology*, 7:144–150.

Walker, G. A. (2002). *Common Statistical Methods for Clinical Research with SAS examples, 2nd edition.* SAS Institute Inc, Cary, North Carolina.

Wallace, C. and Clayton, D. (2003). Estimating the Relative Recurrence Risk Ratio Using a Global Cross-Ratio Model. *Genetic Epidemiology*, 25:293–302.

Wallace, C. Clayton, D. and Fine, P. (2003). Estimating the relative recurrence risk ratio for leprosy in Karonga District, Malawi. *Leprosy Review*, 74:133–140.

Yule, G. U. (1912). On the methods of measuring the association between two attributes. *Journal of the Royal Statistical Society*, 75:579–652.

Zellner, A. and Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, 25:365–393.