PARAMETER REDUNDANCY IN LOG-LINEAR MODELS

Serveh Sharifi Far

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2017

Identifiers to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/10023-11739
http://hdl.handle.net/10023/11739

Parameter Redundancy in Log-linear Models

University of
St Andrews

Serveh Sharifi Far

Thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

in the School of Mathematics and Statistics

UNIVERSITY OF ST ANDREWS

August 2017

Declaration

1. Candidate's declarations:
I, Serveh Sharifi Far, hereby certify that this thesis, which is approximately 30,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.
I was admitted as a research student and as a candidate for the degree of PhD in Statistics in September 2013; the higher study for which this is a record was carried out in the University of St Andrews between 2013 and 2017.

Date ————- signature of candidate ————-

2. Supervisor's declaration:
I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ————- signature of supervisor ————-

3. Permission for publication:
In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions

that may be required in order to allow such access and migration, or have requested the appropriate embargo below.
The following is an agreed request by candidate and supervisor regarding the publication of this thesis:
Embargo on all of both print copy and electronic copy for a period of two years on the following ground: publication would preclude future publication.

Date ————- signature of candidate ————-

signature of supervisor ————

Serveh Sharifi Far
August 2017

Acknowledgements

I would like to express my gratitude to my supervisors Michail Papathomas and Ruth King, for their constant guidance, support and encouragement along this work. They devoted countless hours to meetings, discussions and answering queries, and without them this thesis would not have been possible.

I am extremely grateful to the School of Mathematics and Statistics, and the Engineering and Physical Sciences Research Council (EPSRC), for the financial support of my studies. A very special appreciation goes to everyone in CREEM for their friendly manner and assistance. Many thanks to my officemates, and friends, Claudia, Christina, Andreia, Ameneh, and Laleh for their friendship and camaraderie.

I want to thank my mother Nina, my brother Siamand and his wife Azadeh, for encouraging me in my academic development. Thanks to my sister and my best friend Silvana, and my nieces Viyana and Rozhina who made my vacations back home relaxing and unforgettable. Special thanks to my aunts Esmat and Amin for all their support and kindness. Finally, I gratefully acknowledge my late father who taught me to enjoy the endless world of research, reading and learning.

Abstract

Log-linear models are widely used to analyse categorical variables arranged in a contingency table. Sampling zero entries in the table can cause the problem of large standard errors for some model parameter estimates. This thesis focuses on the reason of this problem and suggests a solution by utilising the parameter redundancy approach. This approach detects whether a model is non-identifiable and parameter redundant, and specifies a smaller set of parameters or combinations of them that all are estimable. The parameter redundancy method is adapted here for Poisson log-linear models which are parameter redundant because of the number and pattern of the zero observations in the contingency table. Furthermore, it is shown that in some parameter redundant log-linear models, the presence of constraints referred to as esoteric constraints can make more parameters estimable. It is proven in a theorem that for a saturated Poisson log-linear model fitted to an l^m table with one zero cell count, which model parameters are not estimable. Three examples of real data in sparse contingency tables are presented to demonstrate the process of identifying the estimable parameters and reducing the model.

An alternative approach is the existence of the MLE method that checks for the existence of the maximum likelihood estimates of the cell means in the log-linear model after observing the zero entries. The method considers the log-linear model as a polyhedral cone and provides conditions to detect the estimability of the cell means. This method is compared here with the parameter redundancy approach and their similarities and differences are explained and illustrated using examples. In parameter redundant models with existent MLE, it is observed that the presence of the esoteric constraints makes all the parameters estimable.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Categorical variables and contingency tables

Statistical methods for analysing categorical data have always been of interest to statisticians and relevant methodology has mostly been developed since the beginning of the 20th century. A **categorical variable** is used to describe qualitative or quantitative data that are classified in a set of possible categories or levels. Categorical variables are commonly used in the social and biomedical sciences for measuring attitudes and various states of variables. Their application, however, is not only restricted to these areas and they occur in ecological, medical and even engineering and industrial sciences. There are three types of categorical variables based on the nature of the set of categories. When the categories of a variable do not have an intrinsic order, the variable is called nominal. For instance, eye colour, gender, and blood type are considered nominal categorical variables. In contrast, variables which do have naturally ordered categories, such as education level and social class, are named ordinal variables. Categorical variables for which there is a numerical distance between two levels is an interval variable. An interval variable is usually a categorised continuous or discrete numerical variable, for example, age, and years of education are interval variables.

A display format for data with categorical measurements is a **contingency table**, first introduced by Karl Pearson in 1904 [Agresti, 2002]. It is designed to assist with the detection of the relationship between two or more categorical variables. In such a table, subjects are cross-classified over different categories of variables, and each cell count represents the number of subjects under a certain cross-classification. A contingency table for two variables X and Y, with I rows and J columns, is referred to as a two-way $I \times J$ table. Table 1.1 is a 2×3 contingency table, taken from Agresti [2002], which shows the relationship between two nominal categorical variables, aspirin use and heart

	Heart attack		
	Fatal	Non-fatal	None
Placebo	18	171	10,845
Aspirin	5	99	10,933

Table 1.1 2×3 contingency table of aspirin use and heart attack.

attack. The joint probability distribution π_{ij} indicates the probability that X and Y realisation belongs to the cell at row i and column j. The marginal distribution for each variable is derived by summing the joint probabilities over the other variables. For a two-way table, the marginal distributions are given by,

$$\pi_{i+} = \sum_{j=1}^{J} \pi_{ij}, \qquad \pi_{+j} = \sum_{i=1}^{I} \pi_{ij},$$

such that $\sum_{i=1}^{I} \sum_{j=1}^{J} \pi_{ij} = \sum_{i=1}^{I} \pi_{i+} = \sum_{j=1}^{J} \pi_{+j} = 1$.

Each cell count in a contingency table can be viewed as a random variable with non-negative integer possible values. For a table with n cells, we denote these random variables as Y_i and their observed values as y_i, $i = 1, \ldots, n$. The distribution of these variables depends on the assumed **sampling distribution**. The two commonly used sampling distributions for the cell counts of a contingency table are the Poisson and multinomial distributions.

The Poisson distribution is the most common distribution for count data. A Poison random variable indicates the number of independent events occurring in a fixed period of time or interval of space. Its probability mass function is,

$$P(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \qquad y_i = 0, 1, 2, \ldots, \tag{1.1}$$

where $\mu_i > 0$ and $\mu_i = \mathrm{E}(Y_i) = \mathrm{Var}(Y_i)$. The joint probability distribution for n independent variables is the product of their probability mass functions.

When the total sample size is fixed, so that $N = \sum_{i=1}^{n} y_i$ is known, the sampling distribution for the table counts is the multinomial distribution. The probability mass function for the n cells of the table is,

$$P\left(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid \sum_{i=1}^{n} y_i = N\right) = \frac{N!}{\prod_{i=1}^{n} y_i!} \prod_{i=1}^{n} \pi_i^{y_i}, \tag{1.2}$$

in which $y_i = 0, 1, 2, \ldots,$ and π_i is the probability that a subject belongs to cell i. A special case of multinomial sampling happens when instead of having a fixed total sample size, the number of observations in each row or column is fixed. This sampling distribution is called "independent multinomial" or "product multinomial" sampling. The multinomial distribution for the cell counts then becomes the product of the multinomial distributions for the rows or columns.

1.2 Log-linear models for contingency table data

In most statistical analyses there is a distinction between dependent or response variables and independent or explanatory variables. The aim is usually to describe the effects of some of the independent variables on the dependent ones. Some prominent models for categorical data belong to the family of **generalized linear models** (GLM). The dependent variable in a generalized linear model is assumed to be from the exponential family distributions. The probability function for a variable with n independent values from a member of the exponential family distributions is,

$$f(y_i|\lambda_i, \phi) = \exp\left\{\frac{y_i\lambda_i - b(\lambda_i)}{a(\phi)} + c(y_i, \phi)\right\}. \tag{1.3}$$

In such a function, ϕ is referred to as the dispersion parameter and λ_i is called the natural parameter. We also know that for these distributions, $\mu_i = \mathrm{E}(Y_i) = b'(\lambda_i)$ [Agresti, 2002]. In a GLM, a link function (g) relates the expected values of the observations to the independent variables by the formula,

$$g(\boldsymbol{\mu}) = \mathsf{A}\boldsymbol{\theta}. \tag{1.4}$$

$\mathsf{A}\boldsymbol{\theta}$ terms are known as the linear predictors. In this thesis, we show matrices with upper case letters and show vectors with bold symbols or letters. A is the design matrix or model matrix and includes the values of the independent variables and $\boldsymbol{\theta}$ is the vector of model parameters. The model's parameter estimates are obtained by maximising the likelihood function and often applying the Newton-Raphson algorithm or the iterative weighted least square method.

The cell counts in a contingency table are often assumed to be from the Poisson distribution, which is a member of the exponential family of distributions with the dispersion parameter $\phi = 1$ and the natural parameter $\lambda_i = \log \mu_i$. Then for each of the n cell counts in a contingency table, the probability mass function is as equation (1.1). The natural GLM link function for a Poisson distribution is the log function. So the

generalised linear model in this case is,

$$\log(\boldsymbol{\mu}) = \mathsf{A}\boldsymbol{\theta}.$$

This model is called a **log-linear model**, in which means of cell counts are related to model parameters $\boldsymbol{\theta}$ via design matrix A. Log-linear models specify the association patterns among the categorical variables and in these models there is no distinction between dependent and independent variables. If such a distinction is required then logit models or logistic regressions can be used instead of log-linear models.

A possible log-linear model for a two-way contingency table with categorical variables X and Y, with I rows and J columns and $n = IJ$ cells is,

$$\log \mu_{ij} = \theta + \theta_i^X + \theta_j^Y + \theta_{ij}^{XY}, \qquad i = 1, \ldots, I,\ j = 1, \ldots, J.$$

This model is known as a **saturated model**, since it is the most general and flexible model and it is containing all possible parameters to associate patterns among the variables. θ is the intercept for the model. θ_i^X represents the effect of each level of variable X (row effects) on the logarithm of cell means. θ_j^Y is representative of the effect of the variable Y levels (column effects) on the logarithm of cell means. θ_{ij}^{XY} allows for possible association between the two variables and describes its effect on the logarithm of the cell means. It is also called an **interaction parameter**. θ_i^X and θ_j^Y are the main effects and θ_{ij}^{XY} is the first order interaction parameter. The higher or lower value for each of them relatively increases or decreases the expected cell counts of the corresponding cells in the contingency table. To make the parameters mathematically estimable, there are two common types of constraints to imply on them. One type is the "sum to zero" or "effect coding" constraints which assume $\sum_{i=1}^I \theta_i^X = \sum_{j=1}^J \theta_j^Y = \sum_{i=1}^I \theta_{ij}^{XY} = \sum_{j=1}^J \theta_{ij}^{XY} = 0$. The other type is "corner point" or "dummy coding" constraints which set one level of each effect or interaction, say the first level, equal to zero to have $\theta_1^X = \theta_1^Y = \theta_{i1}^{XY} = \theta_{1j}^{XY} = 0$.

For the same contingency table assume that the two variables are independent. The joint probability then for each cell is $\pi_{ij} = \pi_{i+}\pi_{+j}$ and therefore the expected cell count is,

$$\mathrm{E}(Y_{ij}) = \mu_{ij} = N\pi_{ij} = N\pi_{i+}\pi_{+j}.$$

Taking the logarithms of both sides of this equation leads to the log-linear model,

$$\log \mu_{ij} = \theta + \theta_i^X + \theta_j^Y, \qquad i = 1, \ldots, I,\ j = 1, \ldots, J.$$

This model only includes the intercept and main effects, or row and column effects, without the parameter representing the association between two variables. This model is called the **independence model** and is a special case of the saturated model.

To define a log-linear model for any contingency table, we follow the notation of Overstall and King [2014]. Let $\mathsf{V} = \{V_1, \ldots, V_m\}$ denote a set of m categorical variables and assume the jth variable has l_j levels. The contingency table corresponding to these variables has $n = \prod_{j=1}^{m} l_j$ cells, and is referred to as a $l_1 \times \cdots \times l_j$ table. We let $\mathbf{y}$ denote a $n \times 1$ vector corresponding to the observed cell counts. Each element of this vector is specified by $y_{\mathbf{i}}$ with $\mathbf{i} = (i_1, \ldots, i_m)$ identifying the combination of variable levels that cross-classify the given cell. We define L as the set of all n cross-classifications and thus the set of all cells in the table. Mathematically, $L = \otimes_{j=1}^{m} [l_j]$, in which $[l_j] = \{0, 1, \ldots, l_j - 1\}$ (note that the variables levels start from level zero). Then by definition $|L| = n$ and $N = \sum_{\mathbf{i} \in L} y_{\mathbf{i}}$ is the total observed number of counts.

We assume the data (i.e. $y_{\mathbf{i}}$s) are observations from independent Poisson distributions with the associated probability mass function (1.1) and $\mu_{\mathbf{i}} = \mathrm{E}(Y_{\mathbf{i}})$. Let E denote a set of subsets of V. By adapting the notation of Johndrow et al. [2014], the log-linear model assumes the form,

$$m_{\mathbf{i}} = \log \mu_{\mathbf{i}} = \sum_{e \in \mathsf{E}} \theta^e(\mathbf{i}), \tag{1.5}$$

$\theta^e(\mathbf{i}) \in \mathbb{R}$ denotes the interaction among the variables in e corresponding to the levels in $\mathbf{i}$. The summation is over all members of E, which could be the set of all subsets of variables (for a saturated model) or a set of desired subsets (for a smaller model only with the desired variables). As a convention, θ corresponds to $e = \varnothing$, which guarantees that there is an intercept in the model. Each model could include parameters which specify the main effects of the variables or the interactions between them on the logarithm of the cell means. For identifiability, we choose corner point constraints, such that the lowest level (the zero level) of each main effect or interaction is set equal to zero. Generally, the log-linear models can be written with the design matrix A and p number of θ parameters, as $\mathbf{m}_{n\times 1} = \log \boldsymbol{\mu}_{n\times 1} = \mathsf{A}_{n\times p} \boldsymbol{\theta}_{p\times 1}$. Model (1.5) is referred to as a hierarchical model if for every $e \in \mathsf{E}$ that $\theta^e(\mathbf{i}) = 0$, we have $\theta^f(\mathbf{i}) = 0$ for all $f \supseteq e$ [Johndrow et al., 2014].

For a log-linear model fitted to an l^m table, another way to order cell counts or cell means in model (1.5) is setting a one to one correspondence between the set $\mathbf{i} = (i_1, ..., i_m)$ and integer numbers $i = 1, ..., l^m$, as,

$$\mathbf{i} = (i_1, ..., i_m) = i_1 l^0 + i_2 l^1 + \cdots + i_{m-1} l^{m-2} + i_m l^{m-1} + 1. \tag{1.6}$$

We will use this in examples and proofs for simplification, and order the cell counts inside the contingency tables using this notation.

Log-linear models are essential and widely used for analysing categorical data in contingency tables with applications in many scientific areas. Some examples in social, medical and biological sciences are given in Agresti [2002], Bishop et al. [1975] and McCullagh and Nelder [1989]. Statistical theory for contingency tables can be traced back to Bartlett [1935] who computed the maximum likelihood estimates (MLE) of log-linear models in $2 \times 2 \times 2$ tables and investigated the independence test of variables for them. The study of log-linear models for three-way and higher-way tables began by Birch [1963] who derived maximum likelihood estimates and sufficient statistics for log-linear models under certain hypothesises. Goodman [1970, 1971] continued the work of estimating multiplicative interactions of log-linear models and analysing marginal tables in m-way tables. Fienberg [1972] estimated the total population size of a multiple recapture census for closed population by fitting a log-linear model to an incomplete 2^k contingency table. A comprehensive study of log-linear models for contingency tables was developed by Haberman [1973].

1.3 The problem of zero observations in contingency tables

Contingency tables might contain zero cell counts. There are two main types of zero observations: structural and sampling zeros. If an observed cell count (y_i) is zero and we know that the expected value (μ_i) for that cell is zero too, then it is a **structural zero**. A positive cell count is impossible to occur for such a cell as the mean and variance of the cell count are both zero. For example, in a contingency table with two variables of sex and cancer type, there must be zero in the cells for male and ovarian cancer, or female and prostate cancer. A structural zero does not contribute to the likelihood function and that cell count and cell mean can be removed from the table and the model. The corresponding contingency table is then known as an incomplete table [Agresti, 2002], although the term is used to refer to tables with unobserved cells too [Overstall and King, 2014, Overstall, et al., 2014]. On the other hand, if we know that the expected value is not necessarily zero ($y_i = 0, \mu_i > 0$), then that zero cell count is a **sampling zero** and still contributes to the likelihood function. A sampling zero is a part of the data set, as it is a possible outcome in both Poisson and multinomial sampling distributions. The estimated value ($\hat{\mu}_i$) for a sampling zero could be either zero or non-zero. A contingency table including many sampling zero cell counts is called a sparse table.

While fitting a log-linear model to a contingency table, some observed cell counts may be zero which in turn can raise problems in estimating the parameters of the model, including slow or non convergence of the routine or large standard errors for the estimates [Fienberg and Rinaldo, 2007] . Sampling zeros are more common than structural zeros and we primarily focus on the effect of their presence on the likelihood surface and identifiability of log-linear models. This section provides the background for two approaches that address the possible problems in a log-linear model caused by zero cell counts. The matter of the existence of the MLE and the parameter redundancy approach are briefly explained here. Chapters 2 and 4 illustrate the required theorems and methods regarding each of these two approaches.

Existence of the MLE

For n independent observations from the Poisson distribution (1.1) for the log-linear model $\log(\boldsymbol{\mu}) = \mathsf{A}\boldsymbol{\theta}$, the log-likelihood function becomes,

$$l(\boldsymbol{\mu}(\boldsymbol{\theta})) = \sum_{i=1}^{n}(y_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta})) - \sum_{i=1}^{n} \log y_i!. \tag{1.7}$$

In order to fit a log-linear model to a contingency table, the parameters of model (1.5), $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$, must be estimated by maximizing the likelihood function (1.7). This likelihood function is a strictly concave function of $\boldsymbol{\mu}$. It means the second derivative of the function is always negative if we assume the cell count variables could only have positive means, i.e., $\mu_i > 0$ [Haberman, 1973]. Agresti [2002] notes that if all cell counts are positive, the MLE of log-linear model parameters exists.

For the generalized linear models in the form of (1.4) the estimates vector obtained by the Iterative Re-weighted Least Squares (IWLS) method is,

$$\hat{\boldsymbol{\theta}}^{(t+1)} = (\mathsf{A}^{\mathsf{T}}\mathsf{W}^{(t)}\mathsf{A})^{-1}\mathsf{A}^{\mathsf{T}}\mathsf{W}^{(t)}\mathbf{z}^{(t)}, \tag{1.8}$$

such that $\mathsf{W}^{(t)} = \text{diag}\left(Var(Y_i)g^{\mathsf{T}}(\mu_i^{(t)})^2\right)^{-1}$, $z_i^{(t)} = g(\mu_i^{(t)}) + (y_i - \mu_i^{(t)})g^{\mathsf{T}}(\mu_i^{(t)})$ and the superscript T shows the transpose of a vector or a matrix. For the log-linear model, the initial values are usually $\mu_i^{(0)} = y_i \neq 0$ to make the initial value of $g(\mu_i)$ and z_i finite. So as long as the cell counts are not zero, the parameter estimates exist. In case of having zero observations, usually a small value is added to the cell counts [Agresti, 2002]. When a model is over-parametrized, the initial W matrix and therefore $\mathsf{A}^{\mathsf{T}}\mathsf{W}^{(t)}\mathsf{A}$ are singular and more than one solution for $\hat{\boldsymbol{\theta}}$ is possible. Obtaining a point estimate for

all parameters is possible even when the matrix is singular but the over-parametrization will be shown by big or zero asymptotic standard errors [Brown and Fuchs, 1983].

Haberman [1973] proves maximum likelihood estimates for log-linear models are unique when they exist and provides a necessary and sufficient condition for the existence of the MLE of cell means ($\boldsymbol{\mu}$) in a theorem given in Section 4.1.1. For some patterns of zero cell counts in the contingency table, the MLE of the log-linear model parameters might not exist. Assuming only positive values for the cell means, the theorem determines whether the MLE for all cell means exists or not for any pattern of zeros in the table. Haberman [1974] states the fact that the parameter of a Poisson distribution could take the value of zero too and this was a motivation to define the extended MLE of cell means, when $\hat{\mu}_i = 0$ and there are infinite estimates for some log-linear model parameters (θs). The word "extended" indicates extending the range of μ_i in (1.1) from positive values to non-negative values. The interest was increased in studying the effect of sampling zeros in contingency tables. Brown and Fuchs [1983] investigate this effect by considering and comparing some iterative methods.

The effect of the presence of zero observations in deriving maximum likelihood estimates for parameters of the log-linear models has been studied in a polyhedral and graphical model framework based on works by Lauritzen [1996]. Eriksson et al. [2006] provide a polyhedral version of the Haberman's necessary and sufficient condition for the existence of the MLE. Fienberg and Rinaldo [2012a] study estimability of parameters under non-existent MLE with extended exponential families and under different sampling schemes. Their work is developed to higher dimensional problems by Wang et al. [2016]. The existence of the MLE approach will be discussed in detail in Chapter 4.

Parameter redundancy

A model is parameter redundant when statistical methods fail to estimate all of its parameters. In some cases, the reason is over-parametrization in the model, whilst sometimes lack of data causes this failure. The concept of parameter redundancy is related to the identifiability of a model. A model is not identifiable if two different sets of parameter values generate the same model for the data and this often happens when a model is over-parametrized [Silvey, 1975]. If the model could be rearranged as a function of a smaller set of parameters, which themselves are a function of the initial parameters, then the model is parameter redundant. Thus, a parameter redundant model could be reduced to a smaller but identifiable model with all estimable parameters. Non-parameter redundant models are referred to as full rank models.

Catchpole and Morgan [1997] define parameter redundant models and provide a symbolic method to detect them. They prove that these models have a flat ridge in their likelihood function which makes the unique existence of some MLEs impossible. Catchpole et al. [1998] describe the method to find the estimable parameters or the estimable combinations of parameters in parameter redundant models. Examples of models which are parameter redundant due to their structure are provided in capture-recapture and mark-recovery areas in Catchpole and Morgan [2001]. They also mention that parameter redundancy might occur because of zero observations, as it may happen in the contingency tables. Their work was developed by Cole et al. [2010], by using exhaustive summaries in parameter redundant biological examples. Choquet and Cole [2012] develop a symbolic-numeric method for detecting parameter redundancy and for obtaining the estimable model parameters for the cases where the symbolic method is not easily applicable and provide examples in capture-recapture and compartment models.

Log-linear models are an example of models that can become parameter redundant as a result of lack of data or zero cell count observations. Both sampling and structural zeros can lead to parameter redundancy but we focus on the presence of sampling zeros. In the presence of structural zeros, the corresponding cell means are removed from the model and the resulting model can then be checked for possible parameter redundancy. In Chapter 2, we adapt the parameter redundancy approach to fit log-linear models to contingency tables data including zero observations.

1.4 Thesis aim and structure

The main objective of this thesis is to study parameter redundancy in log-linear models fitted to contingency tables with some sampling zero observations. We aim to examine if a log-linear model for a given table is parameter redundant or not and to detect which parameters or functions of parameters are estimable in the case of parameter redundancy. This enables us to address how to reduce a parameter redundant model to a smaller identifiable one, remove parameters that are not estimable, and fit a model that is identifiable. In contrast to the other approach, we focus on the estimability of the parameters of the log-linear model ($\boldsymbol{\theta}$) rather than the cell means ($\boldsymbol{\mu}$), although it is possible to obtain one's values from knowing the other one. The log-linear model parameters are of particular interest in investigation on how variables interact and relate to each other. This is revealed by the presence or absence of the interaction terms in the model and their sign and magnitude. We describe the alternative approach which

discusses the existence of the MLE, and compare this method and its results to the parameter redundancy approach.

In Chapter 2 parameter redundancy and the method to detect it in a model are described. We explain how to adapt the existing methodology to Poisson log-linear models, in terms of detecting parameter redundancy and reducing the model to a smaller model composed of only the estimable parameters. We illustrate the method by considering 2×2 contingency tables including one or more zero cell counts and address saturated, non-saturated, non-hierarchical models and Poisson and multinomial sampling distributions. It is explained that for some parameter redundant log-linear models, the unique MLE could still be calculated for all of the parameters after implying some additional constraints.

In chapter 3, we provide general theorems about saturated Poisson log-linear models fitted to a contingency table with only one zero cell. We indicate exactly which model parameters become inestimable after a zero count is observed in a specific cell. The theorems are proved for saturated log-linear models corresponding to 2^m, 3^m and l^m contingency tables.

Chapter 4 explains the approach on the existence of the MLE for log-linear models fitted to a table containing zero cells. The method based on the graphical and polyhedral framework is described in detail. This technique determines whether the MLE exists for a pattern of zero and if it does not, it further determines which cell means are estimable. The results from this approach are compared to the ones based on the proposed parameter redundancy approach, illustrated by examples. We investigate the similarities and explain the differences between the results.

Applications from a variety of scientific fields are presented in Chapter 5. The data in the first illustration is from Silverman [2014] on "an exploratory analysis of the scale of modern slavery in the UK, using the statistical technique of multiple systems estimation". The contingency table in this study is made of five variables with two levels for each and 2744 observations in total. The table contains several sampling zeros. We fit a log-linear model to the table after finding the estimable parameters and estimate the total number of the potential victims which is the aim of this study. The data set for the second example is from Brown and Fuchs [1983]. The sparse contingency table in this case, with five variables and two levels for each, presents a symptoms study of 118 patients after the same ear surgery. We fit the specified desirable model to the data, reduce it to an identifiable model and estimate all estimable model parameters. The data of the third example is from Papathomas et al. [2012], which marks 50 important SNPs in a genome-wide association study of lung cancer. We choose five of those SNPs with

three levels for each that make a sparse table of 4260 subjects. A log-linear model is fitted to the sparse table and the estimable parameters and their estimates are obtained.

We conclude in Chapter 6 with a general discussion containing conclusion, a description of the computational aspects, and mentioning the future work.

Chapter 2

Parameter redundancy in log-linear models

2.1 Introduction

The parameter redundancy concept and the approach to detect and apply parameter redundant models are described in Section 2.2 of this chapter. In Section 2.3, we adapt the parameter redundancy method to Poisson log-linear models to realise how zero entries can change the model and affect the parameter estimation process. Examples of log-linear models fitted to contingency tables containing zero cell counts and details about the model specifications are given in Sections 2.4, 2.5 and 2.6. In Section 2.7, we specify a special case of parameter redundant log-linear models and explain that they can be recognised as non-redundant after considering some extra constraints for the model.

2.2 Parameter redundancy

An identifiable model is defined by Silvey [1975] as a model in which two different sets of parameters never give the same probability distribution for the data. Assume $M(\boldsymbol{\theta})$ is the function that specifies a statistical model containing parameters $\boldsymbol{\theta} \in \Omega$. Then there are two types of identifiability:

Definition 2.1. *A model is globally identifiable if $M(\boldsymbol{\theta}_1) = M(\boldsymbol{\theta}_2)$ implies that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. A model is locally identifiable if there exists an open neighbourhood of any $\boldsymbol{\theta}$ such that this is true. Otherwise a model is non-identifiable.* [Cole et al., 2010, Definition 1]

Identifiability is closely related to parameter redundancy. Catchpole and Morgan [1997] state that the most obvious cause for non-identifiability is the model's over-parametrisation, which is referred to as parameter redundancy. A model with p parameters is parameter redundant if all of its parameters are not estimable, so the model can be rewritten in terms of q estimable parameters such that $q < p$. We summarize the approach introduced by Catchpole and Morgan [1997] and Catchpole et al. [1998] to identify a parameter redundant model and to find the smaller set of estimable parameters. Consider independent observations $\mathbf{y}^\mathsf{T} = (y_1, ..., y_n)$ from a member of the exponential family of distributions (1.3). The cell mean vector $\boldsymbol{\mu}$ is expressible as a function of parameters $\boldsymbol{\theta}^\mathsf{T} = (\theta_1, ..., \theta_p) \in \Omega$. The derivative matrix, $D(\boldsymbol{\theta})$, describes the relation between $\boldsymbol{\mu}$ (or a monotonic function of it) and $\boldsymbol{\theta}$,

$$D_{si}(\boldsymbol{\theta}) = \frac{\partial \mu_i}{\partial \theta_s}, \qquad i = 1, \ldots, n, \quad s = 1, \ldots, p. \tag{2.1}$$

This matrix is said to be **symbolically rank deficient**, if and only if there exists a non-zero vector, $\boldsymbol{\alpha}(\boldsymbol{\theta})$, such that for all $\boldsymbol{\theta}$,

$$\boldsymbol{\alpha}(\boldsymbol{\theta})^\mathsf{T} D(\boldsymbol{\theta}) = \mathbf{0}. \tag{2.2}$$

So $\boldsymbol{\alpha}(\boldsymbol{\theta})$ is the null space of the transpose of the derivative matrix, i.e., $D^\mathsf{T}(\boldsymbol{\theta})\boldsymbol{\alpha}(\boldsymbol{\theta}) = \mathbf{0}$.

Theorem 2.1. *A model is parameter redundant if and only if its derivative matrix is symbolically rank-deficient.* [Catchpole and Morgan, 1997, Theorem 1]

This method has a long history of use as a general way to detect identifiability including Goodman [1974], Shapiro [1986], Thowsen [1978], Pohjanpalo [1982], Delforge [1989] and Chappell and Gunn [1998]. After detecting parameter redundancy in a model and realising that all p model parameters are not estimable, we specify how many estimable parameters or estimable combinations of parameters exist. The rank of the derivative matrix (r) determines this number. Then $d = r - p$ is called **model deficiency** which indicates the number of all possible $\boldsymbol{\alpha}(\boldsymbol{\theta})$ vectors. Zero elements in $\boldsymbol{\alpha}(\boldsymbol{\theta})$ correspond to those parameters of the model that are directly estimable. In order to find other possible estimable combinations of the parameters, the next theorem is used. The method used in this theorem was also developed independently for compartment models by Chappell and Gun [1998] and Evans and Chappell [2000].

Theorem 2.2. *A minimal parameter set, containing $p - d$ parameters, can be found by solving the auxiliary equations of a system of linear first order partial differential*

equations,

$$\sum_{s=1}^{p} \alpha_{sj} \frac{\partial f}{\partial \theta_s} = 0, \qquad j = 1, ..., d. \tag{2.3}$$

[Catchpole et al., 1998, Theorem 1]

A model which is not parameter redundant is called a **full rank model**. In that case, the rank of the derivative matrix equals to the number of parameters and the model deficiency is zero. If the rank of $D(\boldsymbol{\theta})$ equals to p for all $\boldsymbol{\theta} \in \Omega$, the model is essentially full rank and if this is true for some but not all $\boldsymbol{\theta}$, then the model is conditionally full rank. Cole et al. [2010] provide a theorem which determines if the model is essentially or conditionally full rank by taking a decomposition of the derivative matrix.

Theorem 2.3. *For a full rank model, write $D = PLUR$, where P is a permutation matrix, L is a lower triangular matrix with ones on the diagonal, U is an upper triangular matrix and R is a matrix in reduced echelon form. The model is parameter redundant at θ if and only if $Det(U) = 0$ at a point $\theta \in \Omega$ and R is defined at θ.* [Cole et al., 2010, Theorem 4]

Catchpole and Morgan [1997] clarify the relationship between identifiability and parameter redundancy:

- "If a model is parameter redundant, then it is not locally identifiable."
- If a model is essentially full rank, then it is (at least) locally identifiable.
- "It is certainly not true that full rank models are necessarily identifiable, or even locally identifiable" (the models might be conditionally full rank).
- "It is an open question whether or not essentially full rank models must be identifiable (the answer is positive for a particular simple class of models)."

In a parameter redundant model, after detecting estimable parameters and estimable combinations of them, the initial model will be reduced to a model only including estimable parameters and combinations. The reduced model is full rank and maximizing its likelihood function results in obtaining MLE of the parameters. The shape of the likelihood surface in a parameter redundant model is described in the next theorem.

Theorem 2.4. *If a model is parameter redundant, then for any data set, the likelihood surface has a completely flat ridge.* [Catchpole and Morgan, 1997, Theorem 2]

The flat ridge usually makes it impossible to find a unique maximum likelihood estimate for some parameters. However, in some cases the flat ridge might be orthogonal to

some parameter axes, so those parameters still have unique estimates which maximise the likelihood function [Catchpole et al., 1998].

Rothenberg [1971] employed the information matrix instead of the derivative matrix D and proved that for probability function f, if the information matrix with elements $I_{ij} = E\left[\frac{\partial \log f}{\partial \theta_i} . \frac{\partial \log f}{\partial \theta_j}\right]$ is non-singular then the model is locally identifiable. If the matrix in non-singular and f is a member of the exponential family of distributions, then the model is globally identifiable. Catchpole and Morgan [1997] used the derivative matrix and showed that its rank is the same as the rank of the information matrix and noted that although the information matrix is smaller, it is algebraically more difficult to handle.

In defining the model, $M(\boldsymbol{\theta})$ could be the probability function from the exponential family of distributions, or terms of a log-likelihood function or any other functions that represent the model. In some examples, a vector of parameter combinations that uniquely defines the model could be used to make the symbolic computations easier. This parameter vector is called an **exhaustive summary** and is shown as $\boldsymbol{\kappa}(\boldsymbol{\theta})$. In the general definition of the derivative matrix (2.1), we have $\boldsymbol{\kappa}(\boldsymbol{\theta}) = \boldsymbol{\mu}$ [Cole et al., 2010]. Another option which simplifies the symbolic computations and is helpful for large and complicated derivative matrices, is to use the extension theorem [Catchpole and Morgan, 1997, Cole et al., 2010]. Assume $\boldsymbol{\kappa}_1(\boldsymbol{\theta}_1)$ is the exhaustive summary used to make the derivative matrix $D_1(\boldsymbol{\theta}_1) = \left[\frac{\partial \boldsymbol{\kappa}_1(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta}_1}\right]$. If the model is extended by adding a new set of parameters $\boldsymbol{\theta}_2$, the exhaustive summary is also extended to $\boldsymbol{\kappa}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = [\boldsymbol{\kappa}_1(\boldsymbol{\theta}_1), \boldsymbol{\kappa}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)]$. Then the derivative matrix for the extended model is,

$$D = \begin{bmatrix} D_1(\boldsymbol{\theta}_1) & D_2(\boldsymbol{\theta}_1) \\ 0 & D_2(\boldsymbol{\theta}_2) \end{bmatrix},$$

such that $D_2(\boldsymbol{\theta}_1) = \frac{\partial \boldsymbol{\kappa}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_1}$ and $D_2(\boldsymbol{\theta}_2) = \frac{\partial \boldsymbol{\kappa}_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\partial \boldsymbol{\theta}_2}$ [Cole et al., 2010]. The following theorem explains how to detect whether the general D matrix is full rank.

Theorem 2.5. *If the original model is full rank (i.e. $D_1(\boldsymbol{\theta}_1)$ is full rank) and $D_2(\boldsymbol{\theta}_2)$ is full rank, then the extended model is full rank also.* [Cole et al., 2010]

Thus, finding the rank of D is not necessary and calculating the rank of smaller and simpler derivative matrices corresponding to smaller models are enough.

2.3 Adaptation to log-linear models

Parameter redundancy is usually a characteristic of the model's structure and does not depend on data. Examples of such parameter redundant models from biology, compartment models and ring-recovery data are given in Catchpole and Morgan [1997], Catchpole et al. [1998] and Cole et al. [2010]. In practice, a full rank model may turn into a parameter redundant one because of missing data or zero observations. For example, Cole et al. [2010] describe such a model for capture-recapture data and mention the interest to discover how much data could be missed before the model becomes parameter redundant. This type of parameter redundancy is sometimes referred to as "extrinsic" parameter redundancy [Gimenez et al., 2004]. To detect the effect of a zero observation on identifiability of a model, exhaustive summaries or monotonic functions of the variables expectations containing the observations could be used in forming the derivative matrix [Cole et al., 2010]. Cole et al. [2012] use log-likelihood function elements as exhaustive summaries to detect extrinsic parameter redundancy for a mark-recovery model with some zero observations.

The log-likelihood function for independent Poisson observations in (1.7) is a strictly concave function of cell means for positive $\boldsymbol{\mu}$. It is known that in a log-linear model for a contingency table with all positive $y_{\mathbf{i}}$, maximum likelihood estimates exist for all the model parameters [Haberman, 1973]. Nonetheless, in chapter 3 we will prove that such a model is full rank and all of its parameters are estimable. For a log-linear model, zero cell count observations may rise identifiability problem [Catchpole and Morgan, 2001]. We want to detect the effect of one or more zero cells in the contingency table on estimability of the model parameters and indicate which parameters of the model or which combinations of parameters are estimable.

To adapt the method described in Section 2.2 to detect the parameter redundancy of a log-linear model as (1.5), the derivative matrix (2.1) must be adjusted first. In forming the derivative matrix, some monotonic transformation of the cell means could be used [Catchpole and Morgan, 1997]. Thus, instead of taking derivatives of cell means we choose to use the monotonic function of them $y_i \log \mu_i$, such that,

$$D_{si} = \frac{\partial y_i \log \mu_i}{\partial \theta_s}, \qquad i = 1, \ldots, n, \quad s = 1, \ldots, p. \tag{2.4}$$

These matrix elements are not functions of $\boldsymbol{\theta}$ any more, as $\log \mu_i$ is a first order linear function of $\boldsymbol{\theta}$ in a log-linear model. This also applies to the corresponding vector of $\boldsymbol{\alpha}(\boldsymbol{\theta})$ in (2.2) which could be shown as $\boldsymbol{\alpha}$ for a log-linear model.

Including cell counts in the monotonic function of the cell means in equation (2.4) allows us to investigate the effect of zero observations in the identifiability of the model. Each sampling zero cell count turns a column of the matrix to zero and may decrease the rank of the derivative matrix and the number of estimable parameters. Thus, the model is parameter redundant only for a given data set not for any data set. The log-likelihood function elements are usually another option to consider in forming the derivative matrix as exhaustive summaries including the data [Cole et al., 2010]. However, for the Poisson log-linear model those elements are $y_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta})$ as shown in (1.7), in which setting $y_i = 0$ would not decrease the rank of the derivative matrix. Furthermore, these elements are not monotonic functions of cell means or model parameters and do not uniquely define them. Catchpole and Morgan [1997, 2001] form the derivative matrix for contingency table data from a multinomial log-linear model and denote the effect of missing data on the redundancy of the model, which will be mentioned in Section 2.5.2.

Using information matrix instead of a derivative matrix is an alternative for detecting parameter redundancy, as mentioned before in Section 2.2. However, by forming the information matrix with the Poisson log-likelihood function elements which are $y_i \log \mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta})$, the presence of zero entries does not reduce the rank of the matrix. If we consider $y_i \log \mu_i(\boldsymbol{\theta})$ elements to form a Hessian matrix, the second derivatives of the elements with respect to the model parameters are zero. So in this extrinsic parameter redundant model, information matrix suffers from the same problem as the standard derivative matrix constructed with log-likelihood function elements. Potential use of the informations matrix and it's correspondence to the derivative matrix for detecting parameter redundancy in log-linear models with some zero observations needs further investigation.

A log-linear model is full rank if the rank of the derivative matrix is not smaller than the number of model parameters or p, otherwise, the model is parameter redundant. When the model is full rank, it is always essentially full rank for the whole range of parameters since the derivative matrix does not include parameters. For a square and full rank derivative matrix (which can be formed as upper triangular), the PLUR decomposition in Theorem 2.3 provides the upper triangular matrix U equal to the derivative matrix. Thus, when the determinant of U is zero it does not depend on the parameter values and the model is essentially full rank. After calculating $\boldsymbol{\alpha}$ in (2.2) and solving the partial differential equations in (2.3), finding all estimable parameters ($\boldsymbol{\theta}$) and estimable combinations of parameters declares which cell means ($\boldsymbol{\mu}$) are estimable. Some cell means with corresponding zero observations might not be expressible in terms of the estimable parameters, so they are referred to as inestimable cell means. The estimate for these cell means could be considered as zero, then we treat them as

structural zeros and remove them from the model. Nonetheless, some cells with zero entries might have positive estimable cell means. Obtaining the vector of the estimable parameters and combinations of them $(\boldsymbol{\theta}')$ and the vector of estimable cell means $(\boldsymbol{\mu}')$ lead to a new and smaller design matrix (A'). The reduced model is built by using these vectors and matrix as $\log \boldsymbol{\mu}' = \mathsf{A}'\boldsymbol{\theta}'$.

After finding the estimable set of parameters and reducing the model to a smaller model with rank r, the degree of freedom for the new model is the number of usable data (i.e. observations with corresponding estimable cell means) minus r (i.e. the number of estimable parameters).

Discovering the set of estimable parameters in the log-linear model makes the process of model selection easier, since checking all of the model parameters is not required and one can consider only the smaller set of estimable parameters in the search for the best model. In other words, only the set of non-redundant models is searched to obtain the best fit. We use an example to illustrate this process in Chapter 5.

In the next sections of this chapter, examples of fitting log-linear models to relatively small contingency tables will be investigated. We examine the effect of zero observations on estimates of the model parameters and indicate the maximum number of zeros a table could contain before the parameter redundancy issue rises.

2.4 Examples in 2^2 contingency tables

We begin fitting log-linear models to contingency tables containing zero observations by considering a small 2×2 table. A saturated model will be fitted to the table data to observe the maximum number of zero cells that the table could contain before parameter redundancy occurs. We consider a numerical example first to see what parameter estimates are provided by a standard statistical software like `R` for a parameter redundant model. Then the parameter redundancy in the model is studied in a symbolic and general way.

2.4.1 Numerical example for a saturated log-linear model

Assume two categorical variables X and Y, with two levels 0 and 1 for each variable, and observations vector of $\mathbf{y}^\mathsf{T} = (y_1, y_2, y_3, y_4)$ as shown in Table 2.1. The cell counts are ordered according to equation (1.6).

The aim is to fit a saturated model to the table data. According to equation (1.5), we

X	Y	
	0	1
0	y_1	y_3
1	y_2	y_4

Table 2.1 A 2^2 contingency table.

have $\mathsf{V} = \{X,Y\}, m = 2$ and $l_1 = l_2 = 2$, so the saturated log-linear model is,

$$\begin{aligned} m_1 &= \log \mu_1 = \log \mu_{00} = \theta, \\ m_2 &= \log \mu_2 = \log \mu_{10} = \theta + \theta^X, \\ m_3 &= \log \mu_3 = \log \mu_{01} = \theta + \theta^Y, \\ m_4 &= \log \mu_4 = \log \mu_{11} = \theta + \theta^X + \theta^Y + \theta^{XY}, \end{aligned} \tag{2.5}$$

and the vector of parameters is $\boldsymbol{\theta}^\mathsf{T} = (\theta, \theta^X, \theta^Y, \theta^{XY})$ with $p = 4$. There are two levels for each variable, so for example, we have θ_0^X and θ_1^X. Due to the use of corner point constraints, θ_0^X is set equal to zero and as there remains only one corresponding parameter θ_1^X, it is denoted by θ^X. The model in matrix form $\log \boldsymbol{\mu} = \mathsf{A}\boldsymbol{\theta}$ is:

$$\begin{bmatrix} \log \mu_{00} \\ \log \mu_{01} \\ \log \mu_{10} \\ \log \mu_{11} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \theta \\ \theta^X \\ \theta^Y \\ \theta^{XY} \end{bmatrix}.$$

As a numerical example, we adopt Table 2.2 from Agresti [2002] which refers to a 1992 survey by the Wright State University School of Medicine and the United Health Services in Dayton, Ohio. The survey asked 2276 students in their final year of high school in a non-urban area whether they had ever used alcohol, cigarettes, or marijuana. This 2×2 table is only one part of the original table and it represents cigarette and marijuana use for those students who have ever used alcohol.

Cigarette (X)	Marijuana (Y)	
	No (0)	Yes (1)
No (0)	456	44
Yes (1)	538	911

Table 2.2 Cigarette and marijuana use for high school seniors.

We make use of the `glm` function in `R` to fit model (2.5) to this table's data. `R` code is provided in Appendix A. The output including parameter's estimates and standard errors is,

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  6.12249    0.04683 130.741  < 2e-16 ***
A2  0.16537    0.06365   2.598  0.00938 **
A3 -2.33830    0.15786 -14.812  < 2e-16 ***
A4  2.86499    0.16696  17.159  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

So $\hat{\boldsymbol{\theta}}^{\mathsf{T}} = (6.12, 0.16, -2.33, 2.86)$, and subsequently $\hat{\boldsymbol{\mu}}^{\mathsf{T}} = (456, 538, 44, 911)$ since the model is saturated.

Now one cell at a time is set equal to zero to observe how parameter estimates are changed.

Case I. If $y_4 = y_{11} = 0$, the estimates and standard errors are given as:

```
Coefficients:
     Estimate Std. Error z value Pr(>|z|)
A1  6.122e+00  4.683e-02 130.741  < 2e-16 ***
A2  1.654e-01  6.365e-02   2.598  0.00938 **
A3 -2.338e+00  1.579e-01 -14.812  < 2e-16 ***
A4 -2.625e+01  4.225e+04  -0.001  0.99950
```

The standard error for the fourth parameter or θ^{XY} is a large number compared to the estimates and the standard errors for the other parameter estimates, which makes the corresponding parameter estimate unreliable.

Case II. If $y_3 = y_{01} = 0$, the estimates and standard errors are given as:

```
Coefficients:
     Estimate Std. Error z value Pr(>|z|)
A1  6.122e+00  4.683e-02 130.741  < 2e-16 ***
A2  1.654e-01  6.365e-02   2.598  0.00938 **
A3 -2.843e+01  4.225e+04  -0.001  0.99946
A4  2.895e+01  4.225e+04   0.001  0.99945
```

The standard errors for the third and the fourth parameters or θ^{Y} and θ^{XY} are equal and large, so the corresponding parameter estimates are questionable.

Case III. If $y_2 = y_{10} = 0$, the estimates and standard errors are given as:

```
Coefficients:
     Estimate Std. Error z value Pr(>|z|)
A1  6.122e+00  4.683e-02 130.741   <2e-16 ***
A2 -2.843e+01  4.225e+04  -0.001    0.999
A3 -2.338e+00  1.579e-01 -14.812   <2e-16 ***
A4  3.146e+01  4.225e+04   0.001    0.999
```

The standard errors for the second and the fourth parameters or θ^X and θ^{XY} are equal and large, so the corresponding parameter estimates are questionable.

Case IV. And finally if $y_1 = y_{00} = 0$, the estimates and standard errors are given as,

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1   -22.30   42247.17  -0.001    1.000
A2    28.59   42247.17   0.001    0.999
A3    26.09   42247.17   0.001    1.000
A4   -25.56   42247.17  -0.001    1.000
```

The standard errors for all parameters are the same and large, so the parameter estimates are not reliable.

In each case, we lose some information by having a zero cell count. The number of parameters which should be estimated is four but the number of non-zero cell counts is three. So `R` cannot estimate all parameters properly and, for some of them, reports numbers with large standard errors instead. It can suggest that these parameters are not directly estimable. It seems that only one zero cell count is enough to make this saturated log-linear model non-identifiable. To get a more clear sense of parameters behaviour in the presence of one zero observation, we utilise `WinBUGS` to estimate parameters one more time.

`WinBUGS` is a computer package for implementing Bayesian analyses and only the likelihood model and prior distributions need to be specified. The Markov chain Monte Carlo (MCMC) iterations are then performed and posterior summary statistics are produced in addition to some other useful plots and statistics. The purpose of applying the Bayesian approach here is to see if the existence of some prior information on the parameters could turn an inestimable parameter to an estimable one. Although the concept of identifiability in the Bayesian framework is different with the one in classical statistics [Almond, et al., 2015, Rao and Dey, 2005], we do not discuss these in any detail here. Our aim is only to obtain parameter estimates with this alternative

method rather than the maximum likelihood estimation. The corresponding `WinBUGS` code is given in Appendix A. For our example, the prior distribution for $\boldsymbol{\theta}$ parameters is assumed as a flat Normal distribution with zero mean and a large variance, i.e. $N(0, 10000)$ which is an uninformative prior. Three chains are used in the process and the initial values for each parameter in each chain are 0, 10 and -10. The code is run for 100000 iterations and we eliminate the first 30000 iterations as burn-in. Corresponding trace plots and marginal density plots are provided in Appendix B. The trace plots show how iterations converge to the estimated values for three implied chains. The density plots demonstrate the marginal posterior density of the parameters.

We consider Table 2.2 without any zero cells first. In Figures B.1 and B.2, the trace plots indicate an acceptable convergence to the estimates and marginal density plots show normal posterior distributions for the parameters and they are unimodal with modes at the estimates. The parameters estimates and some other statistics are given below. The posterior means and posterior standard deviations are almost the same as estimates and standard errors achieved by `R`.

node	mean	sd	MC error	2.5%	median	97.5%
theta[1]	6.121	0.0467	3.314E-4	6.029	6.122	6.212
theta[2]	0.165	0.0636	4.847E-4	0.0407	0.1654	0.2897
theta[3]	-2.347	0.1583	0.002225	-2.664	-2.343	-2.047
theta[4]	2.874	0.1673	0.002367	2.555	2.871	3.209

Now one cell at a time is set to be equal to zero and we use `WinBUGS` to check the estimates and the convergence behaviour of parameters in the saturated model.

Case I. If $y_4 = y_{11} = 0$, the estimates and standard errors are given below. The estimated posterior mean for parameter θ^{XY} is a number with a big posterior standard deviation compared to the standard deviations for the other estimates. In Figures B.3 and B.4, the trace plot for this parameter does not converge around a specific value and the location of mode in the marginal density plot of the parameter is not exact.

node	mean	sd	MC error	2.5%	median	97.5%
theta[1]	6.121	0.04693	2.082E-4	6.028	6.122	6.213
theta[2]	0.1657	0.06375	2.749E-4	0.04105	0.1655	0.2906
theta[3]	-2.349	0.1588	4.327E-4	-2.671	-2.346	-2.048
theta[4]	-82.77	59.42	0.1374	-225.5	-70.5	-7.541

Case II. If $y_3 = y_{01} = 0$, the estimates and standard errors are given below. The estimates for parameters θ^{Y} and θ^{XY} are numbers with relatively big standard

deviations. In Figures B.5 and B.6, the three chains in the trace plots for these parameters do not converge around specific values and the marginal density plots are not unimodal.

node	mean	sd	MC error	2.5%	median	97.5%
theta[1]	6.121	0.04653	2.643E-4	6.029	6.122	6.211
theta[2]	0.1656	0.06347	4.014E-4	0.0415	0.1653	0.29
theta[3]	-12.74	4.322	0.1532	-23.56	-11.78	-6.573
theta[4]	13.26	4.322	0.1532	7.1	12.31	24.08

Case III. If $y_2 = y_{10} = 0$, the estimates and standard errors are given below. The estimates for parameters θ^X and θ^{XY} are numbers with relatively big standard deviations. In Figures B.7 and B.8, the three chains in the trace plots for these parameters do not converge around specific values and the marginal density plots are not unimodal .

node	mean	sd	MC error	2.5%	median	97.5%
theta[1]	6.122	0.04693	2.979E-4	6.028	6.122	6.213
theta[2]	-11.56	4.687	0.1662	-20.94	-9.599	-5.914
theta[3]	-2.352	0.1583	0.002211	-2.669	-2.35	-2.048
theta[4]	14.6	4.689	0.1662	8.949	12.65	23.96

Case IV. If $y_1 = y_{00} = 0$, the estimates and standard errors are given below. The estimates for all parameters are numbers with relatively big standard deviations. In Figures B.9 and B.10, the three chains in the trace plots for these parameters do not converge around specific values and the marginal density plots are not unimodal .

node	mean	sd	MC error	2.5%	median	97.5%
theta[1]	-4.373	2.615	0.09263	-9.081	-4.407	-0.01017
theta[2]	10.66	2.615	0.09262	6.298	10.7	15.36
theta[3]	8.144	2.615	0.09259	3.775	8.198	12.85
theta[4]	-7.617	2.615	0.09259	-12.33	-7.671	-3.245

These results from `WinBUGS` are consistent with those derived by `R` for parameters which do not have large standard estimation errors. If we make the prior distribution informative by decreasing the variance of the Normal distribution, the standard deviation for those parameters which had large standard deviations will be decreased too. So, with a precise informative prior for inestimable parameters, and considering the fact

that there is no corresponding data to affect that prior, then the posterior estimate is reliable. However, adequate Bayesian learning is of concern here [Lee, 2011]. Brooks et al. [2000] investigate the effect of improving the prior distributions in estimating the parameters in recovery and recapture models. They show that Bayesian estimates could be quite precise when unique maximum likelihood estimations do not exist but the location of the flat ridge in the likelihood function is known.

The next section applies the method described in Section 2.3 to symbolically identify the inestimable parameters in the assumed model in the presence of zero cell counts. We also detect the estimable parameters and the estimable combinations of parameters afterwards.

2.4.2 Symbolic method for a saturated log-linear model

Consider $\mathbf{M}(\boldsymbol{\theta})$, the function that specifies model (2.5), as,

$$\begin{aligned}
M_1(\boldsymbol{\theta}) &= \log \mu_{00} = \theta, \\
M_2(\boldsymbol{\theta}) &= \log \mu_{10} = \theta + \theta^X, \\
M_3(\boldsymbol{\theta}) &= \log \mu_{01} = \theta + \theta^Y, \\
M_4(\boldsymbol{\theta}) &= \log \mu_{11} = \theta + \theta^X + \theta^Y + \theta^{XY}.
\end{aligned}$$

According to Definition 2.1, this model is globally identifiable because two different set of parameters cannot produce the same model and it is true for all values of $\boldsymbol{\theta}$.

The derivative matrix (2.4) for model (2.5) is,

$$D = \left[\frac{\partial y_i \log \mu_i}{\partial \theta_s}\right] = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ 0 & y_2 & 0 & y_4 \\ 0 & 0 & y_3 & y_4 \\ 0 & 0 & 0 & y_4 \end{bmatrix}, \qquad i = 1,2,3,4, \quad s = 1,2,3,4. \tag{2.6}$$

If none of the cell counts are zero, then this matrix is full rank with rank $r = 4$ and model deficiency $d = 0$, so there does not exist an $\boldsymbol{\alpha}$ such as defined in (2.2). This means all model parameters, $\boldsymbol{\theta}$, and subsequently all cell means, $\boldsymbol{\mu}$, are estimable.

To investigate the effect of zero cells, we assume one cell count is zero at a time and then check the parameter redundancy for the model.

Case I. If $y_4 = y_{11} = 0$, the fourth column in the derivative matrix turns to zero and the rank of the matrix decreases to $r = 3$ which is smaller than the number of the model parameters. The model deficiency $d = p - r = 4 - 3 = 1$ indicates there exist one $\boldsymbol{\alpha}$ as defined in (2.2), which is $\boldsymbol{\alpha}^\mathsf{T} = (\alpha_{11}, \alpha_{21}, \alpha_{31}, \alpha_{41}) =$

$(0,0,0,\alpha_{41})$. Note that the last element of the vector could be any non-zero value or any function of the parameters, for simplification purpose we consider it as $\boldsymbol{\alpha}^{\mathsf{T}} = (0,0,0,1)$. The three zero elements represent the three estimable parameters. Thus, the smaller set of the estimable parameters is $\boldsymbol{\theta}'^{T} = (\theta, \theta^X, \theta^Y)$. This denotes that in model (2.5), only the first three cell means are estimable and $\log \mu_{11}$ is not estimable because it is not defined by a combination of parameters in $\boldsymbol{\theta}'$. $y_{11} = 0$ is treated as a structural zero now and is removed from the model. The reduced full rank model with rank 3 and degree of freedom of $d.f = 3-3 = 0$ is,

$$\begin{aligned} m_1 &= \log \mu_1 = \log \mu_{00} = \theta, \\ m_2 &= \log \mu_2 = \log \mu_{10} = \theta + \theta^X, \\ m_3 &= \log \mu_3 = \log \mu_{01} = \theta + \theta^Y. \end{aligned}$$

We fit this reduced model to the data in Table 2.2, assuming $y_{11} = 0$. The data vector for the reduced model is $\mathbf{y}^{\mathsf{T}} = (456, 538, 44)$ and the parameter estimates corresponding to $\boldsymbol{\theta}'$ are,

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  6.12249    0.04683 130.741  < 2e-16 ***
A2  0.16537    0.06365   2.598  0.00938 **
A3 -2.33830    0.15786 -14.812  < 2e-16 ***
---
predictions:
  1   2   3
456 538  44
```

Case II. If $y_3 = y_{01} = 0$, the third column in the derivative matrix turns to zero and the rank of the matrix decreases to $r = 3$ which is smaller than the number of the model parameters. The model deficiency is $d = p - r = 4 - 3 = 1$ and $\boldsymbol{\alpha}^{\mathsf{T}} = (0,0,1,-1)$. Again the elements 1 and -1 of the vector could be any non-zero value or function of the parameters. The two zero elements of the vector represent two parameters which are directly estimable, i.e. θ, θ^X. Because the rank of the matrix is three, there must exist one more estimable parameter. In order to find that one, we refer to Theorem 2.2. The corresponding partial differential equation (PDE) in (2.3) is,

$$\sum_{s=1}^{4} \alpha_{s1} \frac{\partial f}{\partial \theta_s} = 0 + 0 + \frac{\partial f}{\partial \theta^Y} - \frac{\partial f}{\partial \theta^{XY}} = 0.$$

The solution for this PDE is, $f = \theta^Y + \theta^{XY}$, which is the third estimable parameter. Thus, the smaller set of the estimable parameters is $\boldsymbol{\theta}'^T = (\theta, \theta^X, \theta^Y + \theta^{XY})$. This denotes as θ^Y is not estimable in model (2.5), $\log \mu_{01}$ is not estimable either. $y_{01} = 0$ is treated as a structural zero now because its corresponding model equation cannot be written by combinations of estimable parameters in $\boldsymbol{\theta}'$. The reduced full rank model with rank 3 and degree of freedom of $d.f = 3 - 3 = 0$ is,

$$\begin{aligned} m_1 &= \log \mu_1 = \log \mu_{00} = \theta, \\ m_2 &= \log \mu_2 = \log \mu_{10} = \theta + \theta^X, \\ m_4 &= \log \mu_4 = \log \mu_{11} = \theta + \theta^X + \theta^Y + \theta^{XY}. \end{aligned}$$

With the data vector $\mathbf{y}^\mathsf{T} = (456, 538, 911)$, the parameter estimates corresponding to $\boldsymbol{\theta}'$ are,

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  6.12249    0.04683 130.741  < 2e-16 ***
A2  0.16537    0.06365   2.598  0.00938 **
A3  0.69205    0.05736  12.064  < 2e-16 ***
---
Predictions:
  1   2   3
456 538 911
```

Case III. If $y_2 = y_{10} = 0$, we follow the same procedure again. The second column in the derivative matrix turns to zero and the rank of the matrix decreases to $r = 3$. The model deficiency is $d = p - r = 4 - 3 = 1$ and $\boldsymbol{\alpha}^\mathsf{T} = (0, 1, 0, -1)$. The two zero elements of the vector represent two parameters which are directly estimable, i.e. θ, θ^Y. Solving the corresponding partial differential equation in (2.3) gives us the third estimable parameter,

$$\sum_{s=1}^{4} \alpha_{s1} \frac{\partial f}{\partial \theta_s} = 0 + \frac{\partial f}{\partial \theta^X} + 0 - \frac{\partial f}{\partial \theta^{XY}} = 0.$$

The solution for this PDE is $f = \theta^X + \theta^{XY}$. Thus, the smaller set of the estimable parameters is $\boldsymbol{\theta}'^T = (\theta, \theta^X + \theta^{XY}, \theta^Y)$. This denotes as θ^X is not estimable in model (2.5), $\log \mu_{10}$ is not estimable either. $y_{10} = 0$ is treated as a structural zero now. The reduced full rank model with rank 3 and degree

of freedom of $d.f = 3-3 = 0$ is,

$$\begin{aligned} m_1 &= \log \mu_1 = \log \mu_{00} = \theta, \\ m_3 &= \log \mu_3 = \log \mu_{01} = \theta + \theta^Y, \\ m_4 &= \log \mu_4 = \log \mu_{11} = \theta + \theta^X + \theta^Y + \theta^{XY}. \end{aligned}$$

With the data vector $\mathbf{y}^\mathsf{T} = (456, 44, 911)$, the parameter estimates corresponding to $\boldsymbol{\theta}'$ are,

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  6.12249    0.04683  130.74   <2e-16 ***
A2  3.03035    0.15435   19.63   <2e-16 ***
A3 -2.33830    0.15786  -14.81   <2e-16 ***
---
predictions:
  1   2   3
456  44 911
```

Case IV. If $y_1 = y_{00} = 0$, we follow the same procedure but the result is a bit different. The first column in the derivative matrix turns to zero and the rank of the matrix decreases to $r = 3$. The model deficiency is $d = p - r = 4 - 3 = 1$ and $\boldsymbol{\alpha}^\mathsf{T} = (1, -1, -1, 1)$. There are no zero elements in the vector which denotes none of the parameters are directly estimable. As the model rank is 3, we must find three estimable combinations of parameters by solving the corresponding PDE in (2.3) which is,

$$\sum_{s=1}^{4} \alpha_{s1} \frac{\partial f}{\partial \theta_s} = \frac{\partial f}{\partial \theta} - \frac{\partial f}{\partial \theta^X} - \frac{\partial f}{\partial \theta^Y} + \frac{\partial f}{\partial \theta^{XY}} = 0.$$

The minimal linearly independent set of solutions for this PDE is,

$$f = \theta + \theta^X, \theta + \theta^Y, \theta - \theta^{XY},$$

thus the smaller set of the estimable parameters is,

$$\boldsymbol{\theta}'^T = (\theta + \theta^X, \theta + \theta^Y, \theta - \theta^{XY}).$$

y_{00} contributes to estimate θ whose estimate goes to minus infinity here, thus other parameters are not estimable either because all cell means are functions including θ. We are not able to define and estimate $\log \mu_{00}$ by using this set of estimable combinations of parameters. So, $y_{00} = 0$ is treated

as a structural zero now. The reduced full rank model with rank 3 and degree of freedom of $d.f = 3-3 = 0$ is,

$$\begin{aligned} m_2 &= \log\mu_2 = \log\mu_{10} = \theta + \theta^X, \\ m_3 &= \log\mu_3 = \log\mu_{01} = \theta + \theta^Y, \\ m_4 &= \log\mu_4 = \log\mu_{11} = \theta + \theta^X + \theta^Y + \theta^{XY}. \end{aligned}$$

In the matrix form it is shown as,

$$\begin{bmatrix} \log\mu_{10} \\ \log\mu_{01} \\ \log\mu_{11} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \theta + \theta^X \\ \theta + \theta^Y \\ \theta - \theta^{XY} \end{bmatrix}.$$

With the data vector $\mathbf{y}^\mathsf{T} = (538, 44, 911)$ the parameter estimates corresponding to $\boldsymbol{\theta}'$ are,

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  6.28786    0.04311  145.85   <2e-16 ***
A2  3.78419    0.15076   25.10   <2e-16 ***
A3 -3.25751    0.16026  -20.33   <2e-16 ***
---
predictions:
  1   2   3
538  44 911
```

More than one zero observation

The method described in Sections 2.2 and 2.3 is applicable for any number of zero observations in the contingency table. As an example, we consider fitting model (2.5) to Table 2.1, assuming two cells y_{01} and y_{11} are zero. Then all elements in third and fourth columns of the derivative matrix (2.6) become zero, and we have,

$$r = 2, \qquad d = p - r = 4 - 2 = 2, \qquad \alpha^\mathsf{T} = (0,0,1,1), \qquad \boldsymbol{\theta}'^T = (\theta, \theta^X).$$

As only θ and θ^X are estimable, $\log\mu_{01}$ and $\log\mu_{11}$ cannot be defined using the estimable parameters. $y_{01} = 0$ and $y_{11} = 0$ are treated as structural zeros now. The reduced full rank model with rank 2 and degree of freedom of $d.f = 2-2 = 0$ is,

$$\begin{aligned} m_1 &= \log\mu_1 = \log\mu_{10} = \theta, \\ m_2 &= \log\mu_2 = \log\mu_{10} = \theta + \theta^X. \end{aligned}$$

2.5 Alternative specifications

In defining the general log-linear model (1.5), we specified the model as a Poisson log-linear model with corner point constraints on the model parameters. In this section, we briefly consider some alternative specifications. Sum to zero constraints and a multinomial distribution as the sampling distribution are applied to the small model (2.5) which was used before. We also look at the corresponding independence model for Table 2.1, as an example of an unsaturated model and illustrate an example of a non-hierarchical model.

2.5.1 Sum to zero constraints

In this section, we fit a saturated log-linear model to the 2×2 Table 2.1, adopting sum to zero constraints instead of corner point constraints. A sum to zero constraint does not set any level of the parameters equal to zero but assumes that the summation of parameters over each index equals to zero. Therefore, for this example with two existent levels for each variable, the Poisson log-linear model with sum to zero constraints is,

$$\begin{aligned}
m_1 &= \log \mu_1 = \log \mu_{00} = \beta + \beta_0^X + \beta_0^Y + \beta_{00}^{XY}, \\
m_2 &= \log \mu_2 = \log \mu_{10} = \beta + \beta_1^X + \beta_0^Y + \beta_{10}^{XY}, \\
m_3 &= \log \mu_3 = \log \mu_{01} = \beta + \beta_0^X + \beta_1^Y + \beta_{01}^{XY}, \\
m_4 &= \log \mu_4 = \log \mu_{11} = \beta + \beta_1^X + \beta_1^Y + \beta_{11}^{XY}.
\end{aligned}$$

The model parameters vector is $\boldsymbol{\theta}^\mathsf{T} = (\beta, \beta_0^X, \beta_0^Y, \beta_{00}^{XY}, \beta_1^X, \beta_{10}^{XY}, \beta_1^Y, \beta_{01}^{XY}, \beta_{11}^{XY})$. The parameters are shown as β to emphasize the fact that they are different to those in model (2.5). For example, the interpretation of the intercept β here is not the same as θ in the model (2.5). The sum to zero constraints imply that,

$$\begin{aligned}
\sum_{i=0}^{1} \beta_i^X &= 0 \qquad \rightarrow \qquad \beta_1^X = -\beta_0^X, \\
\sum_{j=0}^{1} \beta_j^X &= 0 \qquad \rightarrow \qquad \beta_1^Y = -\beta_0^Y, \\
\sum_{i=0}^{1} \beta_{ij}^{XY} &= 0 \qquad \rightarrow \qquad \beta_{10}^{XY} = -\beta_{00}^{XY}, \quad \beta_{11}^{XY} = -\beta_{01}^{XY}, \\
\sum_{j=0}^{1} \beta_{ij}^{XY} &= 0 \qquad \rightarrow \qquad \beta_{01}^{XY} = -\beta_{00}^{XY}, \quad \beta_{11}^{XY} = -\beta_{10}^{XY}, \qquad \beta_{00}^{XY} = \beta_{11}^{XY}.
\end{aligned}$$

So the log-linear model equations could also be defined by only four parameters in the parameter vector $\boldsymbol{\theta}^{\mathsf{T}} = (\beta, \beta_0^X, \beta_0^Y, \beta_{00}^{XY})$, as,

$$\begin{aligned} m_1 &= \log \mu_1 = \log \mu_{00} = \beta + \beta_0^X + \beta_0^Y + \beta_{00}^{XY}, \\ m_2 &= \log \mu_2 = \log \mu_{10} = \beta - \beta_0^X + \beta_0^Y - \beta_{00}^{XY}, \\ m_3 &= \log \mu_3 = \log \mu_{01} = \beta + \beta_0^X - \beta_0^Y - \beta_{00}^{XY}, \\ m_4 &= \log \mu_4 = \log \mu_{11} = \beta - \beta_0^X - \beta_0^Y + \beta_{00}^{XY}. \end{aligned} \tag{2.7}$$

The parameter estimates for fitting this model to the data in Table 2.2 are obtained as:

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  5.75227    0.04174  137.81   <2e-16 ***
A2 -0.79893    0.04174  -19.14   <2e-16 ***
A3  0.45290    0.04174   10.85   <2e-16 ***
A4  0.71625    0.04174   17.16   <2e-16 ***
```

These estimates are different with the previous ones in the model (2.5), because as mentioned before, the parameter interpretations in these two models are different. However, there is an obvious relation between them as,

$$\begin{aligned} \beta + \beta_0^X + \beta_0^Y + \beta_{00}^{XY} &= \theta, \\ \beta - \beta_0^X + \beta_0^Y - \beta_{00}^{XY} &= \theta + \theta^X, \\ \beta + \beta_0^X - \beta_0^Y - \beta_{00}^{XY} &= \theta + \theta^Y, \\ \beta - \beta_0^X - \beta_0^Y + \beta_{00}^{XY} &= \theta + \theta^X + \theta^Y + \theta^{XY}. \end{aligned}$$

The equations hold in the numerical example. For example, the first equation implies that,

$$5.7522710 - 0.7989298 + 0.4529047 + 0.7162469 = 6.12249.$$

The presence of zero observations

The effect of the presence of zero observations in model (2.7) is a bit different with the one in model (2.5), because of the relation between each cell mean of the model with all the four parameters in $\boldsymbol{\theta}$. To observe the difference, assume $y_2 = y_{10} = 0$ in Table 2.2. Then the parameter estimates are reported as:

```
Coefficients:
    Estimate Std. Error z value Pr(>|z|)
A1    -1.395  10561.791   0.000    1.000
```

```
A2      6.349  10561.791   0.001    1.000
A3     -6.695  10561.791  -0.001    0.999
A4      7.864  10561.791   0.001    0.999
```

Standard errors for all estimates are equal large numbers indicating that estimates are not reliable. In the symbolic way, the derivative matrix (2.4) is represented now as,

$$D = \left[\frac{\partial y_i \log \mu_i}{\partial \theta_s}\right] = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ y_1 & -y_2 & y_3 & -y_4 \\ y_1 & y_2 & -y_3 & -y_4 \\ y_1 & -y_2 & -y_3 & y_4 \end{bmatrix} \qquad i = 1,2,3,4, \quad s = 1,2,3,4. \quad (2.8)$$

After setting the second column to zero, we have $\boldsymbol{\alpha}^{\mathsf{T}} = (-1,1,-1,1)$, which reveals that none of the parameters are directly estimable. The rank of this matrix is 3, so there must be three estimable combinations of parameters. They are $\beta + \beta_0^X, -\beta + \beta_0^Y, \beta + \beta_{00}^{XY}$, derived by solving the corresponding partial differential equations in (2.3). By setting any of the cell counts equal to zero there are three estimable combinations of parameters, unlike model (2.5) in which some of the model parameters in the initial vector of parameters $\boldsymbol{\theta}$ could be directly estimable. Hence, we retain corner point constraints as they provide an easier version in defining the model.

2.5.2 Multinomial sampling distribution

It is well known that fitting a log-linear model to a contingency table data produces identical inferences whether Poisson or multinomial sampling distributions are considered. Haberman [1973] proved in a theorem that if $\hat{\boldsymbol{\mu}}^{\mathsf{m}}$ is the MLE of cell means for a multinomial model and $\hat{\boldsymbol{\mu}}$ is their MLE for a Poisson model, then $\hat{\boldsymbol{\mu}}^{\mathsf{m}} = \hat{\boldsymbol{\mu}}$.

Suppose the sum of the cell counts in Table 2.1 is fixed as N. Then the data vector is from a multinomial distribution as defined in (1.2),

$$P\left(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, Y_4 = y_4 \mid \sum_{i=1}^{4} y_i = N\right) = \frac{N!}{\prod_{i=1}^{4} y_i!} \prod_{i=1}^{4} \pi_i^{y_i}.$$

Only three parameters are free to estimate since $\sum_i \pi_i = 1$. For this distribution, cell means are,

$$\mu_i = N\pi_i, \qquad \pi_i = \frac{\mu_i}{\sum_i \mu_i},$$

and for a saturated model $\hat{\pi}_i = p_i = \frac{y_i}{N}$. The log-likelihood function for this multinomial distribution is,

$$l(\boldsymbol{\mu}(\boldsymbol{\theta})) = \sum_{i=1}^{4} y_i \log \frac{\mu_i}{\sum_i \mu_i}.$$

A multinomial log-linear model is structured using π_i rather that μ_i. For model (2.5), cell probabilities are as follows and the model intercept cancels out.

$$\pi_1 = \pi_{00} = \frac{e^{\theta}}{e^{\theta+\theta^X+\theta^Y+\theta^{XY}} + e^{\theta+\theta^X} + e^{\theta+\theta^Y} + e^{\theta}} = \frac{1}{e^{\theta^X+\theta^Y+\theta^{XY}} + e^{\theta^X} + e^{\theta^Y} + 1},$$
$$\pi_2 = \pi_{10} = \frac{e^{\theta+\theta^X}}{e^{\theta+\theta^X+\theta^Y+\theta^{XY}} + e^{\theta+\theta^X} + e^{\theta+\theta^Y} + e^{\theta}} = \frac{e^{\theta^X}}{e^{\theta^X+\theta^Y+\theta^{XY}} + e^{\theta^X} + e^{\theta^Y} + 1},$$
$$\pi_3 = \pi_{01} = \frac{e^{\theta+\theta^Y}}{e^{\theta+\theta^X+\theta^Y+\theta^{XY}} + e^{\theta+\theta^X} + e^{\theta+\theta^Y} + e^{\theta}} = \frac{e^{\theta^Y}}{e^{\theta^X+\theta^Y+\theta^{XY}} + e^{\theta^X} + e^{\theta^Y} + 1},$$
$$\pi_4 = \pi_{11} = \frac{e^{\theta+\theta^X+\theta^Y+\theta^{XY}}}{e^{\theta+\theta^X+\theta^Y+\theta^{XY}} + e^{\theta+\theta^X} + e^{\theta+\theta^Y} + e^{\theta}} = \frac{e^{\theta^X+\theta^Y+\theta^{XY}}}{e^{\theta^X+\theta^Y+\theta^{XY}} + e^{\theta^X} + e^{\theta^Y} + 1}.$$

These elements could be simplified by dividing each cell probability or cell mean over a non-zero probability or cell mean. Here dividing by μ_1 makes the equations simpler,

$$\log\frac{\pi_1}{\pi_1} = \log\frac{\mu_1}{\mu_1} = \log\frac{e^{\theta}}{e^{\theta}} = \log 1 = 0,$$
$$\log\frac{\pi_2}{\pi_1} = \log\frac{\mu_2}{\mu_1} = \log\frac{e^{\theta+\theta^X}}{e^{\theta}} = \log e^{\theta^X} = \theta^X,$$
$$\log\frac{\pi_3}{\pi_1} = \log\frac{\mu_3}{\mu_1} = \log\frac{e^{\theta+\theta^Y}}{e^{\theta}} = \log e^{\theta^Y} = \theta^Y,$$
$$\log\frac{\pi_4}{\pi_1} = \log\frac{\mu_4}{\mu_1} = \log\frac{e^{\theta+\theta^X+\theta^Y+\theta^{XY}}}{e^{\theta}} = \log e^{\theta^X+\theta^Y+\theta^{XY}} = \theta^X + \theta^Y + \theta^{XY}.$$

Any monotonic function of cell means could be used to form the derivative matrix. Catchpole and Morgan [2001] form the derivative matrix as $D_{si}(\boldsymbol{\theta}) = \dfrac{\partial \log \pi_i}{\partial \theta_s}$. The score vector, shown by $\mathbf{U}(\boldsymbol{\theta})$, is $\mathbf{U}(\boldsymbol{\theta}) = \left(\dfrac{\partial l}{\partial \theta_1}, ..., \dfrac{\partial l}{\partial \theta_p}\right)^{\mathsf{T}} = D(\boldsymbol{\theta})\mathbf{y}$ which implies that the effect of a zero cell count is equivalent to setting the corresponding column of D to zero. To make the derivative matrix, we choose the monotonic function of the cell means as $y_i \log \dfrac{\mu_i}{\mu_1}$ to keep the derivative matrix similar to the one from the Poisson model and also illustrate the fact that the intercept cancels out from a multinomial log-linear model. Then the derivative matrix (2.4), for parameter vector $\boldsymbol{\theta}^{\mathsf{T}} = (\theta^X, \theta^Y, \theta^{XY})$ is,

$$D = \left[\frac{\partial y_i \log(\mu_i/\mu_1)}{\partial \theta_s}\right] = \begin{bmatrix} y_2 & 0 & y_4 \\ 0 & y_3 & y_4 \\ 0 & 0 & y_4 \end{bmatrix}, \qquad i = 2,3,4, \quad s = 1,2,3.$$

The rank of this derivative matrix is 3 which is equal to the number of parameters, so all those three parameters are estimable when none of the cell counts is zero. y_1 does not exist in this matrix and it must not be zero as we have divided all cell probabilities over the first cell's probability. If one cell count is observed as zero, for example $y_2 = 0$, then the rank of the derivative matrix is decreased to 2 and the deficiency is 1. Then, $\boldsymbol{\alpha}^\mathsf{T} = (\alpha_{11}, \alpha_{21}, \alpha_{31}) = (\alpha_{11}, 0, -\alpha_{11})$ which we write it as $\boldsymbol{\alpha}^\mathsf{T} = (1, 0, -1)$. Thus the estimable parameters are θ^Y and $\theta^X + \theta^{XY}$. If $y_3 = 0$, then the rank of the derivative matrix is 2, the deficiency is 1 and we have $\boldsymbol{\alpha}^\mathsf{T} = (0, 1, -1)$. Thus the estimable parameters are θ^X and $\theta^Y + \theta^{XY}$. If $y_4 = 0$, then the rank of the derivative matrix is 2, the deficiency is 1 and we have $\boldsymbol{\alpha}^\mathsf{T} = (0, 0, 1)$. So the estimable parameters are θ^X and θ^Y. After considering the fact that there is not an intercept parameter in this model, these results match the ones from a Poisson model.

Wang et al. [2016] imply the following way to discover estimable and inestimable parameters in fitting a multinomial saturated log-linear model to the data in Table 2.1, including zero observations. By rearranging model (2.5), we can write,

$$\begin{aligned}
\theta &= \log \mu_{00}, \\
\theta^X &= \log \frac{\mu_{10}}{\mu_{00}}, \\
\theta^Y &= \log \frac{\mu_{01}}{\mu_{00}}, \\
\theta^{XY} &= \log \frac{\mu_{11}\mu_{00}}{\mu_{01}\mu_{10}}.
\end{aligned}$$

If we assume that the model includes an intercept and consider the data vector $\mathbf{y}^\mathsf{T} = (0, y_{10}, y_{01}, y_{11})$, then defining the model based on $\hat{\pi}_i = p_i = \frac{y_i}{N}$ gives the parameters estimates as,

$$\begin{aligned}
\hat{\theta} &= \log p_{00} \to -\infty, \\
\hat{\theta}^X &= \log \frac{p_{10}}{p_{00}} \to +\infty, \\
\hat{\theta}^Y &= \log \frac{p_{01}}{p_{00}} \to +\infty, \\
\hat{\theta}^{XY} &= \log \frac{p_{11}p_{00}}{p_{01}p_{10}} \to -\infty,
\end{aligned}$$

indicating none of them is directly estimable. However, $\hat{\theta} + \hat{\theta}^X = \log p_{10} = \log \frac{y_{10}}{N}$ converges to a finite value, so do $\hat{\theta} + \hat{\theta}^Y = \log p_{10}$ and $\hat{\theta} - \hat{\theta}^{XY} = \log \frac{p_{10}p_{01}}{p_{11}}$ which are the three estimable combinations of parameters in this example (the result holds for a Poisson log-linear model; for a multinomial model without an intercept it actually

means that non of the parameters in $\boldsymbol{\theta}$ are estimable). The approach could be considered for having different zero cell counts as well, but it gets difficult to detect the estimable combinations of parameters for larger models.

2.5.3 Independence model

Here the parameter redundancy method is extended to an unsaturated log-linear model. The procedure, which includes defining the model, building the derivative matrix, finding $\boldsymbol{\alpha}$ vector and solving the differential equations remains the same despite of the type of the model.

An independence log-linear model is fitted to Table 2.1 here as an example of an unsaturated model. The saturated model (2.5) is altered to the following model which does not contain a parameter describing the interaction of the two variables,

$$\begin{aligned}
m_1 &= \log \mu_1 = \log \mu_{00} = \theta, \\
m_2 &= \log \mu_2 = \log \mu_{10} = \theta + \theta^X, \\
m_3 &= \log \mu_3 = \log \mu_{01} = \theta + \theta^Y, \\
m_4 &= \log \mu_4 = \log \mu_{11} = \theta + \theta^X + \theta^Y.
\end{aligned}$$

The derivative matrix (2.4) for parameter vector $\boldsymbol{\theta}^\mathsf{T} = (\theta, \theta^X, \theta^Y)$ is,

$$D = \left[\frac{\partial y_i \log \mu_i}{\partial \theta_s}\right] = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ 0 & y_2 & 0 & y_4 \\ 0 & 0 & y_3 & y_4 \end{bmatrix}, \qquad i = 1,2,3,4, \quad s = 1,2,3.$$

The rank of this derivative matrix is 3 which is equal to the number of parameters (the smaller dimension of the matrix), so when none of the cell counts is zero, the model is full rank and all three parameters are estimable. If any of the four cell counts are zero, the rank of the matrix remains three indicating this independence model for a 2×2 table is identifiable and full rank in the presence of one zero observation. Now we investigate the effect of having two zero cells in the table. There are six different combinations for two zero cell counts:

Case I. If $y_1 = y_2 = 0$, then $r = 2$, $d = 1$, $\boldsymbol{\alpha}^\mathsf{T} = (1, 0, -1)$ and $\boldsymbol{\theta}'^\mathsf{T} = (\theta^X, \theta + \theta^Y)$.

Case II. If $y_1 = y_3 = 0$, then $r = 2$, $d = 1$, $\boldsymbol{\alpha}^\mathsf{T} = (1, -1, 0)$ and $\boldsymbol{\theta}'^\mathsf{T} = (\theta + \theta^X, \theta^Y)$.

Case III. If $y_1 = y_4 = 0$, then $r = 2$, $d = 1$, $\boldsymbol{\alpha}^\mathsf{T} = (1, -1, -1)$ and $\boldsymbol{\theta}'^\mathsf{T} = (\theta + \theta^X, \theta + \theta^Y)$.

Case IV. If $y_2 = y_3 = 0$, then $r = 2$, $d = 1$, $\boldsymbol{\alpha}^\mathsf{T} = (0, 1, -1)$ and $\boldsymbol{\theta}'^\mathsf{T} = (\theta, \theta^X + \theta^Y)$.

Case V. If $y_2 = y_4 = 0$, then $r = 2$, $d = 1$, $\boldsymbol{\alpha}^{\mathsf{T}} = (0,1,0)$ and $\boldsymbol{\theta}'^{\mathsf{T}} = (\theta, \theta^Y)$.

Case VI. If $y_3 = y_4 = 0$, then $r = 2$, $d = 1$, $\boldsymbol{\alpha}^{\mathsf{T}} = (0,0,1)$ and $\boldsymbol{\theta}'^{\mathsf{T}} = (\theta, \theta^X)$.

As a numerical example, we fit the independence model to the data in Table 2.2. The parameter estimates for the model and cell mean predictions, without any zero observations in the table, are:

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  5.54127    0.04993 110.983   <2e-16 ***
A2  1.06402    0.05187  20.515   <2e-16 ***
A3 -0.04003    0.04531  -0.883    0.377
---
predictions:
       1        2        3        4
255.0026 738.9974 244.9974 710.0026
```

If one cell count is zero, for example, the first one $y_1 = 0$, estimates for all three parameters are given with reasonable standard errors although the estimates are not the same as the previous ones.

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  2.76351    0.15465   17.87   <2e-16 ***
A2  3.49444    0.15303   22.84   <2e-16 ***
A3  0.57385    0.05391   10.64   <2e-16 ***
---
predictions:
        1         2         3         4
 15.85532 522.14468  28.14468 926.85532
```

By setting more than one cell equal to zero, the large standard errors for some parameter estimates emerge. For example, if $y_1 = y_2 = 0$, then the parameter estimates are:

```
Coefficients:
     Estimate Std. Error z value Pr(>|z|)
A1   -25.8194 52624.4897   0.000        1
A2     3.0304     0.1544  19.633   <2e-16 ***
A3    29.6036 52624.4897   0.001        1
```

As mentioned before, the estimable set of parameters in this case is $\boldsymbol{\theta}'^{T} = (\theta^X, \theta + \theta^Y)$. Defining and estimating $\log \mu_{00}$ and $\log \mu_{10}$ are not possible with only these two estimable parameters . The two cells with zero observations, $y_{00} = 0$ and y_{10} are treated

as structural zeros now. The reduced full rank model with rank 2 and degree of freedom of $d.f = 2-2 = 0$ is,

$$m_3 = \log \mu_3 = \log \mu_{01} = \theta + \theta^Y,$$
$$m_4 = \log \mu_4 = \log \mu_{11} = \theta + \theta^X + \theta^Y.$$

With the data vector $\mathbf{y}^\mathsf{T} = (44, 911)$, the parameter estimates are:

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1   3.0304     0.1544   19.63   <2e-16 ***
A2   3.7842     0.1508   25.10   <2e-16 ***
---
prediction:
  1   2
 44 911
```

The next combination of zero cells create a different situation. If the two zero cell counts are $y_1 = y_4 = 0$, the parameter estimates are:

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1   3.7056     0.1512   24.50   <2e-16 ***
A2   2.5037     0.1568   15.97   <2e-16 ***
A3  -2.5037     0.1568  -15.97   <2e-16 ***
```

Large standard errors are not observed here, although we obtained the estimable set of parameters as $\boldsymbol{\theta}'^T = (\theta + \theta^X, \theta + \theta^Y)$ which suggests $\log \mu_{00}$ and $\log \mu_{11}$ are inestimable. So the model is parameter redundant but the numerical routine has computed the maximum likelihood estimates of the parameters. In Section 2.7, we explain the reason of this occurrence and discuss these kind of models in detail in Chapter 4.

2.5.4 Non-hierarchical models

The parameter redundancy approach is not limited to the hierarchical models. Assume the following non-hierarchical model is fitted to the contingency table 2.1.

$$m_1 = \log \mu_1 = \log \mu_{00} = \theta,$$
$$m_2 = \log \mu_2 = \log \mu_{10} = \theta + \theta^X,$$
$$m_3 = \log \mu_3 = \log \mu_{01} = \theta$$
$$m_4 = \log \mu_4 = \log \mu_{11} = \theta + \theta^X + \theta^{XY}.$$

The derivative matrix (2.4) for parameter vector $\boldsymbol{\theta}^{\mathsf{T}} = (\theta, \theta^X, \theta^{XY})$ is,

$$D = \left[\frac{\partial y_i \log \mu_i}{\partial \theta_s}\right] = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ 0 & y_2 & 0 & y_4 \\ 0 & 0 & 0 & y_4 \end{bmatrix}, \qquad i = 1,2,3,4, \quad s = 1,2,3.$$

The rank of the matrix is 3 for all positive cell counts. One zero cell may or may not reduce the rank as:

Case I. If $y_1 = 0$, then $r = 3,\ d = 0$.

Case II. If $y_2 = 0$, then $r = 2,\ d = 1,\ \boldsymbol{\alpha}^{\mathsf{T}} = (0,1,-1)$ and $\boldsymbol{\theta}'^{\mathsf{T}} = (\theta, \theta^X + \theta^{XY})$.

Case III. If $y_3 = 0$, then $r = 3,\ d = 0$.

Case IV. If $y_4 = 0$, then $r = 2,\ d = 1,\ \boldsymbol{\alpha}^{\mathsf{T}} = (0,0,1)$ and $\boldsymbol{\theta}'^{\mathsf{T}} = (\theta, \theta^X)$.

As a numerical example, this model is fitted to the data in Table 2.2. The parameter estimates for the model and cell mean predictions, without any zero observations in the table, are:

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
x1  5.52146    0.04472 123.464   <2e-16 ***
x2  0.76640    0.06212  12.338   <2e-16 ***
x3  0.52668    0.05437   9.686   <2e-16 ***
---
predictions:
  1   2   3   4
250 538 250 911
```

Assume $y_2 = 0$. Then the estimates for some of the parameters have large standard errors.

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
x1   5.52146    0.04472 123.464   <2e-16 ***
x2 -19.82405  773.78384  -0.026    0.980
x3  21.11713  773.78383   0.027    0.978
---
predictions:
           1            2            3            4
2.500000e+02 6.144212e-07 2.500000e+02 9.110000e+02
```

2.6 Example in a 3^3 contingency table

A larger contingency table is considered now and an unsaturated model is fitted to the data as an example for a parameter redundant model. In the previous examples, if the model was parameter redundant, then the zero cell counts were treated as structural zeros since their means were not estimable according to the estimable parameters of the model. Nonetheless, it is quite possible that one or some cells with sampling zero observations have estimable cell means as a result of having corresponding estimable parameters. In this case, those cells are kept in the model as well, we fit a reduced non-saturated model with a positive degree of freedom and estimate the cell means for all estimable cells of the table.

Table 2.3 is an example taken from Fienberg and Rinaldo [2012a] with three variables X (rows), Y (columns), Z (layers) and three levels $(0,1,2)$ for each variable. Eight cell counts are observed as sampling zeros and the other cell counts are assumed to be positive values from the Poisson distribution. Cell counts are ordered inside the table according to equation (1.6).

0	y_4	y_7
0	y_5	y_8
y_3	y_6	y_9

y_{10}	y_{13}	y_{16}
y_{11}	y_{14}	**0**
y_{12}	0	0

0	y_{22}	**0**
0	y_{23}	y_{26}
y_{21}	y_{24}	y_{27}

Table 2.3 Observations in a 3^3 contingency table.

The desirable hierarchical model to fit to the table data is (XY, XZ, YZ), denoting that only main effects and first-order interactions of variables are present in the model. This log-linear model is shown as,

$$\log \boldsymbol{\mu}_{27\times 1} = \mathsf{A}_{27\times 19}\boldsymbol{\theta}_{19\times 1},$$

such that the 19 model parameters are,

$$\begin{aligned}\boldsymbol{\theta}^{\mathsf{T}} = (&\theta, \theta_1^X, \theta_2^X, \theta_1^Y, \theta_2^Y, \theta_1^Z, \theta_2^Z, \theta_{11}^{XY}, \theta_{21}^{XY}, \theta_{12}^{XY}, \theta_{22}^{XY},\\ &\theta_{11}^{YZ}, \theta_{21}^{YZ}, \theta_{12}^{YZ}, \theta_{22}^{YZ}, \theta_{11}^{XZ}, \theta_{21}^{XZ}, \theta_{12}^{XZ}, \theta_{22}^{XZ}),\end{aligned}$$

and the model's design matrix is shown in Figure 2.1. We use a function in `R`, given in Appendix A, to make the design matrix and the derivative matrix for a specified model with m variables and l levels for each variable.

Now the procedure explained in Sections 2.2 and 2.3 is followed to detect parameter redundancy in this model. We use two procedures in `Maple`, given in Appendix A, to

1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
7	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
10	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
12	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
13	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
14	1	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0
15	1	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0
16	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
17	1	1	0	0	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0
18	1	0	1	0	1	1	0	0	0	0	1	0	1	0	0	0	1	0	0
19	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
20	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
21	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
22	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
23	1	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0
24	1	0	1	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1
25	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0
26	1	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0
27	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1

Fig. 2.1 The design matrix for model (XY, XZ, YZ) fitted to the 3^3 contingency table.

find the $\boldsymbol{\alpha}$ vector and solve the PDEs to achieve the estimable set of parameters. For examples with larger contingency tables, it is easier to find the $\boldsymbol{\alpha}$ vectors in `MATLAB` and use `Maple` only to solve the PDEs in the symbolic way. The rank of the corresponding derivative matrix is 18, meaning that there are only 18 estimable parameters in the model. The $\boldsymbol{\alpha}$ defined in (2.2) is,

$$\boldsymbol{\alpha}^{\mathsf{T}} = (1,0,-1,-1,-1,-1,0,0,1,0,1,1,1,0,0,0,1,0,0).$$

By considering this vector and solving the PDEs in (2.3), the vector of estimable parameters is obtained as,

$$\begin{aligned}\boldsymbol{\theta}'^{T} =&(\theta_1^X, \theta+\theta_2^X, \theta+\theta_1^Y, \theta+\theta_2^Y, \theta+\theta_1^Z, \theta_2^Z, \theta_{11}^{XY}, -\theta+\theta_{21}^{XY}, \theta_{12}^{XY},\\ &-\theta+\theta_{22}^{XY}, -\theta+\theta_{11}^{YZ}, -\theta+\theta_{21}^{YZ}, \theta_{12}^{YZ}, \theta_{22}^{YZ}, \theta_{11}^{XZ}, -\theta+\theta_{21}^{XZ}, \theta_{12}^{XZ}, \theta_{22}^{XZ}).\end{aligned}$$

It determines 21 out of 27 cell means are estimable including cells 17 and 25 with zero observations, which are shown in the table with bold zeros, since,

$$\begin{aligned}&\log\mu_{17} = \log\mu_{121} = \theta+\theta_1^X+\theta_2^Y+\theta_1^Z+\theta_{12}^{XY}+\theta_{21}^{YZ}+\theta_{11}^{XZ},\\ &\log\mu_{25} = \log\mu_{022} = \theta+\theta_2^Y+\theta_2^Z+\theta_{22}^{YZ},\end{aligned}$$

are only containing the estimable combinations of parameters provided in $\boldsymbol{\theta}'$. On the other hand, inestimable cell means include some inestimable combinations of the

1	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	5	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	6	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	8	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	9	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	10	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
9	11	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
10	12	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
11	13	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
12	14	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0
13	16	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
14	17	1	0	0	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0
15	21	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
16	22	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
17	23	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0
18	24	0	1	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1
19	25	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0
20	26	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0
21	27	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1

Fig. 2.2 The reduced design matrix for the log-linear model fitted to the 3^3 contingency table.

parameters,

$$\begin{aligned}
\log \mu_1 &= \log \mu_{000} = \theta, \\
\log \mu_2 &= \log \mu_{100} = \theta + \theta_1^X, \\
\log \mu_{15} &= \log \mu_{211} = \theta + \theta_2^X + \theta_1^Y + \theta_1^Z + \theta_{21}^{XY} + \theta_{11}^{YZ} + \theta_{21}^{XZ}, \\
\log \mu_{18} &= \log \mu_{221} = \theta + \theta_2^X + \theta_2^Y + \theta_1^Z + \theta_{22}^{XY} + \theta_{21}^{YZ} + \theta_{21}^{XZ}, \\
\log \mu_{19} &= \log \mu_{002} = \theta + \theta_2^Z, \\
\log \mu_{20} &= \log \mu_{102} = \theta + \theta_1^X + \theta_2^Z + \theta_{12}^{XZ}.
\end{aligned}$$

Therefore, the initial model must be reduced to a smaller model including 18 estimable parameters in $\boldsymbol{\theta}'$ and 21 estimable cell means excluding cells $1, 2, 15, 18, 19, 20$. Then the reduced full rank model, with rank 18 and with degrees of freedom $d.f = 21 - 18 = 3$, is,

$$\log \boldsymbol{\mu}_{21\times 1} = \mathsf{A}'_{21\times 18}\boldsymbol{\theta}'_{18\times 1}.$$

The reduced design matrix, matching the set of estimable cell means and parameters, is given in Figure 2.2. For a known $\mathbf{y}$, we can fit this model and find estimates for all the 18 estimable quantities and 21 estimable cell means.

2.7 The esoteric constraints

The likelihood function of parameter redundant models includes a flat ridge as stated in Theorem 2.4. This flat ridge is sometimes orthogonal to some parameters' axes, so those parameters still have unique maximum likelihood estimates [Catchpole et al., 1998] and this type of likelihood function imposes some extra constraints on the model

parameters. Those constraints are not arbitrary and are implied through the model's likelihood and permit estimating more model parameters. Knowledge on the existence and nature of these constraints, which are not reported by standard statistical software, reveals the true model that is fitted. The existence of such constraints, which we name them **esoteric constraints**, can be checked after detecting the parameter redundancy.

Consider the log-likelihood function for (1.5) as,

$$l(\boldsymbol{\theta}) = \sum_{\mathbf{i}} (y_{\mathbf{i}} \log \mu_{\mathbf{i}}(\boldsymbol{\theta}) - \mu_{\mathbf{i}}(\boldsymbol{\theta})) - \sum_{\mathbf{i}} \log y_{\mathbf{i}}!$$

and the corresponding score vector as,

$$\mathbf{U}(\boldsymbol{\theta}) = \left(\frac{\partial l}{\partial \theta_1}, \ldots, \frac{\partial l}{\partial \theta_p} \right)^{\mathsf{T}},$$

such that,

$$\frac{\partial l}{\partial \theta_s} = \sum_{\mathbf{i}} \left(\frac{y_{\mathbf{i}}}{\mu_{\mathbf{i}}(\boldsymbol{\theta})} - 1 \right) \frac{\partial \mu_{\mathbf{i}}(\boldsymbol{\theta})}{\partial \theta_s} = \sum_{\mathbf{i}} (y_{\mathbf{i}} - \mu_{\mathbf{i}}(\boldsymbol{\theta})) \frac{\partial \mu_{\mathbf{i}}(\boldsymbol{\theta})}{\partial \theta_s} \frac{1}{\mu_{\mathbf{i}}(\boldsymbol{\theta})}.$$

Whenever a model is parameter redundant and $\boldsymbol{\alpha}_j^{\mathsf{T}}(\boldsymbol{\theta}) D(\boldsymbol{\theta}) = \mathbf{0}$, $j = 1, \cdots, d$, it follows that $\boldsymbol{\alpha}_j^{\mathsf{T}}(\boldsymbol{\theta}) \mathbf{U}(\boldsymbol{\theta}) = 0$, $j = 1, \cdots, d$ [Catchpole and Morgan, 1997].

For a parameter redundant log-linear model, $\boldsymbol{\alpha}^{\mathsf{T}}(\boldsymbol{\theta}) \mathbf{U}(\boldsymbol{\theta}) = 0$ always holds. If $\boldsymbol{\alpha}_j^{\mathsf{T}}(\boldsymbol{\theta}) \mathbf{U}(\boldsymbol{\theta})$ are impossible to be zero with finite θs then the extra constraints do not exist and estimates for some model parameters tend to infinity. However, if imposing one or more constraints can make the expressions zero, then those are the esoteric constraints. Determining these constraints is more straightforward with using the Poisson sampling distribution rather than the multinomial distribution, in which we mentioned that the first cell count is non-zero.

These constraints do not exist for a saturated model with at least one zero cell in the table, thus the examples presented in Section 2.4 do not have esoteric constraints. A `Maple` procedure is given in Appendix A that can be used to find $\boldsymbol{\alpha}^{\mathsf{T}}(\boldsymbol{\theta}) \mathbf{U}(\boldsymbol{\theta})$ and determining the existence of any esoteric constraints. The esoteric constraints do not exist for the model studied in Section 2.6 since we have,

$$\begin{aligned} \boldsymbol{\alpha}^{\mathsf{T}} \mathbf{U}(\boldsymbol{\theta}) = & -e^{\theta + \theta_1^X + \theta_2^Z + \theta_{12}^{XZ}} - e^{\theta + \theta_2^X + \theta_1^Y + \theta_1^Z + \theta_{21}^{XY} + \theta_{11}^{YZ} + \theta_{21}^{XZ}} - e^{\theta + \theta_2^X + \theta_2^Y + \theta_1^Z + \theta_{22}^{XY} + \theta_{21}^{YZ} + \theta_{21}^{XZ}} \\ & - e^{\theta + \theta_2^Z} - e^{\theta} - e^{\theta + \theta_1^X}, \end{aligned}$$

which no constraints can make it equal to zero. Catchpole and Morgan [2001] fit a model with five parameters to a mark-recovery data set that provides a parameter redundant model with three estimable combinations of parameters. They mention that

	Y		
X	0	1	2
0	0	0	1
1	0	0	1
2	1	1	0

(a)

	Y		
X	0	1	2
0	0	0	0
1	0	1	1
2	0	1	1

(b)

Table 2.4 Observations in 3^2 contingency tables

the unique MLE for all five parameters can be obtained after imposing two constraints. The next example presents a model that depending on the place of the zero entries can have esoteric constraints.

Example 2.1. Consider two variables with three levels for each, as shown in Table 2.4. The model that we fit to such a data, based on the corner point constraints, includes only the main effects of the variables,

$$\log \mu_{ij} = \theta + \theta_i^X + \theta_j^Y, \qquad i = 0,1,2,\ j = 0,1,2,$$

and can also be shown as,

$$\begin{aligned}
\log \mu_1 &= \log \mu_{00} = \theta, \\
\log \mu_2 &= \log \mu_{10} = \theta + \theta_1^X, \\
\log \mu_3 &= \log \mu_{20} = \theta + \theta_2^X, \\
\log \mu_4 &= \log \mu_{01} = \theta + \theta_1^Y, \\
\log \mu_5 &= \log \mu_{11} = \theta + \theta_1^X + \theta_1^Y, \\
\log \mu_6 &= \log \mu_{21} = \theta + \theta_2^X + \theta_1^Y, \\
\log \mu_7 &= \log \mu_{02} = \theta + \theta_2^Y, \\
\log \mu_8 &= \log \mu_{12} = \theta + \theta_1^X + \theta_2^Y, \\
\log \mu_9 &= \log \mu_{22} = \theta + \theta_2^X + \theta_2^Y.
\end{aligned}$$

The parameter vector is $\boldsymbol{\theta}^\mathsf{T} = (\theta, \theta_1^X, \theta_2^X, \theta_1^Y, \theta_2^Y)$ and the corresponding derivative matrix is,

$$D = \left[\frac{\partial y_i \log \mu_i}{\partial \theta_s}\right] = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 \\ 0 & y_2 & 0 & 0 & y_5 & 0 & 0 & y_8 & 0 \\ 0 & 0 & y_3 & 0 & 0 & y_6 & 0 & 0 & y_9 \\ 0 & 0 & 0 & y_4 & y_5 & y_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & y_7 & y_8 & y_9 \end{bmatrix}, \quad i = 1,\ldots,9,\ s = 1,\ldots,5.$$

The log-likelihood function for the model is,

$$\begin{aligned} l(\boldsymbol{\theta}) =& \sum_{\mathbf{i}} (y_{\mathbf{i}} \log \mu_{\mathbf{i}}(\boldsymbol{\theta}) - \mu_{\mathbf{i}}(\boldsymbol{\theta})) - \sum_{\mathbf{i}} \log y_{\mathbf{i}}! \\ =& y_1(\theta) + y_2(\theta + \theta_1^X) + y_3(\theta + \theta_2^X) + y_4(\theta + \theta_1^Y) + y_5(\theta + \theta_1^X + \theta_1^Y) \\ &+ y_6(\theta + \theta_2^X + \theta_1^Y) + y_7(\theta + \theta_2^Y) + y_8(\theta + \theta_1^X + \theta_2^Y) + y_9(\theta + \theta_2^X + \theta_2^Y) \\ &- (e^{\theta} + e^{\theta+\theta_1^X} + e^{\theta+\theta_2^X} + e^{\theta+\theta_1^Y} + e^{\theta+\theta_1^X+\theta_1^Y} + e^{\theta+\theta_2^X+\theta_1^Y} + e^{\theta+\theta_2^Y} \\ &\quad + e^{\theta+\theta_1^X+\theta_2^Y} + e^{\theta+\theta_2^X+\theta_2^Y}) - \sum_{\mathbf{i}} \log y_{\mathbf{i}}! \end{aligned}$$

For the zero pattern in part (a) of Table 2.4, the $\boldsymbol{\alpha}$ and $\mathbf{U}(\boldsymbol{\theta})$ vectors are,

$$\boldsymbol{\alpha}^{\mathsf{T}} = (-1, 0, 1, 0, 1), \qquad \boldsymbol{\alpha}^{\mathsf{T}} D = \mathbf{0}, \qquad \boldsymbol{\alpha}^{\mathsf{T}} \mathbf{U}(\boldsymbol{\theta}) = -\frac{dl}{d\theta} + \frac{dl}{d\theta_2^X} + \frac{dl}{d\theta_2^Y} = 0.$$

To check the condition that $\boldsymbol{\alpha}^{\mathsf{T}} \mathbf{U}(\boldsymbol{\theta}) = 0$ for finite θs, we have,

$$\begin{aligned} \boldsymbol{\alpha}^{\mathsf{T}} \mathbf{U}(\boldsymbol{\theta}) =& -(\sum_{i=1}^{9} y_i - (e^{\theta} + e^{\theta+\theta_1^X} + e^{\theta+\theta_2^X} + e^{\theta+\theta_1^Y} + e^{\theta+\theta_1^X+\theta_1^Y} + e^{\theta+\theta_2^X+\theta_1^Y} + e^{\theta+\theta_2^Y} \\ &+ e^{\theta+\theta_1^X+\theta_2^Y} + e^{\theta+\theta_2^X+\theta_2^Y})) + (y_3 + y_6 + y_9 - e^{\theta+\theta_2^X} - e^{\theta+\theta_2^X+\theta_1^Y} - e^{\theta+\theta_2^X+\theta_2^Y}) \\ &+ (y_7 + y_8 + y_9 - e^{\theta+\theta_2^Y} - e^{\theta+\theta_1^X+\theta_2^Y} - e^{\theta+\theta_2^X+\theta_2^Y}) \\ =& -(y_3 + y_6 + y_7 + y_8) + (y_3 + y_6) + e^{\theta} + e^{\theta+\theta_1^X} + e^{\theta+\theta_1^Y} + e^{\theta+\theta_1^X+\theta_1^Y} \\ &+ (y_7 + y_8) - e^{\theta+\theta_2^X+\theta_2^Y} \\ =& e^{\theta} + e^{\theta+\theta_1^X} + e^{\theta+\theta_1^Y} + e^{\theta+\theta_1^X+\theta_1^Y} - e^{\theta+\theta_2^X+\theta_2^Y}. \end{aligned}$$

It equals to zero, if

$$e^{\theta_1^X} + e^{\theta_1^Y} + e^{\theta_1^X+\theta_1^Y} - e^{\theta_2^X+\theta_2^Y} = -1. \tag{2.9}$$

This relation among the parameters creates a flat ridge in the likelihood function which is orthogonal on some parameters. Thus, the unique MLE exists for all the parameters of this model. Without this constraint only 4 quantities given in $\boldsymbol{\theta}'^{\mathsf{T}} =$

$(\theta_1^X, \theta+\theta_2^X, \theta_1^Y, \theta+\theta_2^Y)$ are estimable. Note that equation (2.9) can be rearranged as $\theta_2^X+\theta_2^Y = \log\left(e^{\theta_1^X}+e^{\theta_1^Y}+e^{\theta_1^X+\theta_1^Y}+1\right)$ which is made of only estimable parameters. Therefore, three combinations of parameters $(\theta+\theta_2^X, \theta+\theta_2^Y, \theta_2^X+\theta_2^Y)$ are estimable. Because the number of equations and unknowns are equal, all three unknowns $(\theta, \theta_2^X, \theta_2^Y)$ are estimable along with θ_1^X and θ_1^Y in $\boldsymbol{\theta}'$, which make all the model parameters estimable. A numerical routine can estimate all model parameters for a Poisson or a multinomial sampling scheme (even though the first cell count is zero) despite not knowing (2.9) as a relation among the parameters .

For the zero pattern in part (b) of Table 2.4, there are two $\boldsymbol{\alpha}$ vectors:

$$\begin{matrix}\boldsymbol{\alpha}_1^\mathsf{T} = (-1,0,0,1,1)\\ \boldsymbol{\alpha}_2^\mathsf{T} = (-1,1,1,0,0)\end{matrix}, \quad \boldsymbol{\alpha}^\mathsf{T} D = \mathbf{0}, \quad \boldsymbol{\alpha}^\mathsf{T}\mathbf{U}(\boldsymbol{\theta}) = \begin{bmatrix} -\dfrac{dl}{d\theta}+\dfrac{dl}{d\theta_1^Y}+\dfrac{dl}{d\theta_2^Y}=0 \\ -\dfrac{dl}{d\theta}+\dfrac{dl}{d\theta_1^X}+\dfrac{dl}{d\theta_2^X}=0 \end{bmatrix}.$$

To check the condition that $\boldsymbol{\alpha}^\mathsf{T}\mathbf{U}(\boldsymbol{\theta}) = 0$ for finite θs, we have,

$$\begin{aligned}\boldsymbol{\alpha}_1^\mathsf{T}\mathbf{U}(\boldsymbol{\theta}) =& -(\sum_{i=1}^{9} y_i - (e^{\theta}+e^{\theta+\theta_1^X}+e^{\theta+\theta_2^X}+e^{\theta+\theta_1^Y}+e^{\theta+\theta_1^X+\theta_1^Y}+e^{\theta+\theta_2^X+\theta_1^Y}+e^{\theta+\theta_2^Y}\\ &+e^{\theta+\theta_1^X+\theta_2^Y}+e^{\theta+\theta_2^X+\theta_2^Y})) + (y_4+y_5+y_6-e^{\theta+\theta_1^Y}-e^{\theta+\theta_1^X+\theta_1^Y}-e^{\theta+\theta_2^X+\theta_1^Y})\\ &+(y_7+y_8+y_9-e^{\theta+\theta_2^Y}-e^{\theta+\theta_1^X+\theta_2^Y}-e^{\theta+\theta_2^X+\theta_2^Y})\\ =& e^{\theta}+e^{\theta+\theta_1^X}+e^{\theta+\theta_2^X},\end{aligned}$$

$$\begin{aligned}\boldsymbol{\alpha}_2^\mathsf{T}\mathbf{U}(\boldsymbol{\theta}) =& -(\sum_{i=1}^{9} y_i - (e^{\theta}+e^{\theta+\theta_1^X}+e^{\theta+\theta_2^X}+e^{\theta+\theta_1^Y}+e^{\theta+\theta_1^X+\theta_1^Y}+e^{\theta+\theta_2^X+\theta_1^Y}+e^{\theta+\theta_2^Y}\\ &+e^{\theta+\theta_1^X+\theta_2^Y}+e^{\theta+\theta_2^X+\theta_2^Y})) + (y_2+y_5+y_8-e^{\theta+\theta_1^X}-e^{\theta+\theta_1^X+\theta_1^Y}-e^{\theta+\theta_1^X+\theta_2^Y})\\ &+(y_3+y_6+y_9-e^{\theta+\theta_2^X}-e^{\theta+\theta_2^X+\theta_1^Y}-e^{\theta+\theta_2^X+\theta_2^Y})\\ =& e^{\theta}+e^{\theta+\theta_1^Y}+e^{\theta+\theta_2^Y}.\end{aligned}$$

Neither of them could be zero with imposing constraints on θs, so there does not exist any esoteric constraints to make all the parameters estimable. As functions appear in $\boldsymbol{\alpha}^\mathsf{T}\mathbf{U}(\boldsymbol{\theta}) = 0$ are exponential, the existence of the esoteric constraints depend on the sign of the $\boldsymbol{\alpha}$ elements. However, this model is parameter redundant and $\boldsymbol{\alpha}^\mathsf{T}\mathbf{U}(\boldsymbol{\theta}) = 0$ occurs because $\theta \to -\infty$. In other words, the intercept is not estimable and if we want to consider an estimate for that, it would be a number with a large standard deviation occurred by the flat ridge in the likelihood surface. The slope of the likelihood

surface with respect to the parameters makes the equations $-\frac{dl}{d\theta}+\frac{dl}{d\theta_1^Y}+\frac{dl}{d\theta_2^Y}=0$ and $-\frac{dl}{d\theta}+\frac{dl}{d\theta_1^X}+\frac{dl}{d\theta_2^X}=0$ hold. Thus, the only 3 estimable quantities in this model are $\boldsymbol{\theta}'^{\mathsf{T}}=(-\theta_1^X+\theta_2^X,\theta+\theta_1^X+\theta_1^Y,\theta+\theta_1^X+\theta_2^Y)$.

Chapter 3

The saturated Poisson log-linear model

3.1 Introduction

In this chapter, a saturated Poisson log-linear model is fitted to a contingency table with m variables and l levels or categories for each variable. The aim is to prove two theorems. For the first theorem, all cell counts in the table are assumed to be positive and we prove that the model is full rank in this case. Although this is already known, the process reveals the derivative matrix formation and how the described parameter redundancy method works for log-linear models. In the second theorem, we set one table cell count equal to zero and find out exactly which model parameters become inestimable in the result of that. The estimable combinations of parameters can be derived by solving the corresponding partial differential equations, but that is not the focus here. To enhance clarity with respect to the notations and the process of the proofs, we prove the theorems for a 2^m and a 3^m model first and then for an l^m model. The induction method is implemented to prove the theorems. The theorem's statement is shown to be true for a starting point, it is assumed to be true for the number of variables equals to m, then we prove it is also true for $m+1$.

Row vectors are shown with bold symbols and letters in this chapter. The fitted saturated log-linear model is model (1.5) with E as the set of all subsets of V. We set $D_{\mathsf{r}}(\boldsymbol{\theta}_{\mathsf{r}}) = \dfrac{d\boldsymbol{\mu}_{\mathsf{r}}}{d\boldsymbol{\theta}_{\mathsf{r}}}$ in which $\boldsymbol{\mu}_{\mathsf{r}}$ and $\boldsymbol{\theta}_{\mathsf{r}}$ are vectors of those cell means and parameters which are added to the model because of adding the rth variable to the table. We then define $D_{\mathsf{r}} = D_{\underline{r}}(\underline{\boldsymbol{\theta}_{\mathsf{r}}}) = \dfrac{d\underline{\boldsymbol{\mu}_{\mathsf{r}}}}{d\underline{\boldsymbol{\theta}_{\mathsf{r}}}}$ as the derivative matrix for $\underline{\boldsymbol{\mu}_{\mathsf{r}}} = \boldsymbol{\mu}_1 \cup \boldsymbol{\mu}_2 \cup \cdots \cup \boldsymbol{\mu}_{\mathsf{r}}$ and $\underline{\boldsymbol{\theta}_{\mathsf{r}}} = \boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2 \cup \cdots \cup \boldsymbol{\theta}_{\mathsf{r}}$, which are the union of elements of cell mean and model parameter vectors for having variables 1 to r. Furthermore, instead of $y_{\mathbf{i}}$ and 0 in the derivative matrix we write 1 and 0 for simplicity, and it also forms a relationship between the derivative matrix for m variables and the one for $m+1$ variables. As stated

in the last chapter, a zero cell turns a corresponding column to zero in the derivative matrix and we still apply it despite having 1 and 0 in the derivative matrix.

The next two definitions will be used in proving this chapter's theorems.

Definition 3.1. *For a saturated log-linear model we define the parameter corresponding to cell $y_i, i = 1, \ldots, n$, as the θ which has the maximum number of variables in its superscript in* $\log \mu_i = \mathsf{A}_{(i)}\boldsymbol{\theta}$*, where* $\mathsf{A}_{(i)}$ *is the ith row of the design matrix* A.

For example, in a saturated model fitted to a 3^3 contingency table, the parameter corresponding to cell y_{201} or cell y_{12} according to (1.6), is θ_{21}^{XZ}.

Definition 3.2. *For a given parameter, parameters associated with a higher order interaction are all those which could be specified by including all the additional variables in the given parameter's superscript.*

For example, in a saturated model fitted to a 3^3 contingency table, the parameters associated with a higher order interaction of parameter θ_{21}^{XZ}, are θ_{211}^{XYZ} and θ_{221}^{XYZ}.

3.2 The 2^m contingency table

In this section, a 2^m contingency table is considered. The simplest possible model would be a saturated model for the table 2^1 which has only one variable, say X, with two levels. Then the log-linear model has an intercept and one other parameter, as,

$$\log \mu_i = \theta + \theta_i^X, \qquad i \in \{0,1\}.$$

As corner point constraints are considered, we have $\theta_0^X = 0$ and θ_1^X is shown as θ^X. So the derivative matrix (2.4) is,

$$m = 1, \qquad D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = \left[\begin{array}{c|cc} & \mu_0 & \mu_1 \\ \hline \theta & 1 & 1 \\ \theta^X & 0 & 1 \end{array}\right], \qquad \boldsymbol{\theta}_1 = (\theta, \theta^X), \quad \boldsymbol{\mu}_1 = (\mu_0, \mu_1).$$

If any of the two table cell counts are observed as zero, the rank of the matrix is reduced by one and by solving (2.2) and (2.3) the parameters that are not directly estimable are obtained as,

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_0 = 0$	$\boldsymbol{\alpha}_{11} = (1, -1)$	θ, θ^X
$y_1 = 0$	$\boldsymbol{\alpha}_{12} = (0, 1)$	θ^X

For a 2^2 table with two variables X and Y, the saturated log-linear model is,

$$\log \mu_{ij} = \theta + \theta_i^X + \theta_j^Y + \theta_{ij}^{XY}, \qquad i, j \in \{0,1\}^2.$$

The model's derivative matrix is,

$$m=2, \qquad D_2 = D_{\underline{2}}(\underline{\boldsymbol{\theta}}_2) = \left[\begin{array}{c|cc:cc} & \mu_{00} & \mu_{10} & \mu_{01} & \mu_{11} \\ \hline \theta & 1 & 1 & 1 & 1 \\ \theta^X & 0 & 1 & 0 & 1 \\ \hdashline \theta^Y & 0 & 0 & 1 & 1 \\ \theta^{XY} & 0 & 0 & 0 & 1 \end{array}\right]$$

$$= \begin{bmatrix} D_{\underline{1}}(\underline{\boldsymbol{\theta}}_1) & D_2(\underline{\boldsymbol{\theta}}_1) \\ \mathbf{0} & D_2(\boldsymbol{\theta}_2) \end{bmatrix} = \begin{bmatrix} D_1 & D_2(\underline{\boldsymbol{\theta}}_1) \\ \mathbf{0} & D_1 \end{bmatrix},$$

$$\boldsymbol{\theta}_1 = (\theta, \theta^X), \qquad \boldsymbol{\theta}_2 = (\theta^Y, \theta^{XY}), \qquad \underline{\boldsymbol{\theta}}_2 = (\theta, \theta^X, \theta^Y, \theta^{XY}),$$
$$\boldsymbol{\mu}_1 = (\mu_{00}, \mu_{10}), \qquad \boldsymbol{\mu}_2 = (\mu_{01}, \mu_{11}), \qquad \underline{\boldsymbol{\mu}}_2 = (\mu_{00}, \mu_{10}, \mu_{01}, \mu_{11}).$$

If any of the four table cell counts are observed as zero, then the inestimable parameters are,

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{00} = y_1 = 0$	$\boldsymbol{\alpha}_{21} = (1,-1,-1,1) = (\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11})$	$\theta, \theta^X, \theta^Y, \theta^{XY}$
$y_{10} = y_2 = 0$	$\boldsymbol{\alpha}_{22} = (0,1,0,-1) = (\boldsymbol{\alpha}_{12}, \boldsymbol{\alpha}_{12})$	θ^X, θ^{XY}
$y_{01} = y_3 = 0$	$\boldsymbol{\alpha}_{23} = (0,0,1,-1) = (\mathbf{0}, \boldsymbol{\alpha}_{11})$	θ^Y, θ^{XY}
$y_{11} = y_4 = 0$	$\boldsymbol{\alpha}_{24} = (0,0,0,1) = (\mathbf{0}, \boldsymbol{\alpha}_{12})$	θ^{XY}

Note that we omit the sign in assigning $\boldsymbol{\alpha}_{2i}$, so $\boldsymbol{\alpha}_{21} = (\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11})$ is correct in terms of places of zero and non-zero elements which indicate the estimable and inestimable parameters. The pattern continues by enlarging the model.

For a 2^3 table with three variables X, Y and Z, the saturated log-linear model is,

$$\log \mu_{ijk} = \theta + \theta_i^X + \theta_j^Y + \theta_{ij}^{XY} + \theta_k^Z + \theta_{ik}^{XZ} + \theta_{jk}^{YZ} + \theta_{ijk}^{XYZ}, \qquad i,j,k \in \{0,1\}^3.$$

The model's derivative matrix is,

$$m=3, \quad D_3 = D_{\underline{3}}(\underline{\boldsymbol{\theta}}_3) = \left[\begin{array}{c|cccc:cccc} & \mu_{000} & \mu_{100} & \mu_{010} & \mu_{110} & \mu_{001} & \mu_{101} & \mu_{011} & \mu_{111} \\ \hline \theta & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \theta^X & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \theta^Y & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ \theta^{XY} & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ \hdashline \theta^Z & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ \theta^{XZ} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ \theta^{YZ} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \theta^{XYZ} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$= \begin{bmatrix} D_{\underline{2}}(\underline{\boldsymbol{\theta}_2}) & D_3(\underline{\boldsymbol{\theta}_2}) \\ \mathbf{0} & D_3(\boldsymbol{\theta}_3) \end{bmatrix} = \begin{bmatrix} D_2 & D_3(\underline{\boldsymbol{\theta}_2}) \\ \mathbf{0} & D_2 \end{bmatrix},$$

$$\boldsymbol{\theta}_3 = (\theta^Z, \theta^{XZ}, \theta^{YZ}, \theta^{XYZ}), \quad \underline{\boldsymbol{\theta}_3} = (\theta, \theta^X, \theta^Y, \theta^{XY}, \theta^Z, \theta^{XZ}, \theta^{YZ}, \theta^{XYZ}),$$
$$\boldsymbol{\mu}_3 = (\mu_{001}, \mu_{101}, \mu_{011}, \mu_{111}), \qquad \underline{\boldsymbol{\mu}_3} = (\mu_{000}, \mu_{100}, \mu_{001}, \mu_{110}, \mu_{001}, \mu_{101}, \mu_{011}, \mu_{111}).$$

If any of the eight table cell counts are observed as zero, then the inestimable parameters are,

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{000} = 0$	$\boldsymbol{\alpha}_{31} = (-1,1,1,-1,1,-1,-1,1) = (\boldsymbol{\alpha}_{21}, \boldsymbol{\alpha}_{21})$	$\theta, \theta^X, \theta^Y, \theta^{XY}, \theta^Z, \theta^{XZ}, \theta^{YZ}, \theta^{XYZ}$
$y_{100} = 0$	$\boldsymbol{\alpha}_{32} = (0,1,0,-1,0,-1,0,1) = (\boldsymbol{\alpha}_{22}, \boldsymbol{\alpha}_{22})$	$\theta^X, \theta^{XY}, \theta^{XZ}, \theta^{XYZ}$
$y_{010} = 0$	$\boldsymbol{\alpha}_{33} = (0,0,1,-1,0,0,-1,1) = (\boldsymbol{\alpha}_{23}, \boldsymbol{\alpha}_{23})$	$\theta^Y, \theta^{XY}, \theta^{YZ}, \theta^{XYZ}$
$y_{110} = 0$	$\boldsymbol{\alpha}_{34} = (0,0,0,-1,0,0,0,1) = (\boldsymbol{\alpha}_{24}, \boldsymbol{\alpha}_{24})$	$\theta^{XY}, \theta^{XYZ}$
$y_{001} = 0$	$\boldsymbol{\alpha}_{35} = (0,0,0,0,1,-1,-1,1) = (\mathbf{0}, \boldsymbol{\alpha}_{21})$	$\theta^Z, \theta^{XZ}, \theta^{YZ}, \theta^{XYZ}$
$y_{101} = 0$	$\boldsymbol{\alpha}_{36} = (0,0,0,0,0,-1,0,1) = (\mathbf{0}, \boldsymbol{\alpha}_{22})$	$\theta^{XZ}, \theta^{XYZ}$
$y_{011} = 0$	$\boldsymbol{\alpha}_{37} = (0,0,0,0,0,0,-1,1) = (\mathbf{0}, \boldsymbol{\alpha}_{23})$	$\theta^{YZ}, \theta^{XYZ}$
$y_{111} = 0$	$\boldsymbol{\alpha}_{38} = (0,0,0,0,0,0,0,1) = (\mathbf{0}, \boldsymbol{\alpha}_{24})$	θ^{XYZ}

By increasing m, the number of variables in the model, the observed pattern continues among $\boldsymbol{\alpha}$ vectors, D matrices and inestimable parameters.

Although it is known that a log-linear model with all positive data is a full rank model and all of its parameters are estimable and it is not structurally parameter redundant, we prove this in the next theorem by checking the rank of the derivative matrices. The contingency table is assumed to have m variables each categorized in two levels, so $L = \otimes_{j=1}^{m} [l_j]$ such that $[l_j] = \{0, 1\}$.

Theorem 3.1. *A saturated Poisson log-linear model for a 2^m contingency table ($m \geq 1$) is full rank, when $y_{\boldsymbol{i}} > 0, \forall \boldsymbol{i} \in L$.*

Proof. For the simplest possible model with $m = 1$ which has only an intercept and one other parameter, the derivative matrix is,

$$D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\theta}_1 = (\theta, \theta^X).$$

When all $y_{\mathbf{i}} > 0$, no matrix column is zero so the derivative matrix rank is 2 and the model is full rank. As $m = 1$ can be considered too trivial regarding contingency tables,

we show that the theorem statement is also true for $m = 2$ with variables X and Y.

$$D_2 = D_{\underline{2}}(\underline{\boldsymbol{\theta}_2}) = \left[\begin{array}{cc:cc} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ \hdashline 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right] = \begin{bmatrix} D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) & D_2(\underline{\boldsymbol{\theta}_1}) \\ \mathbf{0} & D_2(\boldsymbol{\theta}_2) \end{bmatrix} = \begin{bmatrix} D_1 & D_2(\underline{\boldsymbol{\theta}_1}) \\ \mathbf{0} & D_1 \end{bmatrix},$$

$$\boldsymbol{\theta}_1 = (\theta, \theta^X), \qquad \boldsymbol{\theta}_2 = (\theta^Y, \theta^{XY}), \qquad \underline{\boldsymbol{\theta}_2} = (\theta, \theta^X, \theta^Y, \theta^{XY}).$$

When all $y_{\mathbf{i}} > 0$, no matrix column is zero so the derivative matrix rank is 4 and this model is full rank. The derivative matrix for $m = k$ variables takes the form D_k and is assumed to be full rank if no observation is zero,

$$D_k = D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) = \begin{bmatrix} D_{\underline{k-1}}(\underline{\boldsymbol{\theta}_{k-1}}) & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix} = \begin{bmatrix} D_{k-1} & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_{k-1} \end{bmatrix}.$$

Now, we must prove that the derivative matrix is full rank for $m = k+1$. According to the pattern among the derivative matrices, D_{k+1} is made of D_k, as,

$$D_{k+1} = D_{\underline{k+1}}(\underline{\boldsymbol{\theta}_{k+1}}) = \begin{bmatrix} D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix} = \begin{bmatrix} D_k & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_k \end{bmatrix}.$$

D_k was assumed to be full rank, so according to Theorem 2.5 the matrix D_{k+1} is also full rank. □

In the next theorem, induction method is used again to prove that by having one zero cell count in the 2^m contingency table with a saturated Poisson log-linear model fitted to it, the parameter corresponding to the zero cell and given it, all other parameters associated with a higher order interaction, according to Definitions 3.1 and 3.2, become inestimable. Estimable combinations of parameters can easily be derived by solving the corresponding partial differential equations (2.3).

Theorem 3.2. *In a saturated Poisson log-linear model for a 2^m contingency table ($m \geq 1$), if $\exists \boldsymbol{i}, \boldsymbol{i} \in L$ such that $y_{\boldsymbol{i}} = 0$, then the corresponding parameter to that cell and all other parameters associated with a higher order interaction given that parameter, are inestimable.*

Proof. The theorem statement is true for the simplest possible model with $m = 1$.

$$D_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\theta}_1 = (\theta, \theta^X).$$

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_0 = 0$	$\boldsymbol{\alpha}_{11} = (1, -1)$	θ, θ^X
$y_1 = 0$	$\boldsymbol{\alpha}_{12} = (0, 1)$	θ^X

If $y_1 = 0$, then θ^X is inestimable and there is no parameter associated with a higher order interaction given it. If $y_0 = 0$, then θ, the parameter corresponding to the first cell is inestimable along with θ^X, which is associated with a higher order interaction given the intercept. In each case, there is only one $\boldsymbol{\alpha}$ since only one column has turned to zero and the rank of the derivative matrix is decreased by one, so $d = 1$.

We assume the theorem statement is true for the model with $m = k$. The derivative matrix, made of the derivative matrix with one fewer variable, is,

$$D_k = \begin{bmatrix} D_{k-1} & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_{k-1} \end{bmatrix}.$$

Consider those k variables as $X, Y, \ldots, U, W$. When each cell count $y_{\mathbf{i}}, \mathbf{i} \in \{0,1\}^k$ is zero, the parameters stated in the following table are assumed to be inestimable. It also indicates the $\boldsymbol{\alpha}$ vectors because there exists only a unique one in each case.

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{0\ldots0} = 0$	$\boldsymbol{\alpha}_{k1} = (\boldsymbol{\alpha}_{(k-1)1}, \boldsymbol{\alpha}_{(k-1)1})$	$\theta, \theta^X, \theta^Y, \ldots, \theta^W, \theta^{XW}, \ldots, \theta^{XY\ldots UW}$
$\vdots$	$\vdots$	$\vdots$
$y_{11\ldots10} = 0$	$\boldsymbol{\alpha}_{k2^{k-1}} = (\boldsymbol{\alpha}_{(k-1)2^{k-1}}, \boldsymbol{\alpha}_{(k-1)2^{k-1}})$	$\theta^{XY\ldots U}, \theta^{XY\ldots UW}$
$y_{00\ldots01} = 0$	$\boldsymbol{\alpha}_{k(2^{k-1}+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{k1})$	$\theta^W, \theta^{XW}, \theta^{YW}, \ldots, \theta^{XY\ldots UW}$
$\vdots$	$\vdots$	$\vdots$
$y_{1\ldots1} = 0$	$\boldsymbol{\alpha}_{k2^k} = (\mathbf{0}, \boldsymbol{\alpha}_{(k-1)2^{k-1}})$	$\theta^{XY\ldots UW}$

Now the theorem statement must be proven for the model with $m = k+1$ variables. Assume the added variable is Q. The derivative matrix is,

$$D_{k+1} = \begin{bmatrix} D_k & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_k \end{bmatrix},$$

and the inestimable parameters should be,

zero cell	inestimable parameters
$y_{0\ldots0}=0$	$\theta,\theta^X,\ldots,\theta^Q,\theta^{XQ},\ldots,\theta^{XY\ldots WQ}$
$\vdots$	$\vdots$
$y_{11\ldots10}=0$	$\theta^{XY\ldots UW},\theta^{XY\ldots WQ}$
$y_{00\ldots01}=0$	$\theta^Q,\theta^{XQ},\theta^{YQ},\ldots,\theta^{X\ldots WQ}$
$\vdots$	$\vdots$
$y_{1\ldots1}=0$	$\theta^{XY\ldots WQ}$

In order to prove that these are inestimable parameters, obtaining the corresponding $\boldsymbol{\alpha}$ vectors are required. We observed a repetitive pattern of $\boldsymbol{\alpha}$ vectors in making the derivative matrices and increasing the number of variables in the contingency table. According to that pattern, $\boldsymbol{\alpha}$s are made of vectors of the previous step. Therefore the unique $\boldsymbol{\alpha}$ vectors are,

zero cell	$\boldsymbol{\alpha}$ vector
$y_{0\ldots0}=0$	$\boldsymbol{\alpha}_{(k+1)1}=(\boldsymbol{\alpha}_{k1},\boldsymbol{\alpha}_{k1})$
$\vdots$	$\vdots$
$y_{11\ldots10}=0$	$\boldsymbol{\alpha}_{(k+1)2^k}=(\boldsymbol{\alpha}_{k2^k},\boldsymbol{\alpha}_{k2^k})$
$y_{00\ldots01}=0$	$\boldsymbol{\alpha}_{(k+1)(2^k+1)}=(\mathbf{0},\boldsymbol{\alpha}_{k1})$
$\vdots$	$\vdots$
$y_{1\ldots1}=0$	$\boldsymbol{\alpha}_{(k+1)2^{k+1}}=(\mathbf{0},\boldsymbol{\alpha}_{k2^k})$

For the first half of the cases in the above table, having a zero cell count gives $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_{ki},\boldsymbol{\alpha}_{ki})$. Since the theorem is assumed to be true for $m=k$, $\boldsymbol{\alpha}_{ki}$ makes the corresponding parameter to that cell and all other higher order interaction parameters to be inestimable for the last smaller model. Repeating $\boldsymbol{\alpha}_{ki}$ in the $\boldsymbol{\alpha}$ vector shows some other parameters of the new model as inestimable as well, which are the same previous parameters but corresponding to the new variable in this model, say Q. For example, if we had θ^X before as an inestimable parameter, now θ^{XQ} is added to the inestimable parameters set as well. As a result, the corresponding parameter to the zero cell count and all other parameters associated with a higher order interaction given that parameter, are inestimable.

For the second half of the cases, having a zero cell count gives $\boldsymbol{\alpha}=(\mathbf{0},\boldsymbol{\alpha}_{ki})$. $\boldsymbol{\alpha}_{ki}$ shows the corresponding parameter to that cell and all other parameters associated with a higher order interaction are inestimable for the last smaller model, but as it appears after a vector of zeroes here, those parameters will have the new variable, say Q, in their

superscript. In conclusion, the theorem statement is true for $m = k+1$ and the theorem is proven by induction. □

3.3 The 3^m contingency table

In this section, a 3^m contingency table is considered. The simplest possible model would be a saturated model for the table 3^1 which has only one variable, say X, with three levels. Then the log-linear model has an intercept and two other parameters, as,

$$\log \mu_i = \theta + \theta_i^X, \qquad i \in \{0,1,2\}.$$

So the derivative matrix (2.4) is,

$$m = 1, \quad D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = \left[\begin{array}{c|ccc} & \mu_0 & \mu_1 & \mu_2 \\ \hline \theta & 1 & 1 & 1 \\ \theta_1^X & 0 & 1 & 0 \\ \theta_2^X & 0 & 0 & 1 \end{array}\right], \quad \boldsymbol{\theta}_1 = (\theta, \theta_1^X, \theta_2^X),\ \boldsymbol{\mu}_1 = (\mu_0, \mu_1, \mu_2).$$

If any of the three table cell counts are observed as zero, the rank of the matrix is reduced by one and by solving (2.2) and (2.3) the inestimable parameters are obtained as,

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_0 = 0$	$\boldsymbol{\alpha}_{11} = (-1,1,1)$	$\theta, \theta_1^X, \theta_2^X$
$y_1 = 0$	$\boldsymbol{\alpha}_{12} = (0,1,0)$	θ_1^X
$y_2 = 0$	$\boldsymbol{\alpha}_{12} = (0,0,1)$	θ_2^X

For a 3^2 table with two variables X and Y, the saturated log-linear model is,

$$\log \mu_{ij} = \theta + \theta_i^X + \theta_j^Y + \theta_{ij}^{XY}, \qquad i,j \in \{0,1,2\}^2.$$

The model's derivative matrix is,

$$m = 2, \qquad D_2 = D_{\underline{2}}(\underline{\boldsymbol{\theta}_2}) = \left[\begin{array}{c|ccc:ccc:ccc} & \text{Y=0} & & & \text{Y=1} & & & \text{Y=2} & & \\ & \mu_{00} & \mu_{10} & \mu_{20} & \mu_{01} & \mu_{11} & \mu_{21} & \mu_{02} & \mu_{12} & \mu_{22} \\ \hline \theta & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ \theta_1^X & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ \theta_2^X & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ \hdashline \theta_1^Y & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ \theta_{11}^{XY} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \theta_{21}^{XY} & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \theta_2^Y & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ \theta_{12}^{XY} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \theta_{22}^{XY} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$= \begin{bmatrix} D_1(\underline{\boldsymbol{\theta}_1}) & D_2(\underline{\boldsymbol{\theta}_1}) \\ \mathbf{0} & D_2(\boldsymbol{\theta}_2) \end{bmatrix} = \begin{bmatrix} D_1 & D_2(\underline{\boldsymbol{\theta}_1}) \\ \mathbf{0} & D_2(\boldsymbol{\theta}_2) \end{bmatrix},$$

such that,

$$D_2(\boldsymbol{\theta}_2) = \begin{bmatrix} D_1 & \mathbf{0} \\ \mathbf{0} & D_1 \end{bmatrix},$$

and,

$$\boldsymbol{\theta}_1 = (\theta, \theta_1^X, \theta_2^X), \qquad \boldsymbol{\theta}_2 = (\theta_1^Y, \theta_{11}^{XY}, \theta_{21}^{XY}, \theta_2^Y, \theta_{12}^{XY}, \theta_{22}^{XY}), \qquad \underline{\boldsymbol{\theta}_2} = (\boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2),$$
$$\boldsymbol{\mu}_1 = (\mu_{00}, \mu_{10}, \mu_{20}), \qquad \boldsymbol{\mu}_2 = (\mu_{01}, \mu_{11}, \mu_{21}, \mu_{02}, \mu_{12}, \mu_{22}), \qquad \underline{\boldsymbol{\mu}_2} = (\boldsymbol{\mu}_1 \cup \boldsymbol{\mu}_2).$$

If any of the nine table cell counts are observed as zero, the inestimable parameters are,

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{00} = 0$	$\boldsymbol{\alpha}_{21} = (1,-1,-1,-1,1,1,-1,1,1) = (\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11})$	$\theta, \theta_1^X, \theta_2^X, \theta_1^Y, \theta_{11}^{XY}, \theta_{21}^{XY}, \theta_2^Y, \theta_{12}^{XY}, \theta_{22}^{XY}$
$y_{10} = 0$	$\boldsymbol{\alpha}_{22} = (0,-1,0,0,1,0,0,1,0) = (\boldsymbol{\alpha}_{12}, \boldsymbol{\alpha}_{12}, \boldsymbol{\alpha}_{12})$	$\theta_1^X, \theta_{11}^{XY}, \theta_{12}^{XY}$
$y_{20} = 0$	$\boldsymbol{\alpha}_{23} = (0,0,-1,0,0,1,0,0,1) = (\boldsymbol{\alpha}_{13}, \boldsymbol{\alpha}_{13}, \boldsymbol{\alpha}_{13})$	$\theta_2^X, \theta_{21}^{XY}, \theta_{22}^{XY}$
$y_{01} = 0$	$\boldsymbol{\alpha}_{24} = (0,0,0,-1,1,1,0,0,0) = (\mathbf{0}, \boldsymbol{\alpha}_{11}, \mathbf{0})$	$\theta_1^Y, \theta_{11}^{XY}, \theta_{21}^{XY}$
$y_{11} = 0$	$\boldsymbol{\alpha}_{25} = (0,0,0,0,1,0,0,0,0) = (\mathbf{0}, \boldsymbol{\alpha}_{12}, \mathbf{0})$	θ_{11}^{XY}
$y_{21} = 0$	$\boldsymbol{\alpha}_{26} = (0,0,0,0,0,1,0,0,0) = (\mathbf{0}, \boldsymbol{\alpha}_{13}, \mathbf{0})$	θ_{21}^{XY}
$y_{02} = 0$	$\boldsymbol{\alpha}_{27} = (0,0,0,0,0,0,-1,1,1) = (\mathbf{0}, \mathbf{0}, \boldsymbol{\alpha}_{11})$	$\theta_2^Y, \theta_{12}^{XY}, \theta_{22}^{XY}$
$y_{12} = 0$	$\boldsymbol{\alpha}_{28} = (0,0,0,0,0,0,0,1,0) = (\mathbf{0}, \mathbf{0}, \boldsymbol{\alpha}_{12})$	θ_{12}^{XY}
$y_{22} = 0$	$\boldsymbol{\alpha}_{29} = (0,0,0,0,0,0,0,0,1) = (\mathbf{0}, \mathbf{0}, \boldsymbol{\alpha}_{13})$	θ_{22}^{XY}

$\boldsymbol{\alpha}_{21} = (\boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11}, \boldsymbol{\alpha}_{11})$ is correct in terms of places of zero and non-zero elements which indicate the estimable and inestimable parameters. The pattern continues by enlarging the model.

For a 3^3 table with three variables X, Y and Z, the saturated log-linear model is,

$$\log \mu_{ijk} = \theta + \theta_i^X + \theta_j^Y + \theta_{ij}^{XY} + \theta_k^Z + \theta_{ik}^{XZ} + \theta_{jk}^{YZ} + \theta_{ijk}^{XYZ}, \qquad i,j,k \in \{0,1,2\}^3.$$

The model's derivative matrix is,

$$m = 3, \qquad D_3 = D_{\underline{3}}(\underline{\boldsymbol{\theta}_3}) = \begin{bmatrix} D_{\underline{2}}(\underline{\boldsymbol{\theta}_2}) & D_3(\underline{\boldsymbol{\theta}_2}) \\ \mathbf{0} & D_3(\boldsymbol{\theta}_3) \end{bmatrix} = \begin{bmatrix} D_2 & D_3(\underline{\boldsymbol{\theta}_2}) \\ \mathbf{0} & D_3(\boldsymbol{\theta}_3) \end{bmatrix},$$

such that,

$$D_3(\boldsymbol{\theta}_3) = \begin{bmatrix} D_2 & \mathbf{0} \\ \mathbf{0} & D_2 \end{bmatrix},$$

and,

$$\begin{aligned}\boldsymbol{\theta}_3 =&(\theta_1^Z, \theta_2^Z, \theta_{11}^{XZ}, \theta_{21}^{XZ}, \theta_{11}^{YZ}, \theta_{21}^{YZ}, \theta_{12}^{XZ}, \theta_{22}^{XZ}, \theta_{12}^{YZ}, \theta_{22}^{YZ}, \theta_{111}^{XYZ}, \theta_{211}^{XYZ}, \theta_{121}^{XYZ}, \theta_{221}^{XYZ},\\ &\theta_{112}^{XYZ}, \theta_{212}^{XYZ}, \theta_{122}^{XYZ}, \theta_{222}^{XYZ}),\\ \underline{\boldsymbol{\theta}_3} =&(\boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2 \cup \boldsymbol{\theta}_3).\end{aligned}$$

By increasing m, the number of variables in the model, the observed pattern continues among $\boldsymbol{\alpha}$ vectors, D matrices and inestimable parameters.

In the next theorem, we prove that a saturated log-linear model is full rank when all the observations are positive, by checking the rank of the derivative matrices. The contingency table is assumed to have m variables categorized in three levels, so $L = \otimes_{j=1}^m [l_j]$ and $[l_j] = \{0,1,2\}$.

Theorem 3.3. *A saturated Poisson log-linear model for a 3^m contingency table ($m \geq 1$) is full rank, when $y_{\boldsymbol{i}} > 0, \forall \boldsymbol{i} \in L$.*

Proof. For the simplest possible model with $m = 1$ which has only an intercept and one other parameter, the derivative matrix is,

$$D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\theta}_1 = (\theta, \theta_1^X, \theta_2^X).$$

When all $y_{\mathbf{i}} > 0$, no matrix column is zero so the derivative matrix rank is 3 and the model is full rank. As $m = 1$ can be considered too trivial for a contingency table, we show that the theorem statement is also true for $m = 2$ with variables X and Y.

$$D_2 = D_{\underline{2}}(\underline{\boldsymbol{\theta}_2}) = \begin{bmatrix} D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) & D_2(\underline{\boldsymbol{\theta}_1}) \\ \mathbf{0} & D_2(\boldsymbol{\theta}_2) \end{bmatrix} = \begin{bmatrix} D_1 & D_2(\underline{\boldsymbol{\theta}_1}) \\ \mathbf{0} & D_2(\boldsymbol{\theta}_2) \end{bmatrix},$$

such that,

$$D_2(\boldsymbol{\theta}_2) = \begin{bmatrix} D_1 & \mathbf{0} \\ \mathbf{0} & D_1 \end{bmatrix},$$

and,

$$\boldsymbol{\theta}_2 = (\theta_1^Y, \theta_{11}^{XY}, \theta_{21}^{XY}, \theta_2^Y, \theta_{12}^{XY}, \theta_{22}^{XY}), \qquad \underline{\boldsymbol{\theta}_2} = (\boldsymbol{\theta}_1 \cup \boldsymbol{\theta}_2).$$

When all $y_{\mathbf{i}} > 0$, no matrix column is zero so the derivative matrix rank is 9 and this model is full rank. The derivative matrix for $m = k$ variables takes the form D_k and is assumed to be full rank if no observation is zero.

$$D_k = D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) = \begin{bmatrix} D_{\underline{k-1}}(\underline{\boldsymbol{\theta}_{k-1}}) & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix} = \begin{bmatrix} D_{k-1} & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix},$$

such that,

$$D_k(\boldsymbol{\theta}_k) = \begin{bmatrix} D_{k-1} & \mathbf{0} \\ \mathbf{0} & D_{k-1} \end{bmatrix}.$$

Now, we must prove that the derivative matrix is full rank for $m = k+1$. According to the pattern among the derivative matrices, D_{k+1} is made of D_k, as,

$$D_{k+1} = D_{\underline{k+1}}(\underline{\boldsymbol{\theta}_{k+1}}) = \begin{bmatrix} D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix} = \begin{bmatrix} D_k & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix},$$

so that,

$$D_{k+1}(\boldsymbol{\theta}_{k+1}) = \begin{bmatrix} D_k & \mathbf{0} \\ \mathbf{0} & D_k \end{bmatrix}.$$

D_k is assumed to be full rank so $D_{k+1}(\boldsymbol{\theta}_{k+1})$ is full rank as well. Then according to Theorem 2.5, the matrix D_{k+1} is also full rank. □

In the next theorem, we prove the same statement as given in Theorem 3.2 but for a Poisson saturated model fitted to a 3^m contingency table.

Theorem 3.4. *In a saturated Poisson log-linear model for a 3^m contingency table ($m \geq 1$), if $\exists \boldsymbol{i}, \boldsymbol{i} \in L$ such that $y_{\boldsymbol{i}} = 0$, then the corresponding parameter to that cell and all other parameters associated with a higher order interaction given that parameter, are inestimable.*

Proof. The theorem statement is true for the simplest possible model with $m = 1$.

$$D_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\theta}_1 = (\theta, \theta_1^X, \theta_2^X).$$

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_0 = 0$	$\boldsymbol{\alpha}_{11} = (-1,1,1)$	$\theta, \theta_1^X, \theta_2^X$
$y_1 = 0$	$\boldsymbol{\alpha}_{12} = (0,1,0)$	θ_1^X
$y_2 = 0$	$\boldsymbol{\alpha}_{12} = (0,0,1)$	θ_2^X

If $y_1 = 0$, then θ_1^X is inestimable and there is no parameter associated with a higher order interaction given it. If $y_0 = 0$, then θ, the parameter corresponding to the first cell, is inestimable along with θ_1^X and θ_2^X which are parameters with a higher order interaction given the intercept. In each case, there is only one $\boldsymbol{\alpha}$ since only one column has turned to zero and the rank of the derivative matrix is decreased by one, so $d = 1$.

We assume the theorem statement is true for the model with $m = k$. The derivative matrix, made of the derivative matrix with one fewer variable, is

$$D_k = \begin{bmatrix} D_{k-1} & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix},$$

such that,

$$D_k(\boldsymbol{\theta}_k) = \begin{bmatrix} D_{k-1} & \mathbf{0} \\ \mathbf{0} & D_{k-1} \end{bmatrix}.$$

Consider those k variables as $X, Y, \ldots, U, W$. When each cell count $y_{\mathbf{i}}, \mathbf{i} \in \{0,1,2\}^k$ is zero, the parameters stated in the following table are assumed to be inestimable. It also indicates the $\boldsymbol{\alpha}$ vectors, as there exists only a unique one in each case.

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{00\ldots00} = 0$	$\boldsymbol{\alpha}_{k1} = (\boldsymbol{\alpha}_{(k-1)1}, \boldsymbol{\alpha}_{(k-1)1}, \boldsymbol{\alpha}_{(k-1)1})$	$\theta, \theta_1^X, \theta_2^X, \ldots, \theta_{22\ldots22}^{XY\ldots UW}$
$\vdots$	$\vdots$	$\vdots$
$y_{22\ldots20} = 0$	$\boldsymbol{\alpha}_{k3^{k-1}} = (\boldsymbol{\alpha}_{(k-1)3^{k-1}}, \boldsymbol{\alpha}_{(k-1)3^{k-1}}, \boldsymbol{\alpha}_{(k-1)3^{k-1}})$	$\theta_{22\ldots2}^{XY\ldots U}, \theta_{22\ldots21}^{XY\ldots UW}, \theta_{22\ldots22}^{XY dots UW}$
$y_{00\ldots01} = 0$	$\boldsymbol{\alpha}_{k(3^{k-1}+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{(k-1)1}, \mathbf{0})$	$\theta_1^W, \theta_{11}^{XW}, \theta_{21}^{XW}, \ldots, \theta_{22\ldots21}^{XY\ldots UW}$
$\vdots$	$\vdots$	$\vdots$
$y_{22\ldots21} = 0$	$\boldsymbol{\alpha}_{k3^{k-1}\times 2} = (\mathbf{0}, \boldsymbol{\alpha}_{(k-1)3^{k-1}}, \mathbf{0})$	$\theta_{22\ldots21}^{XY\ldots UW}$
$y_{00\ldots02} = 0$	$\boldsymbol{\alpha}_{k(3^{k-1}\times 2+1)} = (\mathbf{0}, \mathbf{0}, \boldsymbol{\alpha}_{(k-1)1})$	$\theta_2^W, \theta_{12}^{XW}, \theta_{22}^{XW}, \ldots, \theta_{22\ldots22}^{XY\ldots UW}$
$\vdots$	$\vdots$	$\vdots$
$y_{22\ldots22} = 0$	$\boldsymbol{\alpha}_{k3^k} = (\mathbf{0}, \mathbf{0}, \boldsymbol{\alpha}_{(k-1)3^{k-1}})$	$\theta_{22\ldots22}^{XY\ldots UW}$

Now the theorem statement must be proven for the model with $m = k+1$ variables. If the added variable is Q, then the derivative matrix is,

$$D_{k+1} = \begin{bmatrix} D_k & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix},$$

such that,

$$D_{k+1}(\boldsymbol{\theta}_{k+1}) = \begin{bmatrix} D_k & \mathbf{0} \\ \mathbf{0} & D_k \end{bmatrix}.$$

The inestimable parameters should be,

zero cell	inestimable parameters
$y_{00\ldots00}=0$	$\theta, \theta_1^X, \ldots, \theta_{22\ldots22}^{XY\ldots WQ}$
$\vdots$	$\vdots$
$y_{22\ldots20}=0$	$\theta_{22\ldots2}^{XY\ldots UW}, \theta_{22,\ldots,21}^{XY\ldots WQ}, \theta_{22,\ldots,22}^{XY\ldots WQ}$
$y_{00\ldots01}=0$	$\theta_1^Q, \theta_{11}^{XQ}, \theta_{21}^{XQ}, \ldots, \theta_{22\ldots21}^{XY\ldots WQ}$
$\vdots$	$\vdots$
$y_{22\ldots21}=0$	$\theta_{22\ldots21}^{XY\ldots WQ}$
$y_{00\ldots02}=0$	$\theta_2^Q, \theta_{12}^{XQ}, \theta_{21}^{XQ}, \ldots, \theta_{22\ldots22}^{XY\ldots WQ}$
$\vdots$	$\vdots$
$y_{22\ldots22}=0$	$\theta_{22\ldots22}^{XY\ldots WQ}$

The corresponding $\boldsymbol{\alpha}$ vectors must be obtained to prove that these are the inestimable parameters. We observed a repetitive pattern of $\boldsymbol{\alpha}$ vectors in making the derivative matrices and increasing the number of variables in the contingency table. According to the pattern, $\boldsymbol{\alpha}$s are made of vectors of the previous step. Therefore the unique $\boldsymbol{\alpha}$ vectors are,

zero cell	$\boldsymbol{\alpha}$ vector
$y_{00\ldots00}=0$	$\boldsymbol{\alpha}_{(k+1)1}=(\boldsymbol{\alpha}_{k1},\boldsymbol{\alpha}_{k1},\boldsymbol{\alpha}_{k1})$
$\vdots$	$\vdots$
$y_{22\ldots20}=0$	$\boldsymbol{\alpha}_{(k+1)3^k}=(\boldsymbol{\alpha}_{k3^k},\boldsymbol{\alpha}_{k3^k},\boldsymbol{\alpha}_{k3^k})$
$y_{00\ldots01}=0$	$\boldsymbol{\alpha}_{(k+1)(3^k+1)}=(\mathbf{0},\boldsymbol{\alpha}_{k1},\mathbf{0})$
$\vdots$	$\vdots$
$y_{22\ldots21}=0$	$\boldsymbol{\alpha}_{(k+1)3^k\times2}=(\mathbf{0},\boldsymbol{\alpha}_{k3^k},\mathbf{0})$
$y_{00\ldots02}=0$	$\boldsymbol{\alpha}_{(k+1)(3^k\times2+1)}=(\mathbf{0},\mathbf{0},\boldsymbol{\alpha}_{k1})$
$\vdots$	$\vdots$
$y_{22\ldots22}=0$	$\boldsymbol{\alpha}_{(k+1)3^{k+1}}=(\mathbf{0},\mathbf{0},\boldsymbol{\alpha}_{k3^k})$

For the first one third of the cases in the above table, having a zero cell count gives $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_{ki},\boldsymbol{\alpha}_{ki},\boldsymbol{\alpha}_{ki})$. Since the theorem is assumed to be true for $m=k$, the first $\boldsymbol{\alpha}_{ki}$ shows the corresponding parameter to that cell and given that, all other parameters associated with a higher order interaction are inestimable for the last smaller model. Repeating $\boldsymbol{\alpha}_{ki}$ two times in the $\boldsymbol{\alpha}$ vector shows some other parameters of the new model

are inestimable too, which are the same previous parameters corresponding to the both levels of the new variable, Q. For example if we had θ_1^X before, now $\theta_{11}^{XQ}, \theta_{12}^{XQ}$ are added to the inestimable parameters set as well. In result, the corresponding parameter to the zero cell count and given that, all other parameters associated with a higher order interaction are inestimable.

For the next two third of the cases, having a zero cell count gives $\boldsymbol{\alpha} = (\mathbf{0}, \boldsymbol{\alpha}_{ki}, \mathbf{0})$ or $\boldsymbol{\alpha} = (\mathbf{0}, \mathbf{0}, \boldsymbol{\alpha}_{ki})$. $\boldsymbol{\alpha}_{ki}$ shows the corresponding parameter to that cell and given that, all other parameters associated with a higher order interaction are inestimable for the last smaller model, but as it appears after one or two vectors of zeroes here, those parameters will have the first or second level of the new variable, say Q, in their superscript and subscript. In conclusion, the statement is true for $m = k+1$, so the theorem is proven by induction.

□

3.4 The l^m contingency table

The theorems in this section differ from the previous ones as an l^m contingency table is considered here, so neither the number of variables, m, nor the number of categories for each variable, l, are fixed. A different version of induction method is adopted here to prove the theorems. Earl [2003] suggests following ways to deal with the induction for two variables. If we consider $P(\mathsf{n}, \mathsf{k})$ for $\mathsf{n}, \mathsf{k} = 1, 2, 3, \ldots$, as the general statement with two variables, then induction method can be used to prove it in one of these two ways:

- Prove $P(1,1)$ and show $P(\mathsf{n}+1, \mathsf{k})$ and $P(\mathsf{n}, \mathsf{k}+1)$ both follow from $P(\mathsf{n}, \mathsf{k})$ for all n and k.

- Prove $P(1, \mathsf{k})$ for all k and show how knowledge of $P(\mathsf{n}, \mathsf{k})$ for all k proves that $P(\mathsf{n}+1, \mathsf{k})$ holds for all k. This reduces the problem to one application of induction for a family of statements. Or we could do the inducting part through k and consider n as the arbitrary variable. So we prove $P(\mathsf{n}, 1)$ for all n and show how knowing $P(\mathsf{n}, \mathsf{k})$ for all n, leads to proving $P(\mathsf{n}, \mathsf{k}+1)$ for all n.

In the next theorem, we prove that a saturated log-linear model is full rank when all the observations are positive, by checking the rank of the derivative matrices. The contingency table is assumed to have m variables categorized in l levels, so $L = \otimes_{j=1}^{m} [l_j]$ and $[l_j] = \{0, 1, \ldots, l-1\}$. We also provide a descriptive proof for the theorem in addition to the induction based proof.

Theorem 3.5. *A saturated Poisson log-linear model for an l^m contingency table ($m \geq 1, l \geq 2$) is full rank, when $y_{\boldsymbol{i}} > 0, \forall \boldsymbol{i} \in L$.*

Proof. The derivative matrix elements are derivatives of logarithm of cell means with respect to the parameters. We can always consider an order of cell means and their corresponding parameters that form the D matrix as an upper triangular matrix with all main diagonal elements equal to 1 (like all of the derivative matrices in this chapter). By implying corner point constraints in the model there is one parameter corresponding to each cell according to Definition 3.1 and derivations with respect to them make a vector of 1s in the diagonal of the derivative matrix. The rank of an upper triangular matrix is the number of none zero elements on the main diagonal. So in the case of having no zero cell counts, the D matrix is always full rank. Since this does not depend on the value of parameters $\boldsymbol{\theta}$, the model is essentially full rank. □

In order to prove the same theorem by induction, we use the second method described above on dealing with induction for two variables. The method has two steps. On the first step, we prove that the statement is true for an l^1 table for all integers $l \geq 2$. On the second step, we show that when the statement is assumed true for an l^m table, it is also true for l^{m+1}.

Proof. **Step one:** We prove that the statement is true for the saturated model of an l^1 table, for all integers $l \geq 2$. Consider the only variable in this contingency table is X with $[l] = \{0, 1, \ldots, l-1\}$ levels, therefore the saturated model includes l parameters. The derivative matrix for this model is,

$$D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = \left[\begin{array}{c|cccccc} & \mu_0 & \mu_1 & \mu_2 & \mu_3 & & \mu_{l-1} \\ \hline \theta & 1 & 1 & 1 & 1 & \ldots & 1 \\ \theta_1^X & 0 & 1 & 0 & 0 & \ldots & 0 \\ \theta_2^X & 0 & 0 & 1 & 0 & \ldots & 0 \\ \theta_3^X & 0 & 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_{l-1}^X & 0 & 0 & 0 & 0 & \ldots & 1 \end{array}\right].$$

When all cell counts are positive then the elements on the main diagonal of this matrix are always none zero, so the rank of the matrix equals to the number of parameters of the model and it is full rank.

Step two: The statement is assumed to be true for the saturated model for an l^m table when $m = k$, it should be true when $m = k+1$ as well. For $m = k$ the derivative

matrix which is assumed to be full rank, is,

$$D_k = D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) = \begin{bmatrix} D_{\underline{k-1}}(\underline{\boldsymbol{\theta}_{k-1}}) & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix} = \begin{bmatrix} D_{k-1} & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix},$$

such that,

$$D_k(\boldsymbol{\theta}_k) = \begin{bmatrix} D_{k-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & D_{k-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & D_{k-1} \end{bmatrix}_{l-1\times l-1}.$$

Now, we must prove that the derivative matrix is full rank for $m = k+1$. The derivative matrix in this case is,

$$D_{k+1} = D_{\underline{k+1}}(\underline{\boldsymbol{\theta}_{k+1}}) = \begin{bmatrix} D_{\underline{k}}(\underline{\boldsymbol{\theta}_k}) & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix} = \begin{bmatrix} D_k & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix},$$

such that,

$$D_{k+1}(\boldsymbol{\theta}_{k+1}) = \begin{bmatrix} D_k & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & D_k & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & D_k \end{bmatrix}_{l-1\times l-1}.$$

In this matrix, D_k is an $l^{m-1} \times l^{m-1}$ matrix and assumed to be full rank, so matrix $D_{k+1}(\boldsymbol{\theta}_{k+1})$ is full rank. It follows that D_{k+1} is also full rank according to Theorem 2.5. □

Catchpole and Morgan [1997] proved a relevant theorem which is a special case of Theorem 2.5. "Suppose that a product multinomial model with parameter vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ is full rank for an $r \times c$ table. Let $r' \geq r$ and $c' \geq c$. Suppose that for an $r' \times c'$ experiment, the extension of the table by one row leads to the inclusion of extra parameters $\boldsymbol{\psi} = (\theta_{p'+1}, \dots, \theta_{p'+v})$. Regard this extra row as a function of $\boldsymbol{\psi}$ only, and form its derivative matrix. Now repeat this procedure for an extension by one column. If both of these subsidiary derivative matrices are full rank, then the model is full rank for any $r' \times c'$ table" [Catchpole and Morgan, 1997, Theorem 6]. The theorem states that when the model for a table with only two variables is full rank and adding more levels to those variables makes a full rank subsidiary derivative matrix, then the model for the bigger table is full rank. In order to prove this, the model is assumed to be full rank for an $r \times c$ table with derivative matrix D. After the addition of a row or column,

the derivative matrix is,

$$\mathscr{D} = \begin{bmatrix} D & B \\ 0 & C \end{bmatrix}.$$

The columns of B and C correspond to the extra cell means (or cell probabilities for a multinomial model), and the rows of C correspond to the extra parameters. The matrix C is the subsidiary derivative matrix created due to adding a row or a column to the model and it is formed only based on the extra cell means and the extra parameters. If D and C are both full rank, it follows immediately that $\mathscr{D}$ is full rank. The result follows by induction, therefore the model remains of full rank after adding a row or a column.

In Theorem 3.5, we showed that a saturated model for an l^m table is full rank when all cell counts are positive. In general, adding a variable to the table or adding a level to all table variables result in having a subsidiary derivative matrix akin to C in the main derivative matrix $\mathscr{D}$ above. If this subsidiary derivative matrix is full rank then the general derivative matrix is also full rank.

In the next theorem, we prove the same statement as given in Theorems 3.2 and 3.4, but for a Poisson saturated model fitted to an l^m contingency table. We prove that the theorem's statement is true for an l^1 table, for all integers $l \geq 2$. Then we show that if the statement is true for l^m, then it is true for l^{m+1}.

Theorem 3.6. *In a saturated Poisson log-linear model for an l^m contingency table ($m \geq 1, l \geq 2$), if $\exists \boldsymbol{i}, \boldsymbol{i} \in L$ such that $y_{\boldsymbol{i}} = 0$, then the corresponding parameter to that cell and all other parameters associated with a higher order interaction given that parameter, are inestimable.*

Proof. **Step one:** We prove that the statement is true for l^1 for all integers $l \geq 2$. Assume the only variable in this model is X with $[l] = \{0, 1, \ldots, l-1\}$ levels, therefore the saturated model includes l parameters. The derivative matrix for this model is,

$$D_1 = D_{\underline{1}}(\underline{\boldsymbol{\theta}_1}) = \left[\begin{array}{c|cccccc} & \mu_0 & \mu_1 & \mu_2 & \mu_3 & & \mu_{l-1} \\ \hline \theta & 1 & 1 & 1 & 1 & \ldots & 1 \\ \theta_1^X & 0 & 1 & 0 & 0 & \ldots & 0 \\ \theta_2^X & 0 & 0 & 1 & 0 & \ldots & 0 \\ \theta_3^X & 0 & 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_{l-1}^X & 0 & 0 & 0 & 0 & \ldots & 1 \end{array}\right].$$

For this model, $\boldsymbol{\alpha}$ vectors and the inestimable parameters in the presence of zero cell counts are shown here. Since only one cell count is zero, deficiency is one and there is

one α vector for each case.

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_0 = 0$	$\boldsymbol{\alpha}_{11} = (1,-1,-1,-1,\ldots,1)$	all parameters
$y_1 = 0$	$\boldsymbol{\alpha}_{12} = (0,1,0,0,\ldots,0)$	θ_1^X
$y_2 = 0$	$\boldsymbol{\alpha}_{13} = (0,0,1,0,\ldots,0)$	θ_2^X
$y_3 = 0$	$\boldsymbol{\alpha}_{14} = (0,0,0,1,\ldots,0)$	θ_3^X
$\vdots$	$\vdots$	$\vdots$
$y_{l-1} = 0$	$\boldsymbol{\alpha}_{1l} = (0,0,0,\ldots,0,1)$	θ_{l-1}^X

According to the $\boldsymbol{\alpha}$ vectors, the theorem statement is true for this model.

In order to further clarify the idea, we can fix the number of variables at $m = 2$ and show that the statement is still true for this model with any number of levels. So let the variables in this model to be X, Y with $[l] = \{0, 1, \ldots, l-1\}$ levels. The derivative matrix $D_2 = D_{\underline{2}}(\underline{\boldsymbol{\theta}_2})$ for the model fitted to the l^2 table is,

$$
\left[\begin{array}{c|ccccc|ccccc|c|ccccc}
 & & & Y=0 & & & & & Y=1 & & & & & & Y=l-1 & & \\
 & \mu_{00} & \mu_{10} & \mu_{20} & \cdots & \mu_{l-10} & \mu_{01} & \mu_{11} & \mu_{21} & \cdots & \mu_{l-11} & \cdots & \mu_{0l-1} & \mu_{1l-1} & \mu_{2l-1} & \cdots & \mu_{l-1l-1} \\
\hline
\theta & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & \cdots & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \\
\theta_1^X & 0 & 1 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\theta_2^X & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_{l-1}^X & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 & 0 & \cdots & 1 \\
\hline
\theta_1^Y & 0 & 0 & 0 & \cdots & 0 & 1 & 1 & 1 & \cdots & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\theta_{11}^{XY} & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\theta_{21}^{XY} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_{l-11}^{XY} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\
\hline
\vdots & & & \vdots & & & & & \vdots & & & \vdots & & & \vdots & & \\
\hline
\theta_{l-1}^Y & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & 1 & 1 & \cdots & 1 \\
\theta_{1l-1}^{XY} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\theta_{2l-1}^{XY} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\theta_{l-1l-1}^{XY} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & \cdots & 1
\end{array}\right]
$$

$$
= \begin{bmatrix} D_1 & D_1 & \ldots & D_1 \\ \mathbf{0} & D_1 & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & D_1 \end{bmatrix}.
$$

The derivative matrix is upper triangular and all elements on the main diagonal are 1. Consider $y_{\mathbf{i}(0)}$ as a cell count that its index is finished with zero, we can order these

cell counts also from 1 to l^m according to (1.6). $\boldsymbol{\gamma}_{\mathbf{i}}$ is the set including inestimable parameters. Thus in case of having one zero cell count, the inestimable parameters and unique $\boldsymbol{\alpha}$ vectors are given in the below table which match the theorem's statement.

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\alpha}_{21} = \overbrace{(\boldsymbol{\alpha}_{11}, \ldots, \boldsymbol{\alpha}_{11})}^{\#l}$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_1 = \{\text{all parameters}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_l = 0$	$\boldsymbol{\alpha}_{2l} = (\boldsymbol{\alpha}_{1l}, \ldots, \boldsymbol{\alpha}_{1l})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_l = \{\theta^X_{l-1}, \theta^{XY}_{l-11}, \ldots, \theta^{XY}_{l-1l-1}\}$
$y_{\mathbf{i}(1)} = y_{l+1} = 0$	$\boldsymbol{\alpha}_{2(l+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{11}, \mathbf{0}, \ldots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l+1} = \{\theta^Y_1, \theta^{XY}_{11}, \ldots, \theta^{XY}_{l-11}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{2\times l} = 0$	$\boldsymbol{\alpha}_{2(2\times l)} = (\mathbf{0}, \boldsymbol{\alpha}_{1l}, \mathbf{0}, \ldots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{2\times l} = \{\theta^{XY}_{l-11}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^2-l+1} = 0$	$\boldsymbol{\alpha}_{2(l^2-l+1)} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{11})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^2-l+1} = \{\theta^Y_{l-1}, \theta^{XY}_{1l-1}, \ldots, \theta^{XY}_{l-1l-11}\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^2} = 0$	$\boldsymbol{\alpha}_{2l^2} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{1l})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^2} = \{\theta^{XY}_{l-1l-1}\}$

Step two: The statement is assumed to be true for l^m when $m = k$, it should be true when $m = k+1$ as well. For $m = k$ when any of the cell counts is zero, the corresponding parameter to that cell and given that, all parameters associated with a higher order interaction of variables are assumed to be inestimable. The derivative matrix is,

$$D_k = \begin{bmatrix} D_{k-1} & D_k(\underline{\boldsymbol{\theta}_{k-1}}) \\ \mathbf{0} & D_k(\boldsymbol{\theta}_k) \end{bmatrix}.$$

Derivative matrices are upper triangular and all elements on their main diagonals are 1. Consider $y_{\mathbf{i}(0)}$ as a cell count such that the final element of $\mathbf{i}$ is zero. $\boldsymbol{\gamma}_{\mathbf{i}}$ is the set including the parameter corresponding to that cell and given that, all parameters with a higher order interaction of variables. The order of setting cell counts to zero here is the same order that used in forming the derivative matrix. Thus, the inestimable parameters should be as follows (so are $\boldsymbol{\alpha}$ vectors, due to the repetitive pattern in models and the

point that in each case there is only one $\boldsymbol{\alpha}$ vector),

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\alpha}_{k1} = \overbrace{(\boldsymbol{\alpha}_{(k-1)1}, \ldots, \boldsymbol{\alpha}_{(k-1)1})}^{\#l}$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_1 = \{$all parameters$\}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_{l^{k-1}} = 0$	$\boldsymbol{\alpha}_{kl^{k-1}} = (\boldsymbol{\alpha}_{(k-1)l^{k-1}}, \ldots, \boldsymbol{\alpha}_{(k-1)l^{k-1}})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k-1}}$
$y_{\mathbf{i}(1)} = y_{l^{k-1}+1} = 0$	$\boldsymbol{\alpha}_{k(l^{k-1}+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{(k-1)1}, \mathbf{0}, \ldots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k-1}+1}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{l^{k-1}\times 2} = 0$	$\boldsymbol{\alpha}_{k(l^{k-1}\times 2)} = (\mathbf{0}, \boldsymbol{\alpha}_{(k-1)l^{k-1}}, \mathbf{0}, \ldots, \mathbf{0})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k-1}\times 2}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{(l^{k-1}\times l-1)+1} = 0$	$\boldsymbol{\alpha}_{k((l^{k-1}\times l-1)+1)} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{(k-1)1})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{(l^{k-1}\times l-1)+1}$
$\vdots$	$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^k} = 0$	$\boldsymbol{\alpha}_{kl^k} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{(k-1)l^{k-1}})$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k} = \{$only the highest order parameter$\}$

Now the theorem statement must be proven for $m = k+1$, while the derivative matrix is,

$$D_{k+1} = \begin{bmatrix} D_k & D_{k+1}(\underline{\boldsymbol{\theta}_k}) \\ \mathbf{0} & D_{k+1}(\boldsymbol{\theta}_{k+1}) \end{bmatrix}.$$

The inestimable parameters should be,

zero cell	inestimable parameters
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_1 = \{$all parameters$\}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_{l^k} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k}$
$y_{\mathbf{i}(1)} = y_{l^k+1} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k+1}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{l^k\times 2} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^k\times 2}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{(l^k\times l-1)+1} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{(l^k\times l-1)+1}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^{k+1}} = 0$	$\boldsymbol{\gamma}_{\mathbf{i}} = \boldsymbol{\gamma}_{l^{k+1}} = \{$only the highest order parameter$\}$

In order to prove that these are inestimable parameters, we need to obtain the corresponding $\boldsymbol{\alpha}$ vectors. According to the repetitive pattern of $\boldsymbol{\alpha}$ vectors that was observed before, $\boldsymbol{\alpha}$s are made of vectors of the previous step. Therefore the unique $\boldsymbol{\alpha}$ vectors are,

zero cell	inestimable parameters
$y_{\mathbf{i}(0)} = y_1 = 0$	$\boldsymbol{\alpha}_{(k+1)1} = \overbrace{(\boldsymbol{\alpha}_{k1}, \ldots, \boldsymbol{\alpha}_{k1})}^{\#l}$
$\vdots$	$\vdots$
$y_{\mathbf{i}(0)} = y_{l^k} = 0$	$\boldsymbol{\alpha}_{(k+1)l^k} = (\boldsymbol{\alpha}_{kl^k}, \ldots, \boldsymbol{\alpha}_{kl^k})$
$y_{\mathbf{i}(1)} = y_{l^k+1} = 0$	$\boldsymbol{\alpha}_{(k+1)(l^k+1)} = (\mathbf{0}, \boldsymbol{\alpha}_{k1}, \mathbf{0}, \ldots, \mathbf{0})$
$\vdots$	$\vdots$
$y_{\mathbf{i}(1)} = y_{l^k \times 2} = 0$	$\boldsymbol{\alpha}_{(k+1)(l^k \times 2)} = (\mathbf{0}, \boldsymbol{\alpha}_{kl^k}, \mathbf{0}, \ldots, \mathbf{0})$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{(l^k \times l-1)+1} = 0$	$\boldsymbol{\alpha}_{(k+1)((l^k \times l-1)+1)} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{k1})$
$\vdots$	$\vdots$
$y_{\mathbf{i}(l-1)} = y_{l^{k+1}} = 0$	$\boldsymbol{\alpha}_{(k+1)l^{k+1}} = (\mathbf{0}, \mathbf{0}, \ldots, \boldsymbol{\alpha}_{kl^k})$

For the first $\frac{1}{l}$ of the cases in the above table, having a zero cell count gives $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{ki}, \ldots, \boldsymbol{\alpha}_{ki})$. Since the theorem is assumed to be true for $m = k$, the first $\boldsymbol{\alpha}_{ki}$ shows the corresponding parameter to that cell and given that, all parameters associated with a higher order interaction of variables are inestimable for the last smaller model ($m = k$). Repeating $\boldsymbol{\alpha}_{ki}$, $l - 1$ times in the $\boldsymbol{\alpha}$ vector shows some other parameters of the new model are inestimable as well, which are the same previous parameters corresponding to all levels of the new variable. As a result, the parameters stated in the theorem are inestimable.

For the other cases in the above table, having a zero cell count makes an $\boldsymbol{\alpha}_{ki}$ happen in the vector. $\boldsymbol{\alpha}_{ki}$ shows the corresponding parameter to that cell and given that, all parameters associated with a higher order interaction of variables are inestimable for the last smaller model, but as it appears after one or more vectors of zeroes here, those parameters will have the higher levels of the new variable in their superscript and subscript. In conclusion, the statement is true for $m = k + 1$ and the theorem is proved by induction.

□

$\boldsymbol{\alpha}$ vectors and D matrices

In all the previous examples for saturated log-linear models with one zero cell count in their contingency tables, we observe that when a cell count is zero then the $\boldsymbol{\alpha}$ vector (which is the null space for the transpose of the derivative matrix containing zero columns) matches the corresponding row of the main derivative matrix in terms of the position of zero and non-zero elements. This is Theorem 3.6 in different words. Because when each matrix row is the $\boldsymbol{\alpha}$ vector of setting the corresponding column to zero, then it has non-zero elements for the corresponding parameter to that cell and given that, all parameters associated with a higher order interaction of variables.

For instance, in a saturated log-linear model for a 2^2 table, $\boldsymbol{\alpha}$ vectors and inestimable parameters for setting each cell equal to zero are,

$$D = \left[\begin{array}{c|cc:cc} & \mu_{00} & \mu_{10} & \mu_{01} & \mu_{11} \\ \hline \theta & 1 & 1 & 1 & 1 \\ \theta^X & 0 & 1 & 0 & 1 \\ \hdashline \theta^Y & 0 & 0 & 1 & 1 \\ \theta^{XY} & 0 & 0 & 0 & 1 \end{array}\right],$$

zero cell	$\boldsymbol{\alpha}$ vector	inestimable parameters
$y_{00} = y_1 = 0$	$\boldsymbol{\alpha}_{21} = (1,-1,-1,1) =$ 1st row of D	$\theta, \theta^X, \theta^Y, \theta^{XY}$
$y_{10} = y_2 = 0$	$\boldsymbol{\alpha}_{22} = (0,1,0,-1) =$ 2nd row of D	θ^X, θ^{XY}
$y_{01} = y_3 = 0$	$\boldsymbol{\alpha}_{23} = (0,0,1,-1) =$ 3rd row of D	θ^Y, θ^{XY}
$y_{11} = y_4 = 0$	$\boldsymbol{\alpha}_{24} = (0,0,0,1) =$ 4th row of D	θ^{XY}

More than one zero cell

In fitting the saturated log-linear model to an l^m contingency table, adding more zero cells do not alter any previous inestimable parameter to be an estimable one. So the set of the model's inestimable parameters is the union of the inestimable parameters caused by each zero cell. That set then is the set of the corresponding parameters to all zero cells and given them, all parameters associated to a higher order interaction of variables. Meanwhile, the estimable combinations of the parameters for the model with more than one zero entry could be derived by solving the corresponding partial differential equations (2.3).

Chapter 4

Existence of the MLE and comparison with parameter redundancy

4.1 Maximum likelihood estimation in log-linear models

The prominent purpose of our work so far, has been detecting parameter redundancy and the estimability of $\theta^e(\mathbf{i})$ for a log-linear model defined in (1.5), as

$$m_{\mathbf{i}} = \log \mu_{\mathbf{i}} = \sum_{e \in \mathsf{E}} \theta^e(\mathbf{i}),$$

which also leads to finding the estimable $\mu_{\mathbf{i}}$. In this chapter, we review the alternative approach that seeks the estimable and inestimable cell means of the log-linear model in the presence of zero cell counts by investigating the existence of the MLE for $\mu_{\mathbf{i}}$. We refer to the approach described in Sections 4.1.1 and 4.1.2 as the existence of the MLE (EMLE) method. In Section 4.2, this approach and its results are compared with the parameter redundancy method. Section 4.3 discusses a specific kind of models according to these two methods.

Maximum likelihood estimates of the expected values of cells in a contingency table play a prominent role in the model selection and goodness of fit test [Bishop, 1975]. When sampling zeros occur in the contingency table, the maximum likelihood estimate of some cell means may not exist. The conditions for the existence of the MLEs of the cell means in log-linear models were studied by Birch [1963], Haberman [1973] and Bishop [1975].

Assume the saturated log-linear model for a three-way table with three variables X (rows), Y (columns) and Z (layers). Then the log-likelihood function (1.7) for this

model is,

$$
\begin{aligned}
l(\boldsymbol{\mu}) =& \sum_i \sum_j \sum_k (y_{ijk} \log \mu_{ijk} - \mu_{ijk}) \\
=& y_{+++}\theta + \sum_i y_{i++}\theta_i^X + \sum_j y_{+j+}\theta_j^Y + \sum_k y_{++k}\theta_k^Z + \sum_i \sum_j y_{ij+}\theta_{ij}^{XY} + \sum_i \sum_k y_{i+k}\theta_{ik}^{XZ} \\
&+ \sum_j \sum_k y_{+jk}\theta_{jk}^{YZ} + \sum_i \sum_j \sum_k y_{ijk}\theta_{ijk}^{XYZ} - \sum_i \sum_j \sum_k \exp(\theta + \cdots + \theta_{ijk}^{XYZ}).
\end{aligned}
$$

y_{ijk}s are sufficient statistics for θ_{ijk}^{XYZ} and maximum likelihood estimates for cell means. Estimates of the model parameters do not exist when any $y_{ijk} = 0$ and they are finite only when all cell counts are positive [Agresti, 2002].

As a non-saturated model, consider the same three-way table with the fitted hierarchical model (*XY, XZ, YZ*) which includes the main effects of the variables and first order interactions between them. The corresponding likelihood function is,

$$
\begin{aligned}
l(\boldsymbol{\mu}) =& \sum_i \sum_j \sum_k (y_{ijk} \log \mu_{ijk} - \mu_{ijk}) \\
=& y_{+++}\theta + \sum_i y_{i++}\theta_i^X + \sum_j y_{+j+}\theta_j^Y + \sum_k y_{++k}\theta_k^Z + \sum_i \sum_j y_{ij+}\theta_{ij}^{XY} \\
&+ \sum_i \sum_k y_{i+k}\theta_{ik}^{XZ} + \sum_j \sum_k y_{+jk}\theta_{jk}^{YZ} - \mu_{+++}.
\end{aligned}
$$

Marginal totals $y_{ij+}, y_{i+j}, y_{+jk}$ are maximum likelihood estimates of their expectations [Birch, 1963] and form a complete minimal sufficient statistic set under the Poisson model [Haberman, 1973]. Therefore, cell mean estimates do not exist when any marginal is zero in the set of sufficient marginals [Agresti, 2002]. For example if y_{ij+} is zero then an infinite estimate occurs for θ^{XY} and maximum likelihood estimate of $\mu_{ij+} = \mu_{ij0} + \mu_{ij1} + \ldots$ is zero which is not admissible in a log-linear model. Table 4.1 from Agresti [2002] shows minimal sufficient statistics for different hierarchical log-linear models for a three-way contingency table.

Model	Minimal sufficient statistics
(X,Y,Z)	$y_{i++}, y_{+j+}, y_{++k}$
(XY,Z)	y_{ij+}, y_{++k}
(XY,YZ)	y_{ij+}, y_{i+k}
(XY,XZ,YZ)	$y_{ij+}, y_{i+k}, y_{+jk}$

Table 4.1 Minimal sufficient statistics for log-linear models in a three-way table.

For log-linear models with an explicit formula for $\hat{\mu}_{\mathbf{i}}$ as a function of the observations, positivity of sufficient table marginals is a necessary and sufficient condition for

the existence of the MLE of $\boldsymbol{\mu}$ [Agresti, 2002]. These models are named **decomposable models** and can be interpreted in terms of the relationship among the variables, such as independence and conditional independence [Goodman, 1970, 1971]. For non-decomposable models, $\hat{\mu}_{\mathbf{i}}$ is obtained by iterative methods and positivity of sufficient table marginals is still a necessary condition for the existence of MLE, but it is no longer sufficient. Table 4.2 from Agresti [2002] shows the formula for finding the estimates of cell means for some different hierarchical models fitted to a three-way table. As it is shown, the formula for cell mean estimates of a model with no three order interaction is not known and numerical methods should be used to obtain the estimates.

Model	Probabilistic form	Estimated mean value
(X,Y,Z)	$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi++k$	$\hat{\mu}_{ijk} = \frac{y_{i++}y_{+j+}y_{++k}}{N^2}$
(XY,Z)	$\pi_{ijk} = \pi_{ij+}\pi_{++k}$	$\hat{\mu}_{ijk} = \frac{y_{ij+}y_{++k}}{N}$
(XY,XZ)	$\pi_{ijk} = \frac{\pi_{ij+}\pi i+k}{\pi_{i++}}$	$\hat{\mu}_{ijk} = \frac{y_{ij+}y_{i+k}}{y_{i++}}$
(XY,XZ,YZ)	$\pi_{ijk} = \psi_{ij}\phi_{jk}\omega_{ik}$	Iterative methods
(XYZ)	No restriction	$\hat{\mu}_{ijk} = y_{ijk}$

Table 4.2 Estimated mean values for log-linear models in a three-way table.

Note that knowing $\hat{\mu}_{\mathbf{i}} = y_{\mathbf{i}}$ for a saturated model explains the result stated in Theorem 3.6. When one $\hat{\mu}_{\mathbf{i}} = y_{\mathbf{i}} = 0$, the logarithm of that cell mean tends to minus infinity along with the parameter corresponding to that cell, while all other parameters in that equation have already been estimated. When this parameter appears in the other equations of the log-linear model, which all have finite $\hat{\mu}_{\mathbf{i}}$, only its summation with parameters with the higher order interaction of variables are estimable, not any of those parameters alone.

4.1.1 Haberman's Theorem

Linear manifolds are considered as a general way to describe the log-linear models either from Poisson or multinomial sampling by Haberman [1973]. Assume $\mathscr{M}$ is a p-dimensional linear manifold contained in $\mathbb{R}^L$ and $P_{\mathscr{M}}$ is the orthogonal projection from $\mathbb{R}^L$ to $\mathscr{M}$. The next theorem concerns the MLE of $\mathbf{m}$ in (1.5).

Theorem 4.1. *If an MLE $\hat{\boldsymbol{m}}$ exists, then it is unique and satisfies the equation*

$$P_{\mathscr{M}}\hat{\boldsymbol{\mu}} = P_{\mathscr{M}}\boldsymbol{y}.$$

Conversely, if for some $\hat{\boldsymbol{m}} \in \mathscr{M}$ and $\hat{\boldsymbol{\mu}} = e^{\hat{\boldsymbol{m}}}$ the equation is satisfied, then $\hat{\boldsymbol{m}}$ is the MLE of $\boldsymbol{m}$. [Theorem 3.1, Haberman, 1973]

The theorem gives the MLE for the cell means of the contingency table. For example, in a saturated model each cell count is the MLE for the corresponding cell mean and for

the unsaturated models, certain marginal totals make the MLE. But no condition has been given for the existence of the MLE so far. The next theorem provides a necessary and sufficient condition for the existence of the MLE of $\mathbf{m}$ in (1.5) for Poisson and multinomial models, regardless of the presence of positive or zero table marginals. Define $\mathscr{M}^{\perp}$ as the following set,

$$\mathscr{M}^{\perp} = \left\{ \mathbf{x} \in \mathbb{R}^{|L|} : (\mathbf{x}, \mathbf{m}) = \mathbf{x}^{\mathsf{T}}\mathbf{m} = 0, \forall \mathbf{m} \in \mathscr{M} \right\}.$$

Theorem 4.2. *A necessary and sufficient condition that the MLE $\hat{\boldsymbol{m}}$ of $\boldsymbol{m}$ exists is that there exist $\boldsymbol{\delta} \in \mathscr{M}^{\perp}$ such that $y_{\boldsymbol{i}} + \delta_{\boldsymbol{i}} > 0$ for every $\boldsymbol{i} \in L$.* [Theorem 3.2, Haberman, 1973]

In this theorem, $\boldsymbol{\mu}$ in $\mathbf{m} = \log \boldsymbol{\mu}$ is assumed to be positive. The theorem specifies whether the MLE of all the cell means exists or not for any pattern of zeros in the table. The following examples from Haberman [1973] show the significance of this theorem.

Example 4.1. Consider fitting log-linear model (4.1) below to the contingency Table 4.3 with three variables categorized in two levels. This model is a hierarchical no-second-order interaction model and can be shown as (*XY, XZ, YZ*) . If the table contained only one zero cell, the model would be full rank. But there are two zero cells in the table, with all other cells assumed to have positive observations.

$$\log \mu_{ijk} = \theta + \theta_i^X + \theta_j^Y + \theta_k^Z + \theta_{ij}^{XY} + \theta_{ik}^{XZ} + \theta_{jk}^{YZ}, \qquad i,j,k \in \{0,1\}^2. \tag{4.1}$$

In the matrix form, it is shown as,

$$\begin{bmatrix} m_1 = \log \mu_{000} \\ m_2 = \log \mu_{100} \\ m_3 = \log \mu_{010} \\ m_4 = \log \mu_{110} \\ m_5 = \log \mu_{001} \\ m_6 = \log \mu_{101} \\ m_7 = \log \mu_{011} \\ m_8 = \log \mu_{111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \theta \\ \theta^X \\ \theta^Y \\ \theta^{XY} \\ \theta^Z \\ \theta^{XZ} \\ \theta^{YZ} \end{bmatrix}.$$

None of the marginal totals $y_{ij+}, y_{i+j}, y_{+jk}$ are zero in this example and the model is not decomposable. According to Theorem 4.2, the MLE exists if and only if there is such an $\boldsymbol{\delta}$ vector. The $\boldsymbol{\delta}$ vector could be,

$$\boldsymbol{\delta}^{\mathsf{T}} = (+\delta, -\delta, -\delta, +\delta, -\delta, +\delta, +\delta, -\delta),$$

$Z = 0$

	Y	
X	0	1
0	0	y_3
1	y_2	y_4

$Z = 1$

	Y	
X	0	1
0	y_5	y_7
1	y_6	0

Table 4.3 Observations in a 2^3 contingency table.

as $\boldsymbol{\delta}^{\mathsf{T}}\mathbf{m}$ equals to zero,

$$\begin{aligned}
&\delta m_1 - \delta m_2 - \delta m_3 + \delta m_4 - \delta m_5 + \delta m_6 + \delta m_7 - \delta m_8 \\
=&\delta\theta - \delta\theta - \delta\theta^X - \delta\theta - \delta\theta^Y + \delta\theta + \delta\theta^X + \delta\theta^Y + \delta\theta^{XY} - \delta\theta - \delta\theta^Z \\
&+ \delta\theta + \delta\theta^X + \delta\theta^Z + \delta\theta^{XZ} + \delta\theta + \delta\theta^Y + \delta\theta^Z + \delta\theta^{YZ} \\
&- \delta\theta - \delta\theta^X - \delta\theta^Y - \delta\theta^Z - \delta\theta^{XY} - \delta\theta^{XZ} - \delta\theta^{YZ} = 0.
\end{aligned}$$

But even with assuming $0 < \delta < 1$, $y_i + \delta_i$ is not positive for $i = 8$. The estimated cell means are shown in Table 4.4. If there is a positive estimate δ in the first cell, then the other mean estimates are determined by that δ, which imposes a negative estimate in the last cell. It is clearly impossible to find a δ that yields a positive estimate for the mean of both zero cells, unless it is $\delta = 0$ which means the fitted values are the same as the observations [Bishop, 1975]. Thus, there is not such a $\boldsymbol{\delta}$ as defined in the theorem and the MLE of $\mathbf{m}$ does not exist.

$Z = 0$

	Y	
X	0	1
0	$+\delta$	$y_3 - \delta$
1	$y_2 - \delta$	$y_4 + \delta$

$Z = 1$

	Y	
X	0	1
0	$y_5 - \delta$	$y_7 + \delta$
1	$y_6 + \delta$	$-\delta$

Table 4.4 The estimated cell means.

This example was initially introduced as a "pathological example" since none of the sufficient marginals of the table is zero and the MLE of $\boldsymbol{\mu}$ does not exist [Haberman, 1973]. However, Theorem 4.2 clarifies the reason for this non-existent MLE. Although the sufficient statistics are positive, they impose constraints that make some cell mean estimates (those cells with zero observation) equal to zero. Non-existent MLE is reported in computational packages by failure in convergence or large standard errors for some estimates. More examples of sparse contingency tables with positive margins

but non-existent MLE for the cell means are provided by Fienberg and Rinaldo [2007]. They prove that for a 2^k contingency table and the model of no-$(k-1)$ order interaction, the probability that two randomly placed sampling zeros cause the non-existent MLE without inducing zero margins is,

$$\frac{2^{k-1}-k}{2^k-1}. \tag{4.2}$$

The next example fits the same model (4.1) to a different pattern of zeros in the 2^3 table.

Example 4.2. Consider fitting model (4.1) to the pattern of zeros in Table 4.5. According to Theorem 4.2, the positive MLE for all the cell means in this model exists, as the possible $\boldsymbol{\delta}$ is,

$$\boldsymbol{\delta}^{\mathsf{T}} = (+\delta, -\delta, -\delta, +\delta, -\delta, +\delta, +\delta, -\delta).$$

$\boldsymbol{\delta}^{\mathsf{T}}\mathbf{m}$ equals to zero and by assuming $0 < \delta < 1$ we have $y_i + \delta_i > 0$ even if the positive cell counts are as small as 1. Adding and subtracting these appropriate amounts to the observed data eliminate the zero cells in the table. So the estimated cell means are as shown in Table 4.6 and the MLE of $\mathbf{m}$ does exist. This is the positive vector of values which provides the same vector of margins in (4.3).

$Z = 0$

	Y	
X	0	1
0	0	y_3
1	y_2	0

$Z = 1$

	Y	
X	0	1
0	y_5	y_7
1	y_6	y_8

Table 4.5 Observations in a 2^3 contingency table.

$Z = 0$

	Y	
X	0	1
0	$+\delta$	$y_3 - \delta$
1	$y_2 - \delta$	$+\delta$

$Z = 1$

	Y	
X	0	1
0	$y_5 - \delta$	$y_7 + \delta$
1	$y_6 + \delta$	$y_8 - \delta$

Table 4.6 The estimated cell means.

The extended MLE

Theorem 4.2 assumes that $\boldsymbol{\mu}$ can only be positive and it gives a result that either the MLE exists for all the cell means or not. For model (1.5), we have,

$$\boldsymbol{\mu} = \exp\{\mathbf{m}\}, \qquad or \qquad \log \boldsymbol{\mu} = \mathbf{m}, \qquad \mathbf{m} = \mathsf{A}\boldsymbol{\theta}.$$

This re-parametrization is not able to cover the whole range of μ. μ as the mean of a Poisson variable may take value of zero, but there is no equivalent for that in the log expression. For example, assume that we have only one observation from the Poisson distribution, which happens to be zero. The corresponding probability function is,

$$P(Y_1 = 0) = e^{-\mu}, \qquad \mu \geq 0.$$

This function clearly has a maximum, which happens at $\mu = 0$, so this point is the MLE for cell mean or $\hat{\mu}_{MLE} = 0$. But there is no equivalent for this estimate in the sense of $\log \boldsymbol{\mu} = \mathbf{m}$ as the logarithm of zero is not defined [Baker and Clarke, 1985]. Theorem 4.2 does not allow for a zero estimate for Poisson means as the parameter space is defined for $\boldsymbol{\mu} > 0$ to accommodate for the log expression. To cover such cases, Haberman [1974] introduced the "Extended MLE" for μ which could be zero as well and cancels the restriction previously set for the parameter space. So the MLE for $\boldsymbol{\mu}$ always exists and maximizes the likelihood function, although in some cases there is $\hat{\mu}_{\mathbf{i}} = 0$ and the logarithmic re-parametrization of $\mu_{\mathbf{i}}$ is not defined [Baker and Clarke, 1985]. Feinberg and Rinaldo [2012a] indicates the extended MLE is always well defined in the "extended exponential family of distributions", in which, for example, the parameter of Poisson distribution is a non-negative integer.

This concept enables us to proceed with the models that according to Theorem 4.2 do not have the existent MLE for $\boldsymbol{\mu}$. Examples of such a model could be a saturated log-linear model with at least one zero observation or the model presented in Example 4.1. After detecting the cells in which $\hat{\mu}_{\mathbf{i}} = 0$ and respectively some $\theta \to -\infty$, we can still find the estimates for the other cell means and model parameters.

4.1.2 The polyhedral method

A necessary and sufficient condition for the existence of the unique MLE of the vector of cell means was provided in Theorem 4.2, assuming that each cell mean estimate could only be positive. A geometric and polyhedral equivalent of that theorem to detect the existence of the MLE of cell means for a hierarchical log-linear model is presented by Eriksson et al. [2006]. This new approach gives a simpler way and an algorithm to

check whether the MLE of the cell means exists or not. Detailed definitions and more information on polytopes are provided by Lauritzen [1996], Ziegler [1995] and Gentle [2007].

Consider $\mathscr{M}$ is a p-dimensional log-linear subspace for the log-linear model and A is the corresponding $|L| \times p$ design matrix which is not required to be full rank [Fienberg and Rinaldo, 2005], but for simplicity we assume it as a full rank matrix with rank p [Fienberg and Rinaldo, 2012b]. The Polyhedral cone generated by spanning columns of the matrix A is,

$$C_{\mathsf{A}} = \{\mathbf{t} : \mathbf{t} = \mathsf{A}^{\mathsf{T}}\mathbf{y}, \mathbf{y} \in \mathbb{R}^{|L|}_{\geq 0}\}, \tag{4.3}$$

which is called the **marginal cone** [Fienberg and Rinaldo, 2005]. It is an intersection of a finite number of half spaces [Dotour, 2008], or a set spanned by vectors. The p-dimensional vector $\mathbf{t}$ is the vector of marginal totals including sufficient statistics for the vector of $\boldsymbol{\mu}$ [Fienberg and Rinaldo, 2005]. The necessary and sufficient condition for the existence of the MLE of $\boldsymbol{\mu}$ in Theorem 4.2 could be reduced to the geometric study of these sufficient statistics. Eriksson et al. [2006] prove that under any sampling designs (Poisson, multinomial, Product multinomial) the MLE of $\boldsymbol{\mu}$, thus $\mathbf{m}$, exists if and only if the observed margins, $\mathbf{t} = \mathsf{A}^{\mathsf{T}}\mathbf{y}$, lie in the relative interior of the marginal cone. In this case, the log likelihood function parametrized by cell means ($\mathbf{m} \in \mathscr{M}$) or natural parameters ($\boldsymbol{\theta} \in \mathbb{R}^p$) is a concave function and guarantees a unique maximizer. The relative interior of the marginal cone is $\{\mathbf{t} : \mathbf{t} = \mathsf{A}^{\mathsf{T}}\mathbf{y}, \mathbf{y} \in \mathbb{R}^{|L|}_{>0}\}$. So the MLE of $\boldsymbol{\mu}$ exists if and only if there exist a $\mathbf{y} > 0$ for which $\mathbf{t} = \mathsf{A}^{\mathsf{T}}\mathbf{y}$, and this is equivalent to Theorem 4.2. These concepts are also studied for a more general sampling design known as conditional Poisson scheme, originally introduced by Haberman based on linear restrictions on data $\mathbf{y}$, by Fienberg and Rinaldo [2012a].

Consider model (4.1) for which $\mathbf{t} = \mathsf{A}^{\mathsf{T}}\mathbf{y}$ is,

$$\begin{bmatrix} y_{+++} \\ y_{1++} \\ y_{+1+} \\ y_{11+} \\ y_{++1} \\ y_{1+1} \\ y_{+11} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_{000} \\ y_{100} \\ y_{010} \\ y_{110} \\ y_{001} \\ y_{101} \\ y_{011} \\ y_{111} \end{bmatrix}.$$

For the given observations in Example 4.2, the positive integer vector $\mathbf{y}^{\mathsf{T}} = (y_{000} + \delta, y_{100} - \delta, y_{010} - \delta, y_{110} + \delta, y_{001} - \delta, y_{101} + \delta, y_{011} + \delta, y_{111} - \delta)$ produces exactly the same $\mathbf{t}$, thus the MLE exists. But for the observed cell counts in Example 4.1, this

vector is not all positive any more although the vector of margins is still fixed. Thus for this pattern of zeros the MLE does not exist according to the polyhedral condition.

This polyhedral condition also means that the MLE does not exist if and only if $\mathbf{t} = \mathsf{A}^{\mathsf{T}}\mathbf{y}$ lies on a facet of the marginal cone C_{A}, or in other words, belongs to the relative interior of some proper face of the marginal cone [Eriksson et al., 2006, Fienberg and Rinaldo, 2005 and 2012]. Therefore, to realise if the MLE exists or not, instead of finding $\boldsymbol{\delta}$ (in Theorem 4.2) or $\mathbf{y} > 0$ (such that $\mathbf{t} = \mathsf{A}^{\mathsf{T}}\mathbf{y}$), we can check whether the marginals vector lies on a facet of the cone to expose a zero pattern which is one of the facets of the cone.

A **face** of the marginal cone is defined as a set $\mathsf{F} = \{\mathbf{t} \in C_{\mathsf{A}} : (\mathbf{t}, \boldsymbol{\zeta}) = 0\}$ for some $\boldsymbol{\zeta} \in \mathbb{R}^p$, such that $(\mathbf{t}, \boldsymbol{\zeta}) \geq 0$ for all $\mathbf{t} \in C_{\mathsf{A}}$, and $(\mathbf{t}, \boldsymbol{\zeta})$ represents their inner product. A polyhedral cone has a finite number of faces and it is a face of itself which is called the improper one [Fienberg and Rinaldo, 2005].

The **facial set** $\mathscr{F}$ is a set of cell indexes of the rows of A whose conic hull is precisely F. It means, $\mathscr{F} \subseteq L$ is a facial set of F for any design matrix A for $\mathscr{M}$, if there exists some $\boldsymbol{\zeta} \in \mathbb{R}^p$ such that,

$$\begin{aligned} (\mathsf{A}_{(i)}, \boldsymbol{\zeta}) &= 0, \qquad \text{if} \quad i \in \mathscr{F}, \\ (\mathsf{A}_{(i)}, \boldsymbol{\zeta}) &> 0, \qquad \text{if} \quad i \in \mathscr{F}^c, \end{aligned} \tag{4.4}$$

$\mathsf{A}_{(i)}, i = 1, \ldots, n$ is the ith row of A and $\mathscr{F}^c = L - \mathscr{F}$ is the **co-facial set** of F. So in terms of notations, F is a face, $\mathscr{F}$ is a facial set and $\mathscr{F}^c$ is a co-facial or a **facet** of a cone [Fienberg and Rinaldo, 2012a, 2012b].

Another way to define the facial set is by denoting sub matrices obtained from A, named A_+ and A_0. They are made of rows indexed by $L_+ := \{i : y_i \neq 0\}$ and $L_0 := \{i : y_i = 0\}$ respectively. The vector of marginals belongs to the relative interior of some proper face of the marginal cone if and only if $\mathscr{F}^c \subseteq L_0$. This is equivalent to the existence of a vector $\boldsymbol{\zeta}$ satisfying the following three conditions:

1. $\mathsf{A}_+\boldsymbol{\zeta} = \mathbf{0}$. (4.5)
2. $\mathsf{A}_0\boldsymbol{\zeta} \gneq \mathbf{0}$.
3. The set $\{i : (\mathsf{A}\boldsymbol{\zeta})_{(i)} \neq 0\}$ has maximal cardinality among all sets of the form $\{i : (\mathsf{A}\mathbf{x})_{(i)} \neq 0\}$ with $\mathsf{A}\mathbf{x} \gneq 0$ [Fienberg and Rinaldo, 2012b].

In these three conditions and also in (4.4) the inequality signs greater than zero could be switched to less than zero without loss of generality [Fienberg and Rinaldo, 2005]. In conclusion, if $\text{rank}(\mathsf{A}_+) = \text{rank}(\mathsf{A})$, the MLE exists since there is not such a vector like

$\boldsymbol{\zeta}$ and $\mathscr{F}^c = \varnothing$. If $\text{rank}(\mathsf{A}_+) < \text{rank}(\mathsf{A})$, the MLE may still exist, so we should search for a facial set [Fienberg and Rinaldo, 2012b].

Example 4.3. As a general example, consider fitting model (4.1) to a 2^3 contingency table with two zero cell counts. We can find all the zero patterns that lead to a non-existence MLE in this model. As mentioned before, the marginals $y_{ij+}, y_{+jk}, y_{i+k}$ are sufficient statistics and there are four different marginals of each, so they could make 12 zero patterns with the non-existence MLE. There are four more patterns which are not causing zero marginals but still it is not possible to find a $\boldsymbol{\delta}$ for them that suits Theorem 4.2, or a $\boldsymbol{\zeta}$ that fits (4.4) is possible. In these patterns, zeros could be in cells 1 and 8 (Example 4.1), or cells 2 and 7, or cells 3 and 6, or cells 4 and 5. Therefore, there are 16 facets of the cone, which 12 of them are associated with zero margins. Then, according to (4.2), the probability that two randomly zeros cause the non-existent MLE without inducing zero margins is $\frac{4}{28} = \frac{1}{7}$, as there are 28 different cases of choosing two cells from 8 cells.

Algorithms to detect the existence of the MLE

Linear and non-linear optimisation methods can be applied to detect the existence of the cell mean MLEs based on the polyhedral description of the model. Two of such algorithms from Fienberg and Rinaldo [2012b] are given here. The first one indicates whether the MLE exists and the second one provides the inestimable cells of the table in case of non-existent MLE. More alternative methods of linear and non-linear optimisation for the same purposes are presented by Fienberg and Rinaldo [2012b].

Algorithm 1: The polyhedral condition for the existence of the MLE, that observed margins must lie in the relative interior of the marginal cone, is equivalent to the following linear program,

$$\begin{aligned} & \max s \\ \text{such that} \quad & \mathsf{A}^\mathsf{T}\mathbf{y} = \mathbf{t} \\ & y_{\mathbf{i}} - s \geq 0, \quad \forall \mathbf{i} \\ & s \geq 0. \end{aligned}$$

The MLE does not exist if and only if the optimum s^* is zero, since it means that a strictly positive vector of cell counts satisfying $\mathsf{A}^\mathsf{T}\mathbf{y} = \mathbf{t}$ does not exist.

Algorithm 2: When rank(A) is bigger than rank(A_+), we look for finding the facial set. Kernels of (A_+) can be put in columns of a matrix like Z, then define $B = \mathsf{A}_0 Z$ whose rank is $q = \text{rank}(\mathsf{A}) - \text{rank}(\mathsf{A}_+)$. So,

$$\mathsf{A}Z\mathbf{y} = \begin{bmatrix} \mathsf{A}_+ Z\mathbf{y} \\ \mathsf{A}_0 Z\mathbf{y} \end{bmatrix} = \begin{bmatrix} 0 \\ B\mathbf{y} \end{bmatrix},$$

for some $\mathbf{y} \in \mathbb{R}^q$. It is the same as the non-linear optimisation problem,

$$\begin{aligned} \max \quad & |\text{supp}(B\mathbf{y})| \\ \text{such that} \quad & B\mathbf{y} \geq 0. \end{aligned}$$

The MLE exists if and only if the system $B\mathbf{y} \gneq 0$ is infeasible and any optimal solution $\mathbf{y}^*$ identifies the co-facial set $\mathscr{F}^c = \{i : (B\mathbf{y}^*)_{(i)} \neq 0\}$.

What if the MLE does not exist?

If we conclude that the MLE exists by checking any of the methods explained before, then the log-likelihood function must be maximised to find the estimates for model parameters. Therefore,

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta} \in \mathbb{R}^p} l(\boldsymbol{\theta}) = \text{argmax}_{\boldsymbol{\theta} \in \mathbb{R}^p} \mathbf{t}^\mathsf{T} \boldsymbol{\theta} - \mathbf{1}^\mathsf{T} \exp(\mathsf{A}\boldsymbol{\theta}).$$

If A is full rank a numerical method like the Newton-Raphson could be implied to find a unique maximizer and it will converge from any starting approximation. Then the MLE of the cell mean vector is $\hat{\mathbf{m}} = \exp(\mathsf{A}\hat{\boldsymbol{\theta}})$ [Fienberg and Rinaldo, 2012b].

If the MLE does not exist, then based on the extended MLE theory we can take a subset of size p_F of the original parameters (or a linear combinations of them [Fienberg and Rinaldo, 2012a]) and estimate them. Let $\mathsf{A}_{\mathscr{F}}$ be the matrix whose rows are only those from A that their coordinates are in $\mathscr{F}$, so its rank is p_F and whose column span is $\mathscr{M}_{\mathscr{F}}$ (the methods and algorithms described before are used to find the set $\mathscr{F}$). Then we can replace this $|\mathscr{F}| \times p$ design matrix with any other full rank one named $\mathsf{A}^*_{\mathscr{F}}$ of rank p_F and identical column range to use a minimal representation ($p_\mathsf{F} = \text{rank}(\mathsf{A}_{\mathscr{F}}) = \text{rank}(\mathsf{A}^*_{\mathscr{F}})$). Now the natural parameter space for the reduced model is $\mathbb{R}^{p_\mathsf{F}}$. $\mathsf{A}^*_{\mathscr{F}}$ is made of any linear independent rows from $\mathsf{A}_{\mathscr{F}}$ and can be formed by using Proposition 5.1 in [Fienberg and Rinaldo, 2012b]. The idea of the extended exponential family provides justification to identify a subset of cell means and parameters which are estimable and the inestimable cells in $\mathscr{F}^c$ are treated as being structural zeros. By implementing this reduced design matrix, the log likelihood function is defined which

is concave and admits a unique maximizer. The extended MLE is,

$$\hat{\boldsymbol{\theta}}^e = \text{argmax}_{\boldsymbol{\theta}\in\mathbb{R}^{p_{\mathsf{F}}}} l_F(\boldsymbol{\theta}) = \text{argmax}_{\boldsymbol{\theta}\in\mathbb{R}^{p_{\mathsf{F}}}} \mathbf{t}_F^{\mathsf{T}}\boldsymbol{\theta} - \mathbf{1}^{\mathsf{T}}\exp(\mathsf{A}^*_{\mathscr{F}}\boldsymbol{\theta}), \tag{4.6}$$

in which $\mathbf{t}_F = (\mathsf{A}^*_{\mathscr{F}})^{\mathsf{T}}\mathbf{y}_{\mathscr{F}}$, and the MLE of the cell mean vector is $\hat{\mathbf{m}}^e = \exp(\mathsf{A}^*_{\mathscr{F}}\hat{\boldsymbol{\theta}}^e)$. The Newton-Raphson method is the primary approach to finding the estimates. The degree of freedom for the reduced model is $d.f = |\mathscr{F}| - \text{rank}(\mathsf{A}^*_{\mathscr{F}})$, which is the number of usable data or cell means that are estimable minus the number of estimable log-linear model parameters [Fienberg and Rinaldo, 2012a, 2012b]. Examples to illustrate the existence of the MLE method are provided in section 4.2.2.

Modifying the likelihood function in a straightforward way and maximizing the extended likelihood function by iterative numerical methods to find the MLEs are studied by Fienberg and Rinaldo [2012a, 2012b].

4.2 Comparing the two approaches: The existence of the MLE and the parameter redundancy

4.2.1 Methods comparison

The two approaches are empirically compared here with respect to their processes and produced results. If there are such facial and co-facial sets as defined in (4.4), the MLE does not exist and some of the zero cells are treated as structural zeros. In the parameter redundancy, this is equivalent to $\boldsymbol{\alpha}^{\mathsf{T}}D = \mathbf{0}$ and no esoteric constraints in $\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{U}(\boldsymbol{\theta}) = 0$ which translates to the existence of a flat ridge, non-orthogonal on the parameters' axes, or a flat surface in the likelihood function. When $\text{rank}(\mathsf{A}_+) = \text{rank}(\mathsf{A})$ and the co-facial set is null, the MLE exists. It is equivalent to the model being full rank in the parameter redundancy, so there is a peak point in the likelihood function. When $\text{rank}(\mathsf{A}_+) < \text{rank}(\mathsf{A})$ and there is no co-facial set as described in (4.4), the MLE exists. In the parameter redundancy, it means that the model is parameter redundant but the esoteric constraints exist. It is possible then to find the MLEs for all the parameters as the flat ridge in the likelihood function is orthogonal on some of the parameters' axes. Furthermore, the D matrix in the parameter redundancy approach is the same as the transpose of A_+ matrix in the existence of the MLE approach. We look for finding the $\boldsymbol{\alpha}$ vector satisfying $\boldsymbol{\alpha}^{\mathsf{T}}D = \mathbf{0}$, then $\boldsymbol{\alpha} = \boldsymbol{\zeta}^{\mathsf{T}}$ in (4.5). The second and third conditions in (4.5), which indicate whether the MLE exists or not, perform the same task as $\boldsymbol{\alpha}^{\mathsf{T}}\,\mathbf{U}(\boldsymbol{\theta}) = 0$.

In the parameter redundancy method, we do not use the table marginals until utilising the log-likelihood function to determine the esoteric constraints, despite the polyhedral

method which is based on the observed marginal vector. When the pattern of zeros in the table does not show a marginal zero, the rank of the derivative matrix may still get decreased and make the model parameter redundant. However, for parameter redundant models that the MLE exists, the presence of the esoteric constraints makes all the parameters estimable (and these are the only models that we have observed the esoteric constraints for them). Providing the esoteric constraints by the parameter redundancy method is an advantage, since they present some extra information about the relation among parameters in the likelihood function of the parameter redundant models with the existent MLE.

The existence of the MLE approach focuses on $\boldsymbol{\mu}$ not $\boldsymbol{\theta}$ since θs have different interpretations based on the chosen parameter constraint, while the concept of the cell means is constant. In the parameter redundancy approach, the focus is on $\boldsymbol{\theta}$ as the parameters that describe the relation between variables and their effect on the cell means. Some examples of studies that are interested in estimating θs are given in Chapter 5. After identifying the estimable and inestimable cell means or θs, it is possible to find which of the other ones are estimable. In the parameter redundancy approach the number of estimable cell means is not known in advance, unlike the existence of the MLE approach, in which $|\mathscr{F}|$ shows the number of estimable cell means and p_{F} shows the number of estimable model parameters. Another difference between the two methods is that the polyhedral method is defined only for the hierarchical and the class of graphical models but the parameter redundancy approach is not limited to these type of models.

As it was mentioned before, the existence of the MLE approach converts the problem to an optimisation issue which can be handled with various numerical methods. Wang et al. [2016] report that these methods to find the co-facial sets do not work when the number of variables in the model is larger than 16. In the parameter redundancy approach, the symbolic algebra package `Maple` is used to simultaneously solve a number of corresponding partial differential equations and it is unable to do the calculation for a large number of equations when the model deficiency is about 40.

4.2.2 Examples

In this section, we investigate a few examples by using the two methods; parameter redundancy described in Section 2.3 and the existence of the MLE described in Section 4.1.

Example 4.4. The example presented in Section 2.6, including a contingency table with three variables and three levels for each, was studied by the parameter redundancy

1	3	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	4	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	5	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	6	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	7	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	8	1	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	9	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	10	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
9	11	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
10	12	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
11	13	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
12	14	1	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0
13	16	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0
14	17	1	1	0	0	1	1	0	0	0	1	0	0	1	0	0	1	0	0	0
15	21	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
16	22	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0
17	23	1	1	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0
18	24	1	0	1	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1
19	25	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0
20	26	1	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0
21	27	1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1

Fig. 4.1 $\mathsf{A}_{\mathscr{F}}$ matrix for the log-linear model fitted to the 3^3 contingency table.

method. Now we investigate the same example by using the existence of the MLE method to compare the results. The contingency table is given in Table 2.3 and the design matrix is shown in Figure 2.1 for the model,

$$\log \boldsymbol{\mu}_{27\times 1} = \mathsf{A}_{27\times 19}\boldsymbol{\theta}_{19\times 1}.$$

The $\text{rank}(\mathsf{A}) = 19$ and we have,

$$\begin{aligned} L_0 &= \{1,2,15,17,18,19,20,25\}, \\ L_+ &= L - L_0, \end{aligned}$$

such that the numbers show the corresponding cells in the contingency table. Matrices A_+ and A_0 get the rows from A with corresponding numbers in L_+ and L_0. Then $\text{rank}(\mathsf{A}_+) = 18 < \text{rank}(\mathsf{A}) = 19$, and the MLE may still exist, so we search for a facial set. The facial set and co-facial set of the cone and $\boldsymbol{\zeta}$ according to conditions (4.4) are,

$$\begin{aligned} \mathscr{F}^c &= \{1,2,15,18,19,20\}, \\ \mathscr{F} &= L - \mathscr{F}^c, \\ \boldsymbol{\zeta}^\mathsf{T} &= (1,0,-1,-1,-1,-1,0,0,1,0,1,1,1,0,0,0,1,0,0). \end{aligned}$$

Cells 17 and 25 are included in L_0 but not in $\mathscr{F}^c$, as those corresponding rows of the design matrix multiply in $\boldsymbol{\zeta}$ equal to zero as well. Therefore, cells 17 and 25 are distinguished with bold in the contingency table.

Now we can form $\mathsf{A}_{\mathscr{F}}$ matrix, which is shown in Figure 4.1. The dimension of this matrix is $|\mathscr{F}| \times p = 21 \times 19$ and $\text{rank}(\mathsf{A}_{\mathscr{F}}) = p_\mathsf{F} = 18$, so it is not a full rank matrix.

The matrix $\mathsf{A}^*_{\mathscr{F}}$ which is full rank with dimensions $|\mathscr{F}| \times p_{\mathsf{F}} = 21 \times 18$, should be formed by reducing $\mathsf{A}_{\mathscr{F}}$. If we use the function `drop.coef` in `R`, the reduced design matrix is $\mathsf{A}_{\mathscr{F}}$ without its 17th column. Meaning that θ^{XZ}_{21} is removed from the model and the new set of variables $\boldsymbol{\theta}'$ is containing 18 parameters. Thus the reduced model, non saturated with $d.f = |\mathscr{F}| - p_{\mathsf{F}} = 3$, is,

$$\log \boldsymbol{\mu}'_{21\times 1} = \mathsf{A}^*_{21\times 18}\boldsymbol{\theta}'_{18\times 1}.$$

By comparing the result from this method with what we obtained in Section 2.6, we realise that there are different vectors of 18 estimable parameters, but the estimates for each available parameter and for the cell means would be the same in a numerical example. Which cell means or θ parameters are estimable depends on the design matrix for the reduced model or $\mathsf{A}^*_{\mathscr{F}}$ and the two methods have different ways to construct it. In the polyhedral method, $\mathsf{A}_{\mathscr{F}}$ must be reduced to a full rank matrix. In the parameter redundancy method, we consider the estimable combinations of parameters and find out which cell means could be defined by them and form the reduced design matrix based on that.

For this model, there does not exist a $\boldsymbol{\delta}$ as in Theorem 4.2 or an all positive vector that satisfies (4.3), so the MLE does not exist. Also by checking the esoteric constraint conditions, we are not able to derive any constraint that makes all the cell means estimable.

In the next two examples, we revisit Example 4.1 and Example 4.2 to apply both methods of parameter redundancy and the existence of the MLE. For a parameter redundant model without any possible additional esoteric constraints, as for the model in Example 4.1, the only way to proceed with the initially considered model is to reduce it. But for a model which is parameter redundant and the MLE does exist according to the existence of the MLE method, there are more options to consider. For the model in Example 4.2, there is an esoteric constraint that can make all the parameters estimable. This suggests two ways to deal with these type of models. We could reduce the model to a smaller identifiable one, or use the esoteric constraint which is equivalent to using the numerical methods and maximize the likelihood function to obtain the MLE for all the parameters.

Example 4.5. Consider the model and the data presented in Example 4.1. We showed that the MLE does not exist for this model according to Theorem 4.2. We revisit this example using both methods of parameter redundancy and the EMLE.

Parameter redundancy. If the table contained only one zero cell, the model would be full rank. But there are two zero cells in the table and all other cells are assumed to have positive cell counts. In order to discover the possible parameter redundancy, we apply the method described in Section 2.3. The associated derivative matrix is,

$$D = \left[\begin{array}{c|cccccccc} & \mu_{000} & \mu_{100} & \mu_{010} & \mu_{110} & \mu_{001} & \mu_{101} & \mu_{011} & \mu_{111} \\ \hline \theta & 0 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & 0 \\ \theta^X & 0 & y_2 & 0 & y_4 & 0 & y_6 & 0 & 0 \\ \theta^Y & 0 & 0 & y_3 & y_4 & 0 & 0 & y_7 & 0 \\ \theta^{XY} & 0 & 0 & 0 & y_4 & 0 & 0 & 0 & 0 \\ \theta^Z & 0 & 0 & 0 & 0 & y_5 & y_6 & y_7 & 0 \\ \theta^{XZ} & 0 & 0 & 0 & 0 & 0 & y_6 & 0 & 0 \\ \theta^{YZ} & 0 & 0 & 0 & 0 & 0 & 0 & y_7 & 0 \end{array}\right].$$

The rank of this matrix is 6, so $d = p - r = 7 - 6 = 1$ and from (2.2) we have,

$$\boldsymbol{\alpha}^{\mathsf{T}} = (1, -1, -1, 1, -1, 1, 1).$$

Solving the corresponding equations in (2.3) indicates the estimable parameters are,

$$\boldsymbol{\theta}'^{\mathsf{T}} = (\theta + \theta^X, \theta + \theta^Y, -\theta + \theta^{XY}, \theta + \theta^Z, -\theta + \theta^{XZ}, -\theta + \theta^{YZ}).$$

According to these estimable combinations of parameters, all of the cell means but the first one ($\log \mu_{000} = \theta$) and the last one ($\log \mu_{111} = \theta + \theta^X + \theta^Y + \theta^{XY} + \theta^Z + \theta^{XZ} + \theta^{YZ}$) are estimable. We treat these two cells as they are structural zeros and remove them from the model. Then, we reduce the model to a saturated one with design matrix of rank 6, and the MLE for model parameters and cell means would be derived by maximizing the likelihood function of the reduced model. The reduced model is,

$$\begin{bmatrix} \log \mu_{100} \\ \log \mu_{010} \\ \log \mu_{110} \\ \log \mu_{001} \\ \log \mu_{101} \\ \log \mu_{011} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta + \theta^X \\ \theta + \theta^Y \\ -\theta + \theta^{XY} \\ \theta + \theta^Z \\ -\theta + \theta^{XZ} \\ -\theta + \theta^{YZ} \end{bmatrix}.$$

To check if there is an esoteric constraint for this parameter redundant model, we have,

$$\boldsymbol{\alpha}^{\mathsf{T}} \mathbf{U}(\boldsymbol{\theta}) = y_{000} + y_{111} - e^{\theta} - e^{\theta + \theta^X + \theta^Y + \theta^{XY} + \theta^Y + \theta^{XZ} + \theta^{YZ}},$$

which cannot be zero for finite θs, as,

$$-e^{\theta} - e^{\theta+\theta^X+\theta^Y+\theta^{XY}+\theta^Y+\theta^{XZ}+\theta^{YZ}} \neq 0.$$

Therefore, there is no esoteric constraint for this model.

The existence of the MLE. This example includes no zero sufficient marginals for the model but positive MLE for the cell means do not exist according to the Haberman's sufficiency and necessary condition for the existence of the MLE. The mentioned polyhedral condition also confirms that the MLE do not exist for this example as there is no $\mathbf{y} > 0$ that yields (4.3).

The sub-matrices of the design matrix are,

$$\mathsf{A}_+ = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}, \qquad \mathsf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The facial and co-facial sets and $\boldsymbol{\zeta}$ vector as defined to satisfy conditions (4.4) are,

$$\begin{aligned} \mathscr{F} &= \{100, 010, 110, 001, 101, 011\}, \\ \mathscr{F}^c &= \{000, 111\}, \\ \boldsymbol{\zeta}^{\mathsf{T}} &= (1, -1, -1, 1, -1, 1, 1), \end{aligned}$$

which also satisfies three conditions of (4.5). $\text{Rank}(\mathsf{A}_+) < \text{rank}(\mathsf{A})$, so the MLE might exist. Due to the existence of $\boldsymbol{\zeta}$, the vector of marginals belongs to the relative interior of the face F, then the MLE does not exist for all of the cell means.

$$\mathsf{A}_+\boldsymbol{\zeta} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0},$$

$$A_0\boldsymbol{\zeta} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

To double check that the vector of marginals ($\mathbf{t}$) belong to the relative interior of the face F, we have,

$$\mathsf{F} = \{\mathbf{t} \in C_{\mathsf{A}} : \boldsymbol{\zeta}^{\mathsf{T}}\mathbf{t} = 0\} = \{\mathsf{A}^{\mathsf{T}}\mathbf{y} : \boldsymbol{\zeta}^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}\mathbf{y} = 0\}.$$

The $\boldsymbol{\zeta}$ that we have found matches the above condition, meaning $\boldsymbol{\zeta}^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}\mathbf{y} = 0$. Because $\boldsymbol{\zeta}^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}$ is zero for those columns of A^{T} that are rows of A_+ as $\mathsf{A}_+\boldsymbol{\zeta} = 0$. Then the other elements of $\boldsymbol{\zeta}^{\mathsf{T}}\mathsf{A}^{\mathsf{T}}\mathbf{y}$ are zero, since the $y_{\mathbf{i}}$s corresponding to them are zero.

The MLE does not exist for this model and we must reduce the model. In this example, $\mathsf{A}_{\mathscr{F}} = \mathsf{A}_+$. The design matrix for the reduced model is $\mathsf{A}^*_{\mathscr{F}}$ which is a $|\mathscr{F}| \times p_{\mathsf{F}} = 6 \times 6$ matrix. It could be found by using the suggested proposition in Fienberg and Rinaldo [2012b] and using their `MATLAB` function `Recompute-Basis`[1]. Given a matrix U not of full-column rank, this function produces a matrix of full column rank whose columns are a subset of the columns of U. Then, based on the reduced design matrix, the final model becomes,

$$\begin{bmatrix} \log\mu_{100} \\ \log\mu_{010} \\ \log\mu_{110} \\ \log\mu_{001} \\ \log\mu_{101} \\ \log\mu_{011} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta \\ \theta^X \\ \theta^Y \\ \theta^{XY} \\ \theta^Z \\ \theta^{XZ} \end{bmatrix}.$$

The estimable cell means are the same as derived by the parameter redundancy approach, but θ^{YZ} is dropped from the model. So six estimable parameters remain (instead of making six estimable combinations of parameters). The estimates for this set of parameters are derived by maximizing the corresponding likelihood function as it is shown in (4.6).

Function `drop.coef` in `R` could also be used to transform matrix $\mathsf{A}_{\mathscr{F}}$ to another full rank matrix by dropping some of its columns. The matrix given by this function

[1] Available at http://www.stat.cmu.edu/~arinaldo/?page_id=137

is $\mathsf{A}_{\mathscr{F}}$ without the 6th column. Hence, one of the parameters (θ^{XZ}) is removed as the dimension of the parameters vector must be 6×1.

In a numerical example, the maximum likelihood estimates for all six estimable cell means are the same in these two methods, and log-linear model parameter estimates are likewise consistent. For example, the estimated value for $\theta + \theta^X$ in the first reduced model equals to $\hat{\theta} + \hat{\theta}^X$ in the second one. Although both methods reduce the model to a model with only six estimable parameters, the parameters interpretations are different. The parameters derived by parameter redundancy approach are exactly the ones in the initial model. But for instance, θ in the second reduced model is not the intercept of the initial log-linear model.

Example 4.6. Consider the model and the data table presented in Example 4.2. We showed that the MLE exists for this model according to Theorem 4.2. We revisit this example by using both methods of parameter redundancy and the EMLE.

Parameter redundancy. For applying the parameter redundancy approach, the derivative matrix is derived as,

$$D = \left[\begin{array}{c|cccccccc} & \mu_{000} & \mu_{100} & \mu_{010} & \mu_{110} & \mu_{001} & \mu_{101} & \mu_{011} & \mu_{111} \\ \hline \theta & 0 & y_2 & y_3 & 0 & y_5 & y_6 & y_7 & y_8 \\ \theta^X & 0 & y_2 & 0 & 0 & 0 & y_6 & 0 & y_8 \\ \theta^Y & 0 & 0 & y_3 & 0 & 0 & 0 & y_7 & y_8 \\ \theta^{XY} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_8 \\ \theta^Z & 0 & 0 & 0 & 0 & y_5 & y_6 & y_7 & y_8 \\ \theta^{XZ} & 0 & 0 & 0 & 0 & 0 & y_6 & 0 & y_8 \\ \theta^{YZ} & 0 & 0 & 0 & 0 & 0 & 0 & y_7 & y_8 \end{array}\right].$$

The rank of this matrix is 6 again and $d = p - r = 7 - 6 = 1$. Then,

$$\boldsymbol{\alpha}^{\mathsf{T}} = (1, -1, -1, 0, -1, 1, 1),$$

which indicates the estimable parameters are,

$$\boldsymbol{\theta}'^{\mathsf{T}} = (\theta + \theta^X, \theta + \theta^Y, \theta^{XY}, \theta + \theta^Z, -\theta + \theta^{XZ}, -\theta + \theta^{YZ}).$$

By considering these estimable combinations, $\log \mu_{000}$ and $\log \mu_{110}$ are not estimable. So we reduce the initial model to a saturated one with the design matrix of rank 6. The

reduced model is,

$$\begin{bmatrix} \log\mu_{100} \\ \log\mu_{010} \\ \log\mu_{001} \\ \log\mu_{101} \\ \log\mu_{011} \\ \log\mu_{111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \theta+\theta^X \\ \theta+\theta^Y \\ \theta^{XY} \\ \theta+\theta^Z \\ -\theta+\theta^{XZ} \\ -\theta+\theta^{YZ} \end{bmatrix}.$$

To check if there is an esoteric constraint for this parameter redundant model, we have,

$$\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{U}(\boldsymbol{\theta}) = y_{000} - y_{110} - e^{\theta} + e^{\theta+\theta^X+\theta^Y+\theta^{XY}} = 0,$$

which means the esoteric constraint is,

$$\theta^X + \theta^Y + \theta^{XY} = 0, \qquad \text{or,} \qquad \log\mu_{000} = \log\mu_{110}. \tag{4.7}$$

Imposing this constraint on model (4.1) makes all parameters estimable, although the model is parameter redundant and has a flat ridge in the likelihood surface. From equation (4.7), $\theta^{XY} = -\theta^X - \theta^Y$ and we know that θ^{XY} is estimable. So the three estimable combinations of parameters $(-\theta^X - \theta^Y, \theta+\theta^X, \theta+\theta^Y)$ make a system of three equations and three unknowns and in result $(\theta, \theta^X, \theta^Y)$ are estimable. Estimability of θ makes other parameters $\theta^Z, \theta^{XZ}, \theta^{YZ}$ estimable.

The existence of the MLE. The existence of the MLE in this example is also confirmed by the polyhedral condition as the observed marginals lie in the relative interior of the marginal of the polyhedral cone. There exists a $\mathbf{y} > 0$ such that satisfies (4.3), Let $0 < \delta < 1$, then,

$$\mathbf{y}^{\mathsf{T}} = (y_1+\delta, y_2-\delta, y_3-\delta, y_4+\delta, y_5-\delta, y_6+\delta, y_7+\delta, y_8-\delta).$$

Or in other words, there is no $\boldsymbol{\zeta}$ which satisfies conditions in (4.4). The sub-matrices of the design matrix are,

$$\mathsf{A}_+ = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \qquad \mathsf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

$\text{Rank}(\mathsf{A}_+) < \text{rank}(\mathsf{A})$, so the MLE might still exist. To try to find $\boldsymbol{\zeta}$, we have,

$$\mathsf{A}_+\boldsymbol{\zeta} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{0},$$

but,

$$\mathsf{A}_0\boldsymbol{\zeta} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

which is not strictly positive (or negative). So we are unable to find a $\boldsymbol{\zeta}$ vector that satisfies the three conditions (4.5). It means the facial and co-facial sets do not exist in this case and the MLE exists. Thus, we can calculate the MLEs for all parameters of this model by numerical methods. The esoteric constraint (4.7) will still be true as a relation among the parameters but the numerical methods reach the estimates without reporting this constraint.

4.3 Parameter redundant models with existent MLE

We make use of a numerical example for a parameter redundant model with existent MLE to investigate the various options of handling such models. Three possible ways of dealing with such a model are reviewed here:

- Reducing the parameter redundant model to a smaller non-redundant one.
- Estimate the MLE for all the parameters using the numerical methods or the explicit formulas to estimate cell means of decomposable models, knowing that an extra set of constraints exists among the parameters.
- Since the esoteric constraints do reduce the number of initial parameters (by setting constraints among them), we might be able to find another model with

	Y	
X	0	1
0	0	44
1	538	0

(a) Observations

	Y	
X	0	1
0	40.67	3.33
1	497.3	40.67

(b) Cell mean estimates

	Y	
X	0	1
0	–	44
1	538	–

(c) Cell mean estimates in the reduced model

Table 4.7 A 2^2 table for the model of no-second-order interaction.

the same number of parameters which fits the data better by providing a smaller residual deviance or a smaller Akaike Information Criterion (AIC).

The example which is used here, was briefly mentioned in Section 2.5.3. Assume we are fitting the independence model or the model with no-first-order interaction for two variables,

$$\log \mu_{ij} = \theta + \theta_i^X + \theta_j^Y, \qquad i, j \in \{0, 1\}, \tag{4.8}$$

which could be shown as,

$$\begin{aligned}
m_1 &= \log \mu_1 = \log \mu_{00} = \theta, \\
m_2 &= \log \mu_2 = \log \mu_{10} = \theta + \theta^X, \\
m_3 &= \log \mu_3 = \log \mu_{01} = \theta + \theta^Y, \\
m_4 &= \log \mu_4 = \log \mu_{11} = \theta + \theta^X + \theta^Y.
\end{aligned}$$

The observed data are shown in Table 4.7, part (a).

In Section 2.5.3, the estimates with reasonable standard errors for this model parameters $\boldsymbol{\theta}^\top = (\theta, \theta^X, \theta^Y)$ were obtained by the `glm` function in `R` as,

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1   3.7056     0.1512   24.50   <2e-16 ***
A2   2.5037     0.1568   15.97   <2e-16 ***
A3  -2.5037     0.1568  -15.97   <2e-16 ***
---
predictions:
         1         2         3         4
  40.67354 497.32646   3.32646  40.67354
---
Residual deviance:  311.83  on 1  degrees of freedom
AIC: 331.58
BIC: 329.7379
```

However, it is a decomposable model and we can calculate the estimates directly as below,

$$\log\hat{\mu}_{00} = \hat{\theta} = \log\frac{y_{0+}y_{+0}}{N} = \log\frac{44\times 538}{44+538} = \log 40.67 = 3.7,$$
$$\log\hat{\mu}_{10} = \hat{\theta}+\hat{\theta}^X = \log\frac{y_{1+}y_{+0}}{N} = \log\frac{538\times 535}{44+538} = \log 497.3 = 6.2,$$
$$\log\hat{\mu}_{01} = \hat{\theta}+\hat{\theta}^Y = \log\frac{y_{0+}y_{+1}}{N} = \log\frac{44\times 44}{44+538} = \log 3.33 = 1.2,$$
$$\log\hat{\mu}_{11} = \hat{\theta}+\hat{\theta}^X+\hat{\theta}^Y = \log\frac{y_{1+}y_{+1}}{N} = \log\frac{44\times 538}{44+538} = \log 40.67 = 3.7$$

Therefore, the expected cell counts are the values obtained by R in Table 4.7, part (b). These equations force $\theta^X+\theta^Y=0$ and we get the same estimates for θ^X and θ^Y with opposite signs and the same standard errors. This is the esoteric constraint defined in Section 2.7 that makes all parameters estimable by obliging the flat ridge to be orthogonal on some parameter axes. In this example we have,

$$\boldsymbol{\alpha}^\top\mathbf{U} = y_{00}-y_{11}-e^{\theta}+e^{\theta+\theta^X+\theta^Y} = 0.$$

It means $\theta^X=-\theta^Y$ or $\log\mu_{00}=\log\mu_{11}$. Thus, we draw the conclusion that the fitted model here is actually a model with only two parameters due to the esoteric constraint, which could be presented as,

$$\begin{aligned}\log\mu_1 &= \log\mu_{00} = \theta,\\ \log\mu_2 &= \log\mu_{10} = \theta+\theta^X,\\ \log\mu_3 &= \log\mu_{01} = \theta-\theta^X,\\ \log\mu_4 &= \log\mu_{11} = \theta.\end{aligned} \tag{4.9}$$

Fitting this model gives the same estimates as before for $\boldsymbol{\theta}^\top=(\theta,\theta^X)$ as shown below, but the model has 2 degrees of freedom not 1.

```
Coefficients:
    Estimate Std. Error z value Pr(>|z|)
x31   3.7056     0.1028   36.03   <2e-16 ***
x32   2.5037     0.1109   22.58   <2e-16 ***
---
predictions:
        1         2         3         4
 40.67354 497.32646   3.32646  40.67354
---
Residual deviance:  311.83  on 2  degrees of freedom
AIC: 329.58
```

```
BIC: 328.3516
```

In order to check the parameter redundancy for model (4.8), the derivative matrix is formed as,

$$D = \left[\begin{array}{c|cccc} & \mu_{00} & \mu_{10} & \mu_{01} & \mu_{11} \\ \hline \theta & 0 & 538 & 44 & 0 \\ \theta^X & 0 & 538 & 0 & 0 \\ \theta^Y & 0 & 0 & 44 & 0 \end{array}\right].$$

The rank of the matrix is 2 and there are $d = p - r = 3 - 2 = 1$ vector of $\boldsymbol{\alpha}(\boldsymbol{\theta})$, which is $\boldsymbol{\alpha}^\top = (1, -1, -1)$. After solving the corresponding partial differential equation, the estimable parameters are $\theta + \theta^X, \theta + \theta^Y$ and estimable cell means are μ_{01}, μ_{10}. We reduce the initial model to a model with only these two estimable parameters and two estimable cells, which is a saturated model,

$$\begin{aligned} \log \mu_2 = \log \mu_{10} &= \theta + \theta^X, \\ \log \mu_3 = \log \mu_{11} &= \theta + \theta^Y. \end{aligned} \tag{4.10}$$

The estimates for model parameters $\boldsymbol{\theta}'^\top = (\theta + \theta^X, \theta + \theta^Y)$ and two cell means are obtained with reasonable standard errors as below and are also shown in Table 4.7, part (c).

```
Coefficients:
   Estimate Std. Error z value Pr(>|z|)
A1  6.28786    0.04311   145.8   <2e-16 ***
A2  3.78419    0.15076    25.1   <2e-16 ***
---
predictions:
  1   2
538  44
---
Residual deviance: -4.7073e-14  on 0  degrees of freedom
AIC: 17.752
BIC: 15.13819
```

Models (4.9) and (4.10) are not comparable, as the later one describes only two cell counts but the first one describes all four cell counts. Although model (4.9) is obtained automatically from the esoteric constraint and is maximising the likelihood function, the goodness of fit measurements do not seem very promising. We want to realise if there exist other possible models with four cells and only two parameters which improve these estimates and goodness of fit measurements. A possible model

which is empirically derived can be,

$$\begin{aligned}
\log \mu_1 &= \log \mu_{00} = \theta - \theta^X, \\
\log \mu_2 &= \log \mu_{10} = \theta + \theta^X, \\
\log \mu_3 &= \log \mu_{01} = \theta, \\
\log \mu_4 &= \log \mu_{11} = \theta - \theta^X.
\end{aligned} \tag{4.11}$$

Estimates for parameters $\boldsymbol{\theta}^\top = (\theta, \theta^X)$ and cell means, and some goodness of fit measurements for this model are,

```
Coefficients:
    Estimate Std. Error z value Pr(>|z|)
A1   3.5547     0.1375   25.85   <2e-16 ***
A2   2.7415     0.1419   19.33   <2e-16 ***
---
predictions:
         1          2          3          4
  2.255291 542.510581  34.978837   2.255291
---
Residual deviance:   11.208  on 2  degrees of freedom
AIC: 28.96
BIC: 27.73222
```

which indicates a better fit compared to model (4.9).

Therefore, fitting the independence model to data in Table 4.7 results in three different possible models. As we know that the model is parameter redundant because of the two zero cells, it could be reduced to a smaller saturated model as (4.10). But fitting the model in a software gives us the model (4.9) which has only two parameters instead of having the three parameters in the initial model. It is because of the esoteric constraint which maximizes the likelihood function. However, models may exist with two parameters that would fit the data better, such as model (4.11), but it can not be interpreted as a usual log-linear model.

Considering the approximate shape of the log-likelihood function in each of the models helps to clarify the idea of the existence of the MLE. The independence log-linear model (4.8) has three parameters. To provide an approximate illustration of the log-likelihood function, we fix one of the parameters and plot the log-likelihood against the remaining parameters. Figure 4.2 part (a) shows the log-likelihood function's shape when θ is fixed at its maximum likelihood estimate. In this model, all the cell counts are positive and a peak is obvious in the plot, which is the unique MLE for parameters θ^X and θ^Y. Part (b) of the figure shows the log-likelihood surface for the same model and

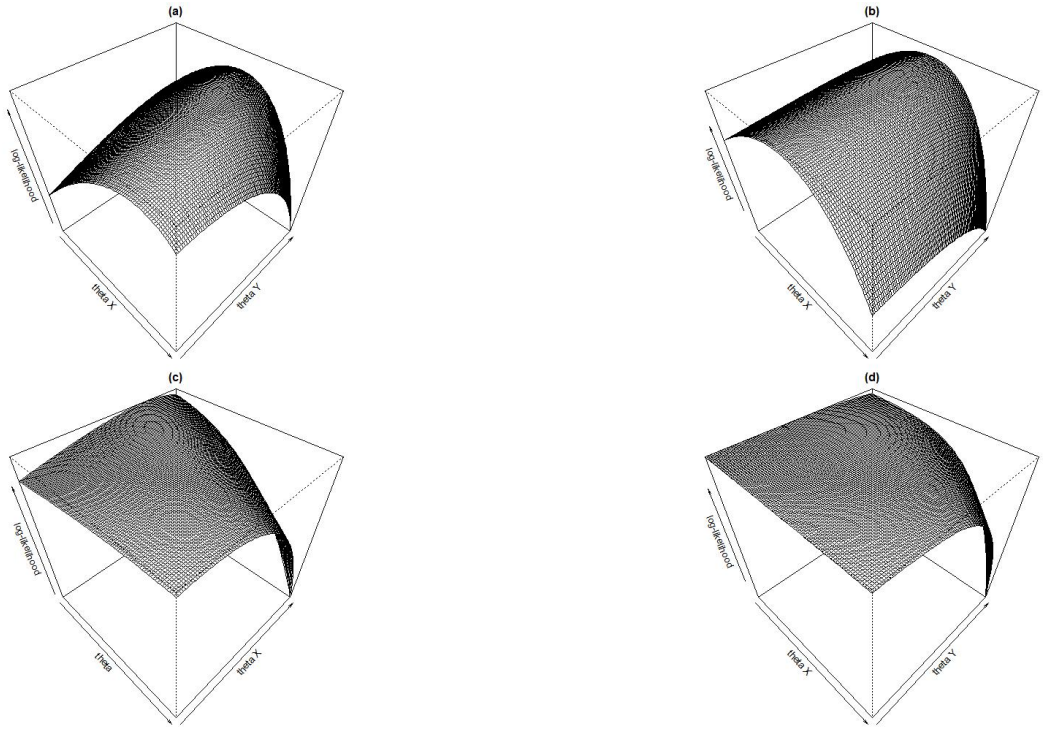

Fig. 4.2 Log-likelihood functions for (a) The independence model with $\mathbf{y} > 0$, (b) The independence model with $y_1 = y_4 = 0$, (c) Model (4.11) with two parameters, (d) The independence model with $y_1 = y_2 = 0$.

fixed intercept, for observations in Table 4.7 when $y_1 = y_4 = 0$. A flat ridge orthogonal on the θ^X axis is present. Part (c) of the figure is the log-likelihood surface for the proposed model (4.11), which only has two parameters and the surface includes a maximum point. Part (d) presents the likelihood function for model (4.8) with a fixed value for θ, when $y_1 = y_2 = 0$ which causes a marginal zero. This function includes a flat surface, indicating that the model must be reduced.

Chapter 5

Applications

5.1 Introduction

This chapter includes three examples of fitting log-linear models to categorical data in sparse contingency tables. In each example, we identify whether the initial model is parameter redundant. The estimable parameters and the estimable linear combinations of parameters are derived and the model is reduced to a smaller non-redundant model with all estimable parameters. We also check if there exist any esoteric constraints in the model as defined in Section 2.7, so the MLE could exist. The data in each example are taken from published papers and the results are compared with those from the papers when it is relevant. The first example includes a 2^5 contingency table which is cross-classifying the number of 2744 potential victims of human trafficking in the UK, over five different sources of identifying the victims [Silverman, 2014]. The second example is presenting again a 2^5 contingency table, classifying 119 patients based on five variables observed after an ear surgery [Brown and Fuchs, 1983]. In the third example, a 3^5 contingency table includes the frequency of 3841 individuals on the combinations of three levels of five different single nucleotide polymorphisms (SNPs) in two chromosomes [Papathomas et al., 2012]. An extra variable indicating the presence of cancer in each individual is added as well to make a sparse $3^5 \times 2^1$ contingency table.

5.2 Modern slavery study

5.2.1 The data

An analysis of the scale of modern slavery in the UK is reported by Silverman [2014]. According to this paper, the National Crime Agency in the UK carried out a strategic assessment in 2013 to locate and identify "potential victims of trafficking" to ensure

they receive the suitable and required support. These victims are said to be living in modern slavery.

The information about potential victims in this assessment is based on different sources, but it cannot present a whole picture of the victim population. Gathering the data is difficult due to the sensibility of the matter and the criminal nature of human trafficking. The part of this population which has not been found and counted in the process of identifying the victims is called "dark figure" and the aim of the paper is to estimate it and subsequently estimate the whole population size. Another example of estimating the dark figure is provided by Overstall et al. [2014], in which the aim is to estimate the number of people who inject drugs in Scotland.

The method applied to estimate the dark figure size is appointed as multiple system estimation, which is an extension of the mark-recapture approach when there are more than two lists [Silvermen, 2014]. In a contingency table, each variable can be considered as a list. The information about potential victims in this assessment is collected from a lot of different organisations, but they are summarised into five main lists. An individual might be on one or more than one of these lists. Then there are 31 possible situations for each individual to belong to one or more than one of theses lists, or they may belong to none of the lists which is the dark figure side that is supposed to be estimated. The five lists or the main sources of information on recognising the potential victims are:

- LA: Local Authority (A),
- NG: Non-governmental organisation (B),
- PF: Police force (C),
- GO: Government organisations (D),
- GP: The general public, through various routes (E).

The individuals who are identified by any or some of those lists are counted and the collected information is shown in Table 5.1 which is taken from Silverman [2014]. The total number of identified potential victims is 2744. The table reports that, for example, 54 individuals are identified by local authority and 15 of them are also in the list provided by the non-governmental organisation.

This data could be presented in a contingency table as each list is a categorical variable with two levels. Level 1 indicates the individual belongs on this list and 0 states otherwise. It forms a 2^5 contingency table given in Table 5.2. y_{00000} is the cell count when all variables are on their 0 level. It is unknown and shows the dark figure size

LA	×					×	×	×								×	×	×
NG		×				×			×	×	×				×	×	×	×
PF			×				×		×			×	×		×	×		×
GO				×				×		×		×		×	×		×	×
GP					×						×		×	×				
y_i	54	463	995	695	316	15	19	3	62	19	1	76	11	8	4	1	1	1

Table 5.1 Number of potential victims in different lists.

or the number of potential victims who are not identified by any of the information sources.

5.2.2 Analysis

The analysis by Silverman [2014] is carried out using the R package `Rcapture`, and the `closedp.MX` routine is used to fit the log-linear models. A forward method is applied to choose the best possible model with the best AIC (Akaike Information Criterion). In this model selection method, interaction effects are added stepwise and at each step the interaction which makes the biggest improvement in the AIC is included in the model. The main effects of each variable on the logarithm of cell means and the ten possible pairwise interactions of variables are considered as the initial parameters of the log-linear model. It is reported in the paper that: "It was found that the resulting model contains one interaction which has a very high standard error and is very far from statistically significant, and so this was dropped from the model" [Silverman, 2014]. The final model including five main effects and six firs-order interactions is given in Table 5.3. The dark figure's estimate is 8569, and the population size estimate derived by this model is 11313 with standard error 802. The 95% confidence interval for the population size obtained by the `profileCI` routine is $(9918, 13046)$.

To check the parameter redundancy, the model we want to fit the data must be chosen first. We try the saturated model, as it enables us to consider all the possible interactions among the five variables. The aim is to find the best log-linear model for the data in contingency Table 5.2 and then estimating the population size of potential victims. According to (1.5), the saturated log-linear model for a 2^5 table is,

$$\log \mu_{ijklm} = \theta + \theta_i^A + \theta_j^B + \theta_k^C + \theta_l^D + \theta_m^E + \theta_{ij}^{AB} + \theta_{ik}^{AC} + \theta_{jk}^{BC} + \cdots + \theta_{ijklm}^{ABCDE},$$

such that, $i, j, k, l, m \in \{0, 1\}^5$. 13 cell counts in the contingency table are zero, so some of the model parameters have estimates with large standard errors in fitting a saturated model. We aim to identify the estimable parameters and then ignore the inestimable ones in building the best model for the data. Note that the esoteric constraint and the

LA	NG	PF	GO	GP	cell	$y_{\mathbf{i}}$
1	1	1	1	1	y_{11111}	0
				0	y_{11110}	1
			0	1	y_{11101}	0
				0	y_{11100}	1
		0	1	1	y_{11011}	0
				0	y_{11010}	1
			0	1	y_{11001}	0
				0	y_{11000}	15
	0	1	1	1	y_{10111}	0
				0	y_{10110}	0
			0	1	y_{10101}	0
				0	y_{10100}	19
		0	1	1	y_{10011}	0
				0	y_{10010}	3
			0	1	y_{10001}	0
				0	y_{10000}	54
0	1	1	1	1	y_{01111}	0
				0	y_{01110}	4
			0	1	y_{01101}	0
				0	y_{01100}	62
		0	1	1	y_{01011}	0
				0	y_{01010}	19
			0	1	y_{01001}	1
				0	y_{01000}	463
	0	1	1	1	y_{00111}	0
				0	y_{00110}	76
			0	1	y_{00101}	11
				0	y_{00100}	995
		0	1	1	y_{00011}	8
				0	y_{00010}	695
			0	1	y_{00001}	316
				0	y_{00000}	-

Table 5.2 Contingency table of the observed number of potential victims in different lists.

Parameter		Silverman's model Estimate	Silverman's model Standard Error	The reduced model Estimate	The reduced model Standard Error
θ	Intercept	9.05591	0.09305	9.02849	0.09331
θ^A	LA	-5.08848	0.15254	-5.06467	0.15845
θ^B	NG	-2.90507	0.09507	-2.87858	0.09512
θ^C	PF	-2.14852	0.08809	-2.12439	0.08811
θ^D	GO	-2.52177	0.09129	-2.49543	0.09145
θ^E	GP	-3.30533	0.10827	-3.27274	0.10896
θ^{AB}	LA*NG	1.52395	0.27625	1.64387	0.29661
θ^{BE}	NG*GP	-2.92170	1.00582	-2.87716	1.00609
θ^{CE}	PF*GP	-1.24675	0.31883	-1.23345	0.31912
θ^{AC}	LA*PF	0.92243	0.26209	1.10502	0.27738
θ^{DE}	GO*GP	-1.19052	0.36926	-1.18087	0.36950
θ^{BD}	NG*GO	-0.55335	0.22399	-0.61809	0.22778
θ^{ABC}	LA*NG*PF			-1.70973	1.06433
$\theta^{ABCD}+\theta^{ACD}$	LA*NG*PF*GO + LA*PF*GO			3.11352	1.42952
Residual deviance		16.35 on 19 d.f		2.82 on 4 d.f	

Table 5.3 The parameters estimates for the final log-linear models.

MLE do not exist for this model since it is a saturated model with at least one zero cell in the table. As discussed in Chapter 4, the cell counts are sufficient statistics for a saturated model, thus the MLE does not exist here. Theorem 3.6 identifies the parameters that are not directly estimable after setting each cell observation to zero in a saturated model.

The design matrix for this model is formed in `R` and transposing it makes the derivative matrix only including 0 ans 1s. The rank of this matrix is 32 but after considering the zero cell counts, in accordance with the derivative matrix defined in (2.4), it reduces to 19. Since the number of parameters in this saturated model is $p = 32$, the deficiency is $d = p - r = 32 - 19 = 13$. We find those 13 $\boldsymbol{\alpha}$ vectors, given in Appendix B, and then solve the corresponding differential equations (2.3), using `Maple`, to obtain the 19 estimable parameters which are,

$$\begin{aligned}\boldsymbol{\theta}'^{\mathsf{T}} =&(\theta, \theta^A, \theta^B, \theta^C, \theta^D, \theta^E, \\ &\theta^{AB}, \theta^{AC}, \theta^{BC}, \theta^{AD}, \theta^{BD}, \theta^{CD}, \theta^{BE}, \theta^{CE}, \theta^{DE}, \\ &\theta^{ABC}, \theta^{ABD}, \theta^{BCD}, \theta^{ABCD}+\theta^{ACD}).\end{aligned} \tag{5.1}$$

The other parameters are not estimable. These estimable parameters show that the 13 cells with zero cell counts cannot have estimable means. The 19 cell means with non-zero cell counts are estimable. We make the reduced design matrix by considering

19 non-zero cell counts and 19 estimable parameters, which is full rank with the rank 19. There exist 19 non-zero cell counts but since the cell count corresponding to the intercept is unknown, we eliminate the corresponding row (with first element of 1 and all others 0) from the matrix. Now there are 18 cell counts, an 18×19 design matrix and 19 estimable parameters. By using the `glm` function, the parameter estimates for the reduced model $\log \boldsymbol{\mu}'_{18\times 1} = \mathsf{A}'_{18\times 19}\boldsymbol{\theta}'_{19\times 1}$ are given as below. One of the parameters is not estimated but not due to the parameter redundancy caused by zero cell counts, as in this stage we only have estimable cell means. This problem appears simply due to singular or under-determined system of equations, as the number of equations (18) is smaller than the number of parameters (19). To fix the problem, the number of parameters must be decreased.

```
Coefficients: (1 not defined because of singularities)
        Estimate Std. Error z value Pr(>|z|)
A1    9.5684     0.5801  16.494  < 2e-16 ***
A2   -5.5794     0.5959  -9.364  < 2e-16 ***
A3   -3.4306     0.5782  -5.933 2.98e-09 ***
A4   -2.6656     0.5792  -4.602 4.19e-06 ***
A5   -3.0245     0.5789  -5.225 1.74e-07 ***
A6   -3.8126     0.5828  -6.542 6.09e-11 ***
A7    2.1497     0.6477   3.319 0.000904 ***
A8    1.6211     0.6377   2.542 0.011020 *
A9    0.6550     0.5632   1.163 0.244832
A10   0.1341     0.8288   0.162 0.871479
A11  -0.1688     0.5294  -0.319 0.749807
A12   0.4524     0.5665   0.799 0.424479
A13  -2.3251     1.1565  -2.010 0.044384 *
A14  -0.6922     0.6554  -1.056 0.290909
A15  -0.6518     0.6806  -0.958 0.338206
A16  -2.3185     1.2062  -1.922 0.054592 .
A17   0.3512     1.3034   0.269 0.787609
A18       NA         NA      NA       NA
A19   2.2556     1.8405   1.226 0.220381
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2.8711e+04  on 18  degrees of freedom
Residual deviance: 4.2188e-14  on  0  degrees of freedom
AIC: 123.25
```

Model selection; Forward method

This model includes many parameters and it is possible to improve the model by removing some of them. We apply a forward selection method to choose the best possible model. The knowledge of estimable and inestimable parameters makes model selection easier by dealing with only 19 parameters instead of 32. First, we consider the model with only main effects. Second, we add 13 interaction effects stepwise and find the interaction which makes the biggest improvement in the AIC. The process terminates when AIC does not decrease any more. By using this procedure the remained parameters in the final model are,

$$\begin{aligned}&(\theta, \theta^A, \theta^B, \theta^C, \theta^D, \theta^E,\\ &\theta^{AB}, \theta^{AC}, \theta^{BD}, \theta^{BE}, \theta^{CE}, \theta^{DE},\\ &\theta^{ABC}, \theta^{ABCD} + \theta^{ACD}).\end{aligned}$$

In order to fit this model, building a proper design matrix A' is required. We have 18 cell counts and 14 parameters, so an 18×14 design matrix is formed. The parameters estimates subsequently are as below, and they are also shown in Table 5.3.

```
Coefficients:
      Estimate Std. Error z value Pr(>|z|)
A1    9.02849    0.09331  96.754  < 2e-16 ***
A2   -5.06467    0.15845 -31.965  < 2e-16 ***
A3   -2.87858    0.09512 -30.262  < 2e-16 ***
A4   -2.12439    0.08811 -24.111  < 2e-16 ***
A5   -2.49543    0.09145 -27.287  < 2e-16 ***
A6   -3.27274    0.10896 -30.036  < 2e-16 ***
A7    1.64387    0.29661   5.542 2.99e-08 ***
A8    1.10502    0.27738   3.984 6.78e-05 ***
A9   -0.61809    0.22778  -2.714 0.006656 **
A10  -2.87716    1.00609  -2.860 0.004240 **
A11  -1.23345    0.31912  -3.865 0.000111 ***
A12  -1.18087    0.36950  -3.196 0.001394 **
A13  -1.70973    1.06433  -1.606 0.108189
A14   3.11352    1.42952   2.178 0.029405 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 28711.2319  on 18  degrees of freedom
Residual deviance:     2.8209  on  4  degrees of freedom
AIC: 118.07
```

According to this log-linear model, the intercept is 9.02849 and the number of individuals who are not in any lists is 8337 since,

$$\log \mu_{00000} = \theta, \qquad \Rightarrow \qquad \hat{\mu}_{00000} = e^{\hat{\theta}} = e^{9.02849} = 8337.$$

By adding those 2744 cases that are already known, the total population size estimation is 11081. The estimated confidence interval for the population size is also computable. The standard error for $\hat{\mu}_{00000}$ is 777.94 as computed by the Delta method in R. Hence, the 95% confidence interval for $\hat{\mu}_{00000}$ is $(6812, 9862)$, as,

$$8337.261 - 1.96 \times 777.9498 = 6812.479,$$
$$8337.261 + 1.96 \times 777.9498 = 9862.043.$$

The 95% confidence interval for the population size is $(9556, 12606)$. The intercept and population estimates in this model are very close to the values ($\hat{\theta} = 9.05591, \hat{\mu}_{00000} = 8569$) obtained by Silverman [2014]. The cell mean estimates provided by this model are given in Table 5.4 and as mentioned before, non of the cell means with a zero observation are estimable.

cell	$y_{\mathbf{i}}$	$\hat{\mu}_i$	cell	$y_{\mathbf{i}}$	$\hat{\mu}_i$
y_{11111}	0	-	y_{01110}	0	-
y_{11110}	1	1.00	y_{01110}	4	2.48
y_{11101}	0	-	y_{01101}	0	-
y_{11100}	1	1.00	y_{01100}	62	56.00
y_{11011}	0	-	y_{01011}	0	-
y_{11010}	1	0.68	y_{01010}	19	20.82
y_{11001}	0	-	y_{01001}	1	1
y_{11000}	15	15.31	y_{01000}	463	468.00
y_{10111}	0	-	y_{00111}	0	-
y_{10110}	0	-	y_{00110}	76	82.15
y_{10101}	0	-	y_{00101}	11	11.00
y_{10100}	19	19.00	y_{00100}	995	996.34
y_{10011}	0	-	y_{00011}	8	8.00
y_{10010}	3	4.38	y_{00010}	695	687.49
y_{10001}	0	-	y_{00001}	316	316.00
y_{10000}	54	52.00	y_{00000}	-	-

Table 5.4 The cell counts and estimated cell means.

The last two interactions in this model have respectively a large p-value and a large standard error. We can eliminate them from the model and fit the model again, which provides the following estimates and slightly increases the AIC.

```
Coefficients:
     Estimate Std. Error z value Pr(>|z|)
A1   9.02981    0.09330  96.782  < 2e-16 ***
A2  -5.02762    0.15413 -32.620  < 2e-16 ***
A3  -2.88147    0.09513 -30.289  < 2e-16 ***
A4  -2.12556    0.08811 -24.124  < 2e-16 ***
A5  -2.49684    0.09143 -27.309  < 2e-16 ***
A6  -3.27407    0.10895 -30.052  < 2e-16 ***
A7   1.45705    0.27729   5.255 1.48e-07 ***
A8   0.94333    0.26131   3.610 0.000306 ***
A9  -0.57770    0.22407  -2.578 0.009931 **
A10 -2.87427    1.00609  -2.857 0.004278 **
A11 -1.23229    0.31912  -3.862 0.000113 ***
A12 -1.17946    0.36949  -3.192 0.001412 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 28711.232  on 18  degrees of freedom
Residual deviance:     8.579  on  6  degrees of freedom
AIC: 119.83
```

In this case, the estimate for the number of individuals who are not in any lists is 8348, since,

$$\log \mu_{00000} = \theta, \qquad \Rightarrow \qquad \hat{\mu}_{00000} = e^{\hat{\theta}} = e^{9.02981} = 8348.$$

By adding those 2744 cases that are already known, the total population size estimation is 11092. Although now the model includes the exact same parameters as the Silverman's model, their parameter estimates are slightly different. It may be the result of using the different R functions `closedp.MX` and `glm`.

5.2.3 Results

Silverman [2014] estimates the dark figure size between 7,000 to 10,000, which means the actual population size is from 10,000 to 13,000. Thus, the National Crime Agency is aware of 20% to 30% of the potential victims. As shown in Table 5.3, the interactions between LA and NG, and LA and PF are positive. It suggests that being known by local authorities increases the chance of victims to be known by NGOs and police forces. The interactions between GP and NG, and PF and GO are negative. There is also a negative interaction between NG and GO, which may reflect the fact of not sharing information on the potential victims between NGOs and government organisations [Silverman, 2014].

In our model with parameter estimates in Table 5.3, the dark figure and the total population size estimates are only a little smaller than what was predicted before. The variables interaction signs are the same as the Silverman's model. We have two more interactions in the model. One is a second-order interaction among variables LA, NG, FP which has a negative estimate, and the other one is the summation of two interactions LA*NG*PF*GO and LA*PF*GO, which has a positive estimate.

Instead of considering a saturated model with 32 parameters, we could begin to investigate finding the best model by considering an initial model only including the main effects and the firs-order interactions of variables which make 16 parameters. In this case, the rank of the derivative matrix would be 15, suggesting that only one parameter, θ^{AE} or interaction LA*GP, is not estimable. We assume that must be the same parameter mentioned by Silverman [2014] as the parameter which has a very high standard error.

5.3 Ear surgery outcome

5.3.1 The data

This example is taken from Brown and Fuchs [1983] described as "an example of errors that arise in fitting models to sparse tables". The data set presents a study of 118 patients who had the same ear surgery. The variables of interest are five binary variables related to the success of the surgery. They are shown as D,B,M,N,E and rated in two levels. For example, D indicates the dryness of the ear as dry (1) or not dry (0). The data is presented in Table 5.5 as a 2^5 contingency table. The table includes 13 sampling zeros.

5.3.2 Analysis

Two numerical methods, the Iterative Proportional Fitting (IPF) and the iteratively re-weighted Newton-Raphson algorithm (NR) are compared in fitting the log-linear model to the mentioned sparse table by Brown and Fuchs [1983]. They focus on comparing the efficiency of these two methods in estimating the expected cell counts, determining the degree of freedom of the model, and deriving the parameters estimates. The initial log-linear model chosen to fit the data is the hierarchical model $(DEB,DN,DM,ENMB)$. Choosing the best possible model is not of interest here as the aim is to show the effect of zero observations on the estimability of parameters. This model includes 22 parameters, an intercept, five main effects, ten first-order interactions of variables, five second-order interactions, and one third-order interaction. Since 13 cell counts are zero in the table, some of the model parameters are not estimable. The marginal table

D	B	M	N	E	Cell	$y_{\mathbf{i}}$
1	1	1	1	1	y_{11111}	33
				0	y_{11110}	32
			0	1	y_{11101}	8
				0	y_{11100}	8
		0	1	1	y_{11011}	0
				0	y_{11010}	1
			0	1	y_{11001}	1
				0	y_{11000}	0
	0	1	1	1	y_{10111}	0
				0	y_{10110}	1
			0	1	y_{10101}	0
				0	y_{10100}	0
		0	1	1	y_{10011}	0
				0	y_{10010}	1
			0	1	y_{10001}	0
				0	y_{10000}	0
0	1	1	1	1	y_{01111}	2
				0	y_{01110}	10
			0	1	y_{01101}	3
				0	y_{01100}	6
		0	1	1	y_{01011}	1
				0	y_{01010}	2
			0	1	y_{01001}	0
				0	y_{01000}	2
	0	1	1	1	y_{00111}	0
				0	y_{00110}	1
			0	1	y_{00101}	0
				0	y_{00100}	4
		0	1	1	y_{00011}	0
				0	y_{00010}	1
			0	1	y_{00001}	0
				0	y_{00000}	2

Table 5.5 Contingency table of observed frequencies.

corresponding to DEB contains two zeros and marginal table $ENMB$ contains four zeros. They cause eight cells to have expected frequencies equal to zero. A formula is given by Brown and Fuchs [1983] for specifying inestimable parameters, which utilises the number of zero cells in marginal subtables corresponding to the each configuration of parameters. By applying this formula the inestimable parameters in the model are reported as $\theta^{EB}, \theta^{DEB}, \theta^{ENB}, \theta^{EMB}, \theta^{ENMB}$ in the paper. The estimable parameters cause 24 cells to have non-zero mean estimates. The model's parameter estimates obtained by the IPF and NR methods with tolerance limit of 0.001 are presented in Table 5.6. The model's degree of freedom is $d.f = (32-8)-(22-5) = 7$, calculated by the number of non-zero cell means minus the number of estimable parameters. They mention that the number of estimable parameters is the rank of $\mathsf{A}^{\mathsf{T}}\mathsf{WA}$ in (1.8), but there might be a "problem of numerical accuracy" in the determination of it. Thus, the rank of this matrix must be the same as the rank of the derivative matrix, although calculating the later is quite straightforward.

We apply the parameter redundancy approach for this example. The unsaturated log-linear model $(DEB, DN, DM, ENMB)$ for the 2^5 table has 22 parameters and is as below,

$$\begin{aligned}\log \mu_{ijklm} =& \theta + \theta_i^E + \theta_j^N + \theta_k^M + \theta_l^B + \theta_m^D + \theta_{il}^{EB} + \theta_{mi}^{DE} + \theta_{ml}^{DB} + \theta_{mj}^{DN} + \theta_{mk}^{DM} + \theta_{ik}^{EM} \\ &+ \theta_{ij}^{EN} + \theta_{lk}^{BM} + \theta_{lj}^{BN} + \theta_{kj}^{MN} + \theta_{mil}^{DEB} + \theta_{ilk}^{EBM} + \theta_{lkj}^{BMN} + \theta_{ikj}^{EMN} + \theta_{lkj}^{BMN} \\ &+ \theta_{ilkj}^{EBMN}, \qquad i,j,k,l,m \in \{0,1\}^5.\end{aligned}$$

To find the estimable parameters, we build the design matrix for this model in `R`, which is a 32×22 matrix and its transpose gives the derivative matrix. After considering the 13 zero cell counts, the rank of the derivative matrix is $r = 17$ and the deficiency is $d = p - r = 22 - 17 = 5$. We can find those five $\boldsymbol{\alpha}$ vectors and then solve the corresponding differential equations using `Maple` to obtain all the estimable parameters. The $\boldsymbol{\alpha}$ vectors are,

$$\begin{aligned}\boldsymbol{\alpha}_1^{\mathsf{T}} &= (0,-1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0),\\ \boldsymbol{\alpha}_2^{\mathsf{T}} &= (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1,0,0,0,0,1),\\ \boldsymbol{\alpha}_3^{\mathsf{T}} &= (0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1,0,0,0,0,0,1,0),\\ \boldsymbol{\alpha}_4^{\mathsf{T}} &= (0,0,0,0,0,0,0,-1,0,0,0,0,0,0,0,0,0,0,1,0,0,0),\\ \boldsymbol{\alpha}_5^{\mathsf{T}} &= (0,0,0,0,0,0,-1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0).\end{aligned}$$

Parameter	Brown and Fuchs' model IPF method	Brown and Fuchs' model NR method	The reduced model IWLS
θ	0.189	-1.600	0.64891
θ^{E}	-0.294	-2.083	-
θ^{N}	0.080	0.111	-0.10303
θ^{M}	0.854	0.741	0.53062
θ^{B}	0.661	2.450	-0.13222
θ^{D}	-0.106	-0.061	-3.09602
θ^{EB}	-	1.789	-
θ^{DE}	0.307	0.352	-
θ^{DB}	0.381	0.337	1.45083
θ^{DN}	0.324	0.324	1.25508
θ^{DM}	0.340	0.340	1.62498
θ^{EM}	0.035	-0.078	-
θ^{EN}	-0.074	-0.043	-
θ^{BM}	0.438	0.551	0.90864
θ^{BN}	0.130	0.099	0.16796
θ^{MN}	-0.013	0.027	-0.97434
θ^{DEB}	-	-0.045	-
θ^{EBM}	-	0.113	-
θ^{EBN}	-	-0.031	-
θ^{EMN}	0.020	0.059	-
θ^{BMN}	0.204	0.164	1.20094
θ^{EBMN}	-	-0.040	-
$\theta^{E}+\theta^{EB}$			-1.03154
$\theta^{DE}+\theta^{DEB}$			1.24974
$\theta^{EM}+\theta^{EBM}$			-0.01258
$\theta^{EN}+\theta^{EBN}$			-0.76270
$\theta^{EMN}+\theta^{EBMN}$			0.52102

Table 5.6 Parameter estimates of the log-linear model $(DEB, DN, DM, ENMB)$.

Cell	Cell means estimates IPF	NR	IWLS
y_{11111}	32.3	32.3	31.3
y_{11110}	32.8	32.5	32.5
y_{11101}	8.5	8.7	8.5
y_{11100}	6.9	7.3	6.9
y_{11011}	0.8	-	0.7
y_{11010}	1.0	1.2	1.2
y_{11001}	0.5	1.0	0.4
y_{11000}	0.4	-	0.3
y_{10111}	0	-	0
y_{10110}	0.9	1.2	0.8
y_{10101}	0	-	0
y_{10100}	0.7	-	0.7
y_{10011}	0	-	0
y_{10010}	0.3	0.8	0.2
y_{10001}	0	-	0
y_{10000}	0.1	-	0.08
y_{01111}	2.7	2.7	2.6
y_{01110}	9.2	9.5	9.4
y_{01101}	2.5	2.3	2.4
y_{01100}	7.1	6.7	7.0
y_{01011}	0.2	1.0	0.2
y_{01010}	1.0	0.8	1.7
y_{01001}	0.5	-	0.5
y_{01000}	1.6	2.0	1.6
y_{00111}	0	-	0
y_{00110}	1.1	0.8	1.1
y_{00101}	0	-	0
y_{00100}	3.3	0.4	3.2
y_{00011}	0	-	0
y_{00010}	1.7	1.2	1.7
y_{00001}	0	-	0
y_{00000}	1.9	2.0	1.9

Table 5.7 Cell means estimates of the log-linear model $(DEB, DN, DM, ENMB)$.

Solving the partial differential equations (2.3) specifies the 17 estimable parameters of the model as,

$$\begin{aligned}\boldsymbol{\theta}'^{\mathsf{T}} =&(\theta, \theta^N, \theta^M, \theta^B, \theta^D, \theta^{DB}, \theta^{DN}, \theta^{DM}, \theta^{BM}, \theta^{BN}, \theta^{MN}, \theta^{BMN},\\ &\theta^E + \theta^{EB}, \theta^{EN} + \theta^{EBN}, \theta^{EM} + \theta^{EBM}, \theta^{DE} + \theta^{DEB}, \theta^{EMN} + \theta^{EBMN}).\end{aligned}$$

Due to having marginal zeros in the model, there does not exist any esoteric constraints to make it possible to obtain the MLEs for all the parameters. By checking the existence of the MLE approach for this example, we find out the MLE does not exist and the corresponding co-facial set as defined in (4.4) is,

$$\mathscr{F}^c = \{00001, 00011, 00101, 00111, 10001, 10011, 10101, 10111\}.$$

The set of estimable parameters and estimable combination of parameters makes 24 cell means to be estimable. Estimates for cells $y_{01001}, y_{10000}, y_{10100}, y_{11000}, y_{11011}$ are possible to be computed, although their observed frequencies are zero. The reduced model formed by the vector of estimable parameters ($\boldsymbol{\theta}'$), the vector of estimable cell means ($\boldsymbol{\mu}'$) and the corresponding design matrix (A') is shown as,

$$\log \boldsymbol{\mu}'_{24\times 1} = \mathsf{A}'_{24\times 17}\boldsymbol{\theta}'_{17\times 1},$$

with degree of freedom $d.f = 24 - 17 = 7$.

If we fit the log-linear model with all 22 parameters to the 32 cell counts, the estimates for some parameters have large standard errors which reveal the over-parametrization of the model.

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
A(Intercept)    0.6489     0.7086   0.916  0.35982
AE             -20.2576 10129.4383  -0.002  0.99840
AN             -0.1030      1.0056  -0.102  0.91839
AM              0.5306      0.8771   0.605  0.54520
AB             -0.1322      1.0070  -0.131  0.89554
AD             -3.0960      1.0783  -2.871  0.00409 **
AEN            -0.4998  15081.3500   0.000  0.99997
AEM            -1.3184  15337.8713   0.000  0.99993
AMN            -0.9743      1.3391  -0.728  0.46686
AEB            19.2261  10129.4384   0.002  0.99849
ABN             0.1680      1.3651   0.123  0.90208
ABM             0.9086      1.1657   0.779  0.43571
ADN             1.2551      0.4935   2.543  0.01099 *
ADM             1.6250      0.7999   2.032  0.04219 *
```

```
ADE             1.6560  11711.7642   0.000  0.99989
ADB             1.4508      0.8991   1.614  0.10660
AEMN            0.9249  20880.9104   0.000  0.99996
AEBN           -0.2629  15081.3501   0.000  0.99999
AEBM            1.3058  15337.8714   0.000  0.99993
ABMN            1.2009      1.6463   0.729  0.46571
ADEB           -0.4062  11711.7643   0.000  0.99997
AEBMN          -0.4038  20880.9105   0.000  0.99998
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 434.4815  on 32  degrees of freedom
Residual deviance:   9.0345  on 10  degrees of freedom
AIC: 110.06
```

Cell mean estimates for 32 cells are given in Table 5.7. The zero estimates are actually reported as very small numbers, for example $\hat{\mu}_{00001} = 3.048127 \times e^{-09}$.

However, the final model to fit the data is the reduced model. It produces the following estimates and reasonable standard errors. The estimates are also given in Table 5.6 to be comparable with the model by Brown and Fuchs [1983].

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
A(Intercept)    0.64891    0.70864   0.916  0.35982
AE+EB          -1.03154    1.25282  -0.823  0.41030
AN             -0.10303    1.00564  -0.102  0.91839
AM              0.53062    0.87711   0.605  0.54520
AB             -0.13222    1.00701  -0.131  0.89554
AD             -3.09602    1.07832  -2.871  0.00409 **
ADE+DEB         1.24974    0.54568   2.290  0.02201 *
ADB             1.45083    0.89908   1.614  0.10660
ADN             1.25508    0.49352   2.543  0.01099 *
ADM             1.62498    0.79985   2.032  0.04219 *
AEM+EBM        -0.01258    1.31908  -0.010  0.99239
AEN+EBN        -0.76270    1.69710  -0.449  0.65313
ABM             0.90864    1.16573   0.779  0.43571
ABN             0.16796    1.36513   0.123  0.90208
AMN            -0.97434    1.33913  -0.728  0.46686
AEMN+EBMN       0.52102    1.74792   0.298  0.76564
ABMN            1.20094    1.64630   0.729  0.46571
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for poisson family taken to be 1)
```

```
    Null deviance: 418.4815  on 24  degrees of freedom
Residual deviance:   9.0345  on  7  degrees of freedom
AIC: 100.06
```

Cell mean estimates for 24 estimable cells, from y_{00000} to y_{11111} are as below, which are the same as those given in Table 5.7 but now there is no cell mean estimated as zero.

```
          1           2           3           4           5
 1.91345667  1.72612098  3.25285967  1.10756268  1.67648004
          6           7           8           9          10
 0.59759386  1.78893106  0.29741739  7.07071911  2.48889065
         11          12          13          14          15
 9.46386979  2.61609809  0.08654333  0.27387902  0.74714033
         16          17          18          19          20
 0.89243732  0.32351996  0.40240614  1.21106894  0.70258261
         21          22          23          24
 6.92928089  8.51110935 32.53613021 31.38390191
```

5.3.3 Results

The primary focus of the work by Brown and Fuchs [1983] is on the estimability of parameters $\boldsymbol{\theta}$. The models from Brown and Fuchs [1983] in Table 5.6 include 32 cells and 22 parameters. Our reduced model contains only 24 cells with non-zero expected values and 17 estimable parameters and it makes use of the iteratively re-weighted least squares (IWLS) method, which is the default method in the `glm` function in `R`. The degree of freedom for all of those models is 7. In the IPF method output, the estimates for the main effect of variables E, D and for interactions of EN and MN are negative values. The parameters without reported estimates by this numerical method are,

$$\theta^{EB}, \theta^{DEB}, \theta^{ENB}, \theta^{EMB}, \theta^{ENMB}.$$

In the parameter redundancy approach, those parameters and five more parameters are known as not directly estimable and only some combinations of them are estimable as,

$$\theta^{E} + \theta^{EB}, \theta^{DE} + \theta^{DEB}, \theta^{EN} + \theta^{ENB}, \theta^{EM} + \theta^{EMB}, \theta^{EMN} + \theta^{EMNB}.$$

The estimates for the main effect of variables N, B, D, the interaction of MN and linear combination of interactions $E + EB$, $EM + EBM$ and $EN + EBN$ are negative values. In the NR method, all of the parameters have a reported estimate value. As an example, consider that according to the parameter redundancy θ^{E} and θ^{EB} are not individually estimable but according to the IPF method, θ^{E} is estimable. The NR method has split the IPF estimate of θ^{E} between θ^{E} and θ^{EB} and reported estimates for both of them.

The parameter redundancy approach identifies eight cells with inestimable means. The IPF method in the mentioned paper, specifies the same cells have the expected value of zero, which are,

$$\mu_{00001}, \mu_{00011}, \mu_{00101}, \mu_{00111}, \mu_{10001}, \mu_{10011}, \mu_{10101}, \mu_{10111}.$$

The estimate values for other cell means derived by the IPF and the IWLS method in the `glm` function are almost the same, as shown in Table 5.7. The NR method does not estimate any cell means with a corresponding zero observation, which makes 13 cells. Brown and Fuchs [1983] fit a logistic equivalence of the model using the Newton-Raphson method, which provides the same cell means as the IPF method but then only six of the parameters are estimated. However, Brown and Fuchs [1983] do not reduce the model containing zero cell counts and determine the parameter estimates only based on the numerical methods result.

Therefore, the two approaches (parameter redundancy and the method applied by Brown and Fuchs [1983]) which both primarily focus on the estimability of θs, identify the same degree of freedom for the model and the estimable cell means in the parameter redundancy matches the non-zero cell means in the model obtained by IPF method. However, the empirically derived formula by Brown and Fuchs [1983] to determine the inestimable parameters is difficult to calculate, as it requires forming the marginal sub-tables corresponding to all the configuration of variables in the model and then collapsing them over all the possible indices to find all the zeros in them. This formula may not work when the model is parameter redundant or the MLE does not exist but there is no zero marginal in the table. In this example, it provides a set of inestimable parameters which is only a part of the set of inestimable parameters derived by the parameter redundancy method. These points suggest that the different set of estimable θs could be explained as different parametrizations, similar to what occurred in Example 4.5 for the parameter redundancy and the EMLE methods. Although the numerical methods (IPF method and IWLS in `glm`) specify the right estimates for the estimable cell means, a comprehensive understanding of the model structure depends on identifying the estimable θs which is allowed by parameter redundancy method.

5.3.4 The numerical methods in fitting the log-linear models

For hierarchical log-linear models and even for more general log-linear models, different numerical methods are used to estimate the model's parameters. The Newton-Raphson method is a common method to optimise the log-likelihood function of the model, but it is a complex one since requires reversing a matrix and solving a set of equations

at each step. When the MLE exists, it eventually achieves a unique optimiser from any starting point for natural parameters. But it sometimes is not "feasible" when the model is high dimensional or in presence of zero observations [Agresti, 2002, Fienberg and Rinaldo, 2012b]. The iterative proportional fitting (IPF) method is a simple algorithm for calculating cell mean estimates which does not require matrix inversion and is based on keeping the estimated sufficient marginals equal to the observed ones. Since the optimisation occurs in the mean value space, the algorithm converges to a unique optimum regardless of the existence of the MLE [Fienberg and Rinaldo, 2012b]. However, this method has some deficiencies. The algorithm can be very slow to converge if the model is not decomposable, [Agresti, 2002, Fienberg and Rinaldo, 2012b] and it does not provide any indication of non-existent MLE. In the later case, monitoring the slow rate of the convergence [Fienberg and Rinaldo, 2012b] or noticing the zero cell mean estimates [Fienberg and Rinaldo, 2007] are the only ways of detecting the non-existent MLE. As a result, these methods are usually appropriate to estimate cell means but not to identify the number of estimable natural parameters of the model or to estimate them [Fienberg and Rinaldo, 2012b].

In the reviewed example here, Brown and Fuchs [1983] compare the Newton-Raphson and the IPF methods to fit a log-linear model to a sparse contingency table. The model $(DEB, DN, DM, ENMB)$ is non-decomposable because its corresponding graph has a chordless cycle of length more than three [Lauritzen, 1996]. The cell mean estimated by both methods are given in Table 5.7, in which the results based on the IPF and the IWLS methods match. Then identifying the estimable parameters and reducing the model is necessary to estimate the log-linear parameters, θ. Aston and Wilson [1984] comment on the Brown and Fuchs' [1983] paper and mention the importance of removing the cells with zero estimated values from the model. So for the same model, they fit 24 cells with 22 parameters by implementing the iteratively reweighted Newton-Raphson method. The cell mean estimates are almost the same as those already derived by the IPF and the IWLS in Table 5.7. However, since we know that the number of estimable parameters is 17 not 22, they obtain zero estimates for the same five θs that are not estimated by the IPF in Table 5.6. They declare these parameters as "extrinsically aliased" and that the algorithm can not detect all the redundancies while contracting the model's design matrix.

IWLS method is an application of the Newton-Raphson method [Fienberg and Rinaldo, 2007] which replaces the Hessian matrix of the procedure with the expectation of the Hessian matrix [Agresti, 2002]. It is applied by the `glm` function in `R` and seems to be a good numerical method, at least for the small dimensional tables, since it shows the non-existent MLE with numbers tending to zero for inestimable cell means and

distinguishes the parameters that are not directly estimable with large standard errors. Fienberg and Rinaldo [2007] compare the IWLS and the IPF methods' behaviour by using two examples of 2^3 and 3^3 contingency tables with non-existent MLE and positive marginals. The IWLS method converges to the extended MLE faster than the IPF, but the estimated parameters tend to "explode" when the MLE does not exist. Fienberg and Rinaldo [2005] suggest modifications to improve the Newton-Raphson procedure for larger models and eliminate the exploding behaviour. For an example of a $2^3 \times 3$ contingency table with a decomposable model and a marginal zero observation, the IPF converges in the first iteration and the only indication of the non-existent MLE is that some of the fitted values are zero [Fienberg and Rinaldo, 2007].

By applying the parameter redundancy approach, we aim to identify the cell means and parameters that are mathematically estimable given the observed data set, and then compute the point and interval estimates for them which needs proper standard errors.

5.4 A genome-wide association study of lung cancer

5.4.1 The data

A genome-wide association study of lung cancer is provided in Hung et al. [2008]. 317,139 single nucleotide polymorphisms (SNP) in chromosomes 6 and 15 are considered to perform genotyping. Each SNP is categorised at three levels of 0, 1 and 2 which indicate the number of their minor allele. The study sample consists of 4260 individuals from six European countries and age, gender, smoking status and country of origins for each individual are also provided. Removing the individuals from the data set with some missing information reduced the total number of them to 3841. 500 SNPs are chosen as the top ones which have the highest p-value in the test for association with lung cancer. This number of SNPs is decreased to 50 SNPs selected via applying profile regression by Papathomas et al. [2012]. Those 50 SNPs are again reduced to 12 and then to 3 most important ones by Papathomas and Richardson [2016] to compare all possible graphical models by using reversible jump MCMC. Their final aim is to investigate the presence of gene-environment with log-linear model comparison.

The Spearman's correlation among the chosen 50 SNPs is shown in Figure 5.1, which suggests about 19 groups of correlated variables. We select the following five uncorrelated SNPs as variables that will form the contingency table to estimate their main effects and their interactions effects on the cell means in a log-linear model.

- `rs7748167_C` (A),

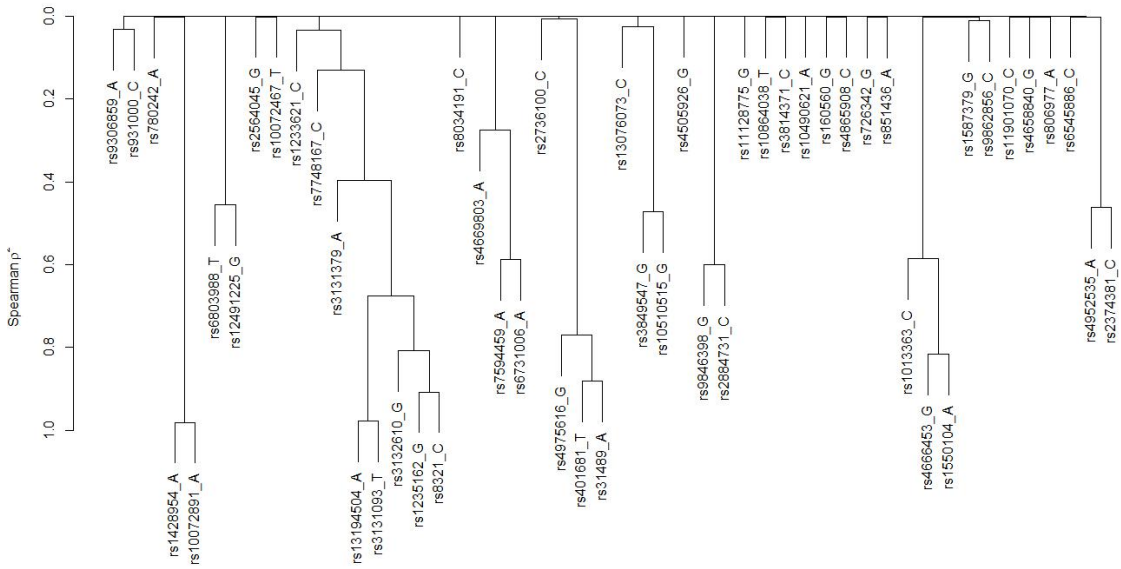

Fig. 5.1 Spearman correlation (ρ^2) for 50 SNPs.

- `rs4975616_G` (B),
- `s6803988_T` (C),
- `rs11128775_G` (D),
- `rs9306859_A` (E).

The observed cell counts for the 243 possible cross-classifications of the five variables is provided in Appendix B.

5.4.2 Analysis

We choose to fit a log-linear model with main effects, firs-order, and second-order interactions of the variables to the selected data in the 3^5 contingency table. The corresponding log-linear model, as shown in (1.5), has 243 cell counts and 131 parameters. The parameter vector is made of the intercept, 10 main effects, 40 parameters describing firs-order interactions and 80 parameters describing second-order interactions of variables.

The contingency table includes 132 sampling zero cells. To check the parameter redundancy, the design matrix and the derivative matrix must be formed. After involving the zero cells the rank of the derivative matrix is 95, indicating only 95 parameters or linear combinations of parameters is estimable in this model. The deficiency of the model is $d = p - r = 131 - 95 = 36$. As this model is bigger than the previous

examples, `Maple` is not able to find the $\boldsymbol{\alpha}$s and we use `MATLAB` to find them. Using the 36 $\boldsymbol{\alpha}$ vectors, given in Appendix B, and solving the corresponding differential equations in (2.3) specify the 95 estimable parameters as,

$$\begin{aligned}\boldsymbol{\theta}'^{\mathsf{T}} =(&\theta, \theta_1^A, \theta_2^A, \theta_1^B, \theta_2^B, \theta_1^C, \theta_2^C, \theta_1^D, \theta_2^D, \theta_1^E, \theta_2^E,\\ &\theta_{11}^{AB}, \theta_{21}^{AB}, \theta_{12}^{AB}, \theta_{22}^{AB}, \theta_{11}^{AC}, \theta_{21}^{AC}, \theta_{12}^{AC}, \theta_{22}^{AC}, \theta_{11}^{AD}, \theta_{21}^{AD}, \theta_{12}^{AD}, \theta_{11}^{AE}, \theta_{21}^{AE}, \theta_{11}^{BC}, \theta_{21}^{BC},\\ &\theta_{12}^{BC}, \theta_{22}^{BC}, \theta_{11}^{BD}, \theta_{21}^{BD}, \theta_{12}^{BD}, \theta_{11}^{BE}, \theta_{21}^{BE}, \theta_{12}^{BE}, \theta_{22}^{BE}, \theta_{11}^{CD}, \theta_{21}^{CD}, \theta_{12}^{CD}, \theta_{11}^{CE}, \theta_{21}^{CE}, \theta_{12}^{CE},\\ &\theta_{11}^{DE}, \theta_{21}^{DE}, \theta_{12}^{DE},\\ &\theta_{111}^{ABC}, \theta_{211}^{ABC}, \theta_{121}^{ABC}, \theta_{112}^{ABC}, \theta_{212}^{ABC}, \theta_{122}^{ABC}, \theta_{222}^{ABC}, \theta_{111}^{ABD}, \theta_{211}^{ABD}, \theta_{121}^{ABD}, \theta_{221}^{ABD}, \theta_{112}^{ABD},\\ &\theta_{111}^{ACD}, \theta_{121}^{ACD}, \theta_{221}^{ACD}, \theta_{112}^{ACD}, \theta_{111}^{BCD}, \theta_{211}^{BCD}, \theta_{121}^{BCD}, \theta_{221}^{BCD}, \theta_{112}^{BCD}, \theta_{22}^{BD} + \theta_{212}^{BCD}, \theta_{111}^{ABE},\\ &\theta_{211}^{ABE}, \theta_{121}^{ABE}, -\theta_{211}^{ACD} + \theta_{221}^{ABE}, \theta_{12}^{AE} + \theta_{112}^{ABE}, \theta_{12}^{AE} + \theta_{122}^{ABE}, \theta_{111}^{ACE}, \theta_{211}^{ACE}, \theta_{121}^{ACE},\\ &-\theta_{211}^{ACD} + \theta_{221}^{ACE}, \theta_{112}^{ACE}, \theta_{111}^{BCE}, \theta_{211}^{BCE}, \theta_{121}^{BDE}, \theta_{221}^{BCE}, \theta_{112}^{BCE}, \theta_{212}^{BCE},\\ &\theta_{22}^{CE} + \theta_{122}^{BCE}, \theta_{22}^{CE} + \theta_{222}^{BCE}, \theta_{111}^{ADE}, \theta_{211}^{ACD} + \theta_{211}^{ADE}, \theta_{111}^{BDE}, \theta_{211}^{BDE}, \theta_{112}^{BDE}, \theta_{111}^{CDE},\\ &\theta_{211}^{CDE}, \theta_{121}^{CDE}, \theta_{121}^{BDE} + \theta_{122}^{BCD} + \theta_{22}^{CD} + \theta_{221}^{CDE}, \theta_{112}^{CDE}).\end{aligned}$$

These parameters and linear combinations of parameters make 12 cell means estimable of those 132 cells with zero cells entries. Thus, 95 quantities and 123 $(111+12)$ cell means of the model are estimable. This leads to reducing the model to a smaller one with the corresponding design matrix A', shown as,

$$\log \boldsymbol{\mu}'_{123\times 1} = \mathsf{A}'_{123\times 95}\boldsymbol{\theta}'_{95\times 1},$$

with degree of freedom of $d.f = 123 - 95 = 28$.

After forming the new design matrix and fitting the model to the data for 123 cells, the parameter estimates obtained by the `glm` function in `R`, respectively to $\boldsymbol{\theta}'$, are provided without noticing large standard errors. There is also no esoteric constraint as defined in Section 2.7 to make all the cell means estimable. By checking the existence of the MLE approach for this example, we find out the MLE does not exist and the co-facial set as defined in (4.4) includes 120 cells of the contingency table.

```
Coefficients:
        Estimate Std. Error z value Pr(>|z|)
Ared1   6.07828    0.04716 128.897  < 2e-16 ***
Ared2  -1.82339    0.11934 -15.279  < 2e-16 ***
Ared3  -4.53620    0.44583 -10.175  < 2e-16 ***
Ared4   0.16896    0.06343   2.664  0.00773 **
Ared5  -1.21935    0.09783 -12.464  < 2e-16 ***
Ared6  -0.37684    0.07269  -5.184 2.17e-07 ***
Ared7  -1.78243    0.12288 -14.506  < 2e-16 ***
Ared8  -1.59376    0.10861 -14.674  < 2e-16 ***
Ared9  -4.82373    0.51943  -9.287  < 2e-16 ***
```

```
Ared10 -1.32967    0.09906 -13.423  < 2e-16 ***
Ared11 -4.08292    0.36755 -11.108  < 2e-16 ***
Ared12 -0.07037    0.15628  -0.450  0.65249
Ared13  0.21000    0.56512   0.372  0.71019
Ared14  0.33359    0.21823   1.529  0.12636
Ared15  0.08116    0.89559   0.091  0.92779
Ared16  0.04792    0.17080   0.281  0.77903
Ared17  0.49483    0.61812   0.801  0.42340
Ared18 -0.15757    0.31309  -0.503  0.61476
Ared19  0.04112    1.01581   0.040  0.96771
Ared20  0.37068    0.21186   1.750  0.08019 .
Ared21  0.85309    0.73472   1.161  0.24560
Ared22  1.48261    0.77232   1.920  0.05490 .
Ared23  0.15311    0.20989   0.729  0.46571
Ared24 -2.54539    1.50802  -1.688  0.09143 .
Ared25  0.04610    0.09620   0.479  0.63179
Ared26  0.27839    0.14142   1.969  0.04900 *
Ared27 -0.30850    0.17431  -1.770  0.07676 .
Ared28 -0.04093    0.25498  -0.161  0.87247
Ared29  0.29023    0.13556   2.141  0.03228 *
Ared30  0.04269    0.21694   0.197  0.84400
Ared31  0.59055    0.62017   0.952  0.34097
Ared32 -0.11480    0.13219  -0.868  0.38515
Ared33 -0.12873    0.20664  -0.623  0.53330
Ared34  0.24534    0.46937   0.523  0.60118
Ared35 -1.43964    1.22404  -1.176  0.23954
Ared36  0.29030    0.14980   1.938  0.05264 .
Ared37  0.07721    0.26479   0.292  0.77060
Ared38  0.62497    0.67600   0.925  0.35522
Ared39  0.17532    0.14144   1.239  0.21516
Ared40 -0.63914    0.30451  -2.099  0.03582 *
Ared41 -1.12068    0.81090  -1.382  0.16696
Ared42  0.14677    0.19061   0.770  0.44130
Ared43  0.76826    0.88180   0.871  0.38362
Ared44 -0.83987    1.10777  -0.758  0.44835
Ared45  0.14543    0.20746   0.701  0.48330
Ared46 -1.53105    0.98680  -1.552  0.12077
Ared47 -0.58927    0.30888  -1.908  0.05642 .
Ared48  0.51318    0.38280   1.341  0.18006
Ared49 -0.25941    1.24694  -0.208  0.83520
Ared50 -0.63577    0.62672  -1.014  0.31037
Ared51  0.68987    1.53772   0.449  0.65370
Ared52 -0.31688    0.23477  -1.350  0.17710
Ared53 -0.65127    0.92291  -0.706  0.48040
Ared54 -0.28173    0.36847  -0.765  0.44452
Ared55 -0.39438    1.46691  -0.269  0.78804
Ared56 -1.19327    1.01177  -1.179  0.23824
Ared57 -0.13365    0.23379  -0.572  0.56755
Ared58  0.09403    0.41723   0.225  0.82169
Ared59  0.88049    1.17506   0.749  0.45367
Ared60 -0.80018    1.02714  -0.779  0.43595
Ared61 -0.31080    0.18312  -1.697  0.08965 .
Ared62 -0.29014    0.28756  -1.009  0.31298
Ared63 -0.19269    0.34056  -0.566  0.57154
Ared64  0.12529    0.48819   0.257  0.79745
Ared65 -0.80282    0.84276  -0.953  0.34079
```

```
Ared66  0.82457    0.68772   1.199  0.23053
Ared67  0.09254    0.23888   0.387  0.69845
Ared68  1.96678    1.18112   1.665  0.09588 .
Ared69  0.08582    0.34530   0.249  0.80372
Ared70  5.12715    2.46393   2.081  0.03744 *
Ared71  0.40637    0.67952   0.598  0.54983
Ared72  3.06370    1.23149   2.488  0.01285 *
Ared73 -0.24453    0.23376  -1.046  0.29553
Ared74  2.17911    1.39025   1.567  0.11701
Ared75 -0.15251    0.44521  -0.343  0.73193
Ared76  2.56379    2.18522   1.173  0.24070
Ared77 -1.01258    1.26281  -0.802  0.42264
Ared78 -0.27004    0.18353  -1.471  0.14120
Ared79  0.18134    0.26602   0.682  0.49543
Ared80  0.70008    0.36774   1.904  0.05694 .
Ared81  1.10395    0.48428   2.280  0.02263 *
Ared82  0.62483    0.92740   0.674  0.50048
Ared83  2.27819    1.48319   1.536  0.12454
Ared84 -0.31874    1.05065  -0.303  0.76161
Ared85  2.48698    1.55267   1.602  0.10921
Ared86  0.25977    0.25687   1.011  0.31189
Ared87 -0.25149    1.30623  -0.193  0.84732
Ared88 -0.17091    0.21397  -0.799  0.42443
Ared89 -0.33016    0.33537  -0.984  0.32488
Ared90  0.94414    1.17431   0.804  0.42140
Ared91 -0.11082    0.21281  -0.521  0.60255
Ared92  0.22132    0.39026   0.567  0.57064
Ared93 -0.42344    1.21343  -0.349  0.72712
Ared94  1.38528    1.20246   1.152  0.24930
Ared95  0.60942    0.90605   0.673  0.50119
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 29437.659  on 123  degrees of freedom
Residual deviance:    36.452  on  28  degrees of freedom
AIC: 669.2
```

5.4.3 Results

The fitted reduced model records 17 of the estimable parameters significant at 0.05 level. Parameters $\theta, \theta_1^B, \theta_{21}^{BC}, \theta_{11}^{BD}, -\theta_{211}^{ACD} + \theta_{221}^{ABE}, \theta_{12}^{AE} + \theta_{122}^{ABE}, \theta_{221}^{BCE}$ have positive estimates, which indicate a positive influence on the logarithm of the cell means. The other 10 significant parameters, $\theta_1^A, \theta_2^A, \theta_2^B, \theta_1^C, \theta_2^C, \theta_1^D, \theta_2^D, \theta_1^E, \theta_2^E, \theta_{21}^{CE}$, have a negative effect on the logarithm of the cell means. The firs-order interactions are specifying a positive estimate for $B2C1$, $B1D1$ and a negative estimate for $C2E1$. The second-order interactions are indicating a positive estimate for $B2C2E1$.

5.4.4 Including the outcome variable

A crucial variable in this study is an outcome variable which describes the presence or absence of cancer in each of the 3841 individuals. By adding this variable (F), the $3^5 \times 2^1$ contingency table has 486 cells. To study the interactions between the 5 SNPs and the outcome variable, we only consider the main effects of variables and the firs-order interactions between them which make 62 parameters. Then the contingency table has 298 zero cell counts which makes a derivative matrix with the rank 59 and $d = 62 - 59 = 3$. It indicates that there are 59 estimable parameters in the model fitted to this sparse table, so the reduction in the number of estimable parameters is relatively small. After finding the three $\boldsymbol{\alpha}$ vectors, given in Appendix B, and solving the corresponding partial differential equations, the estimable parameters are,

$$\begin{aligned}\boldsymbol{\theta}^{'\mathsf{T}} =(&\theta, \theta_1^A, \theta_2^A, \theta_1^B, \theta_2^B, \theta_1^C, \theta_2^C, \theta_1^D, \theta_2^D, \theta_1^E, \theta_2^E, \theta_1^F \\ &\theta_{11}^{AB}, \theta_{21}^{AB}, \theta_{12}^{AB}, \theta_{22}^{AB}, \theta_{11}^{AC}, \theta_{21}^{AC}, \theta_{12}^{AC}, \theta_{22}^{AC}, \theta_{11}^{AD}, \theta_{21}^{AD}, \theta_{12}^{AD}, \theta_{11}^{AE}, \theta_{21}^{AE}, \theta_{12}^{AE}, \theta_{11}^{AF}, \theta_{21}^{AF}, \\ &\theta_{11}^{BC}, \theta_{21}^{BC}, \theta_{12}^{BC}, \theta_{22}^{BC}, \theta_{11}^{BD}, \theta_{21}^{BD}, \theta_{12}^{BD}, \theta_{22}^{BD}, \theta_{11}^{BE}, \theta_{21}^{BE}, \theta_{12}^{BE}, \theta_{11}^{BF}, \theta_{21}^{BF}, \theta_{22}^{BE}, \\ &\theta_{11}^{CD}, \theta_{21}^{CD}, \theta_{12}^{CD}, \theta_{22}^{CD}, \theta_{11}^{CE}, \theta_{21}^{CE}, \theta_{12}^{CE}, \theta_{22}^{CE}, \theta_{11}^{CF}, \theta_{21}^{CF}, \\ &\theta_{11}^{DE}, \theta_{21}^{DE}, \theta_{12}^{DE}, \theta_{11}^{DF}, \theta_{21}^{DF}, \theta_{11}^{EF}, \theta_{21}^{EF}).\end{aligned}$$

Thus, only three parameters $\theta_{22}^{AD}, \theta_{22}^{AE}, \theta_{22}^{DE}$ are not estimable due to the sparseness of the table. These estimable parameters make 360 out of 486 cell means estimable, so 62 cells with the observed zero counts have estimable cell means. The reduced model including only the estimable parameters and the estimable cell means is shown as,

$$\log \boldsymbol{\mu}'_{360\times 1} = \mathsf{A}'_{360\times 59}\boldsymbol{\theta}'_{59\times 1},$$

with degree of freedom of $d.f = 360 - 59 = 301$.

After forming the new design matrix and fitting the model to the data for 360 cells, the parameter estimates obtained by the `glm` function in `R`, respectively to $\boldsymbol{\theta}'$, are provided. There is no esoteric constraint as defined in Section 2.7 to make all the cell means estimable.

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
Ared1   5.42696    0.05344 101.548  < 2e-16 ***
Ared2  -1.87099    0.10154 -18.426  < 2e-16 ***
Ared3  -4.68509    0.37935 -12.350  < 2e-16 ***
Ared4   0.24632    0.06428   3.832 0.000127 ***
Ared5  -1.05975    0.09527 -11.123  < 2e-16 ***
Ared6  -0.22708    0.06828  -3.326 0.000882 ***
Ared7  -1.76466    0.12250 -14.406  < 2e-16 ***
Ared8  -1.56473    0.09055 -17.281  < 2e-16 ***
Ared9  -4.69328    0.38424 -12.215  < 2e-16 ***
```

```
Ared10 -1.21471    0.08512 -14.271  < 2e-16 ***
Ared11 -4.25712    0.36514 -11.659  < 2e-16 ***
Ared12 -0.14261    0.06606  -2.159 0.030849 *
Ared13 -0.02763    0.09801  -0.282 0.778034
Ared14 -0.14612    0.35632  -0.410 0.681742
Ared15  0.06729    0.14231   0.473 0.636311
Ared16 -0.20232    0.56314  -0.359 0.719398
Ared17 -0.04924    0.09579  -0.514 0.607235
Ared18 -0.29659    0.37785  -0.785 0.432497
Ared19 -0.01274    0.17707  -0.072 0.942629
Ared20  0.58813    0.50622   1.162 0.245316
Ared21  0.18004    0.10962   1.642 0.100507
Ared22  0.41781    0.37579   1.112 0.266215
Ared23  0.14501    0.45095   0.322 0.747779
Ared24  0.15718    0.10977   1.432 0.152157
Ared25 -0.02808    0.42447  -0.066 0.947259
Ared26  0.16294    0.41751   0.390 0.696345
Ared27  0.20548    0.09138   2.249 0.024534 *
Ared28  0.50831    0.33907   1.499 0.133836
Ared29 -0.06152    0.07335  -0.839 0.401571
Ared30  0.17252    0.10782   1.600 0.109569
Ared31 -0.13376    0.13574  -0.985 0.324402
Ared32  0.13169    0.19394   0.679 0.497106
Ared33  0.07692    0.08657   0.889 0.374255
Ared34 -0.16413    0.13455  -1.220 0.222502
Ared35 -0.12612    0.35626  -0.354 0.723332
Ared36 -0.27434    0.56294  -0.487 0.626018
Ared37 -0.19701    0.08663  -2.274 0.022955 *
Ared38 -0.02225    0.12506  -0.178 0.858785
Ared39  0.74971    0.37460   2.001 0.045351 *
Ared40  0.73081    0.49742   1.469 0.141777
Ared41 -0.08363    0.07035  -1.189 0.234537
Ared42 -0.28488    0.10567  -2.696 0.007016 **
Ared43  0.06397    0.08512   0.751 0.452366
Ared44  0.09772    0.15517   0.630 0.528858
Ared45  0.32838    0.34581   0.950 0.342317
Ared46 -0.13843    0.75185  -0.184 0.853925
Ared47  0.02720    0.08440   0.322 0.747230
Ared48 -0.08882    0.15927  -0.558 0.577072
Ared49 -0.49060    0.33870  -1.448 0.147481
Ared50 -0.74619    0.73645  -1.013 0.310954
Ared51 -0.15960    0.06895  -2.315 0.020633 *
Ared52 -0.32750    0.12950  -2.529 0.011439 *
Ared53  0.07618    0.09949   0.766 0.443855
Ared54 -0.23437    0.45044  -0.520 0.602841
Ared55 -0.21562    0.41686  -0.517 0.604977
Ared56  0.22181    0.08140   2.725 0.006429 **
Ared57  0.52895    0.33870   1.562 0.118356
Ared58 -0.20169    0.08221  -2.453 0.014151 *
Ared59 -0.72222    0.34331  -2.104 0.035406 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 24843.79  on 360  degrees of freedom
Residual deviance:   287.32  on 301  degrees of freedom
```

```
AIC: 1097.9
```

The significant coefficients at 0.05 level are all the main effects and the following interactions,

$$\begin{aligned}
\theta_{11}^{AF} &= 0.205, & \theta_{11}^{BE} &= -0.197, & \theta_{12}^{BE} &= 0.749,\\
\theta_{21}^{BF} &= -0.284, & \theta_{11}^{CF} &= -0.159, & \theta_{21}^{CF} &= -0.327,\\
\theta_{11}^{DF} &= 0.221, & \theta_{11}^{EF} &= -0.201, & \theta_{21}^{EF} &= -0.722.
\end{aligned}$$

Thus, the presence of cancer has a positive interaction with the level 1 of variables A and D. It has a negative interaction with the level 1 of variables C, E and the level 2 of variables B, C, E.

Chapter 6

Discussion

6.1 Conclusion

Sampling zero observations can cause problems in fitting a log-linear model. Chapter 2 suggested the parameter redundancy method as an approach to check the changes created by zero entries in the log-linear model. If the number of zero observations is enough to make the model parameter redundant, then the method was described to detect all the estimable model parameters and cell means and reduce the model to a smaller identifiable one. Small contingency tables were used to illustrate the idea and also to demonstrate the reasons for choosing Poisson distribution and corner point constraints to construct the model. The esoteric constraints were introduced as hidden constraints imposed on the model by the likelihood function that turn all the parameters estimable when the model is parameter redundant but none of the sufficient statistics is zero and the MLE exists for the cell means.

In Chapter 3, we pursued a general manner towards specification of inestimable model parameters. In the case of positivity of all the cell counts, the saturated model fitted to an l^m table was proved to be full rank. Afterwards, we proved that exactly which model parameters are turned to inestimable by observing a zero entry in each specific cell. The model was assumed to be saturated, as many different configurations of unsaturated models produce different results.

An alternative way to investigate the effect of sampling zero observations in a contingency table is exploring the existence of the maximum likelihood estimates for the hierarchical log-linear model's cell means. This approach was explained in Chapter 4 and was referred to as the existence of the MLE (EMLE) method. Considering a polyhedral resolution by defining an equivalent cone for the log-linear model is the base of this method which divides the models into two groups: the group with the existent MLE and the group with the non-existent MLE that must get reduced. The theoretical methods and the results of this approach were compared to the parameter

redundancy approach. We mentioned that parameter redundant models with esoteric constraints are classified by the EMLE approach as models with the existent MLE, without acknowledging the present extra relations among the parameters. Both the methods have strengths and weaknesses regarding their computational capacities, which are discussed in Section 6.2.

Examples of fitting log-linear models to real data were given in Chapter 5. Saturated and unsaturated models with a different number of variables, levels, and zero patterns were investigated for parameter redundancy. Section 5.3.4 briefed some different numerical methods used for estimating log-linear model parameters and cell means and mentioned how they behave in the presence of zero observations. The motivation in investigating parameter redundancy in log-linear models is to determine the point estimates and confidence intervals for as many model parameters as possible. Although a numerical method like IWLS specifies the cell means with zero estimates, the parameter redundancy method reveals how the log-linear model is changed by zero observations and which cell means and model parameters are technically estimable. This aim is also achievable (with a different parametrization of the model) by applying the existence of the MLE procedure. However, when the model is parameter redundant with existent MLE, the parameter redundancy provides more information about the model by obtaining the esoteric constraints.

6.2 Computational aspects

In Section 4.1.2, we summarised the polyhedral approach toward determining the existence of the MLE and mentioned the linear procedures aimed to find the co-facial set of the marginal cone or polytope. The MLE exists if and only if the observed margin $\mathbf{t}$ lies in the relative interior of the marginal cone and the MLE does not exist if and only if $\mathbf{t}$ belongs to the relative interior of some proper face F of the marginal cone. Wang et al. [2016] mention that "solving a sequence of linear programming problems" to find the co-facial set or the extreme points of the cone does not work when the number of variables in the hierarchical model is larger than 16. Massam and Wang [2015] consider less than 16 variables and describe how to find an outer approximation for F. Wang et al. [2016] show how to find an inner approximation of F and explain that if these two approximations are the same then the face is determined. They apply this methodology to larger hierarchical log-linear models, for example, a model with 16 binary variables and 314 parameters and a model with 100 binary variables and 277 parameters.

The parameter redundancy approach described in Chapter 2 has four main steps. First, constructing the design matrix and the derivative matrix for the desired model.

We do it in `R` for an l^m model and it could also be done for an $l_1^{m_1} \times l_2^{m_2} \times \dots$ model with some justification in the function. For example, the matrix can be computed for 20 binary variables with 6196 parameters up to the 4th-order interactions, but allocating the matrix is not possible when the model is bigger and the matrix size surpasses about 100 GB (on a computer with 3.1 GHz processor and 256 GB memory). Second, we set the corresponding zero columns in the derivative matrix and calculate the rank of the matrix and its null spaces by using `MATLAB`. The third step is solving the corresponding partial differential equations to derive the estimable parameters, which is done in `Maple` as it solves the differential equations in a symbolic way. When the number of equations d gets as large as 40, `Maple` can not solve them simultaneously. The number of equations mainly depends on the number and position of zero cells rather than the model size. The last step is finding the esoteric constraints which requires symbolic calculations applied by `Maple` and can get complicated for large models.

6.3 Future work

We investigated the log-linear model in this thesis, which does not distinguish between explanatory and response variables. If a categorical variable depends on the other model variables, then it is treated as a response variable and the others are explanatory variables. The obtained model is a logit model and it is equivalent to a certain log-linear model for that response variable. For example, the equivalent logit model for model (XY, XZ, YZ) defined in (4.1) with Y as a response variable, is,

$$\text{logit}\,[P(Y=1|X=i,Z=k)] = \alpha + \beta_i^X + \beta_k^Z.$$

The association parameter θ_{ik}^{XZ} of the log-linear model is not included in this model, as it cancels in the difference in logarithms the logit defines [Agresti, 2002]. This correspondence between the two models could be helpful in estimating some parameters when the contingency table is sparse, as is mentioned in Brown and Fuchs [1983]. Investigating the parameter redundancy in logit models for sparse tables by considering their correspondence with log-linear models could be studied. Another legitimate model to fit the count data is a Poisson regression model. It is similar to a log-linear model and when a large proportion of data is zero counts then zero-inflated Poisson distribution is applied which allows for overdispersion. This model could be studied by the parameter redundancy approach to specify the effect of the zero observations on estimability of the parameters.

Improving the parameter redundancy procedure by overcoming computational problems is a part of the future work to be able to use the method for larger models. For

a parameter redundant model, we specify the estimable cell means after determining the estimable parameters and combinations of parameters. Although it is not a difficult task for relatively small tables and the estimable cell means can be identified by monitoring the numerical method output for larger models, having a routine process would be helpful in identifying the estimable cell means and also in forming the reduced design matrix to match the estimable combinations of the parameters.

Models discussed in Section 4.3 can be studied further in terms of the importance of the esoteric constraints and possible ways of dealing with these models. Although the suggested Model (4.11) provides a better fit to the data compared to model (4.9), it is constructed only based on having the same number of parameters and cells as the log-linear model (4.9). One issue with this model is that of interpretability, as the standard log-linear models have a natural interpretability with regard to the parameters and the interactions between the variables. Another question raised regarding these models is about the degree of freedom of them and whether the esoteric constraints decrease the number of free parameters in the model.

Further studies can investigate the parameter redundancy concept in a Bayesian context. However, the meaning of identifiability concept differs in frequentist and Bayesian frameworks. Non-identifiability is not considered a strong Bayesian issue [Almond, et al., 2015, Rao and Dey, 2005] since as long as a proper prior distribution is defined for a parameter, the posterior distribution is proper as well and the parameter is thus estimable. In a Bayesian context, obtaining the proper posterior estimates of parameters for a parameter redundant model is possible even with choosing a uniform prior distribution for the parameters and it is due to the known orientation of the flat ridge [Cole, et al., 2010]. But for an inestimable parameter in the model, its prior and posterior distributions may occur to be almost identical and it means the estimates are very sensitive to the prior distribution [Almond, et al., 2015]. Similar prior and posterior distributions imply Bayesian learning issues. Bayesian learning is accomplished when the prior distribution differs from the posterior distribution, revealing the fact that the data have led us from the prior distribution toward the posterior distribution and have changed our knowledge about the distribution of the parameters [Lee, 2011]. To measure the information gained in the process of moving from the prior to the posterior, one way is to compute the Kullback–Leibler distance [Green, et al., 2003].

Bibliography

Agresti, A. (2002). Categorical data analysis, Second Edition. John Wiley and Sons publication.

Almond, R. G., Mislevy, R. J., Steinberg, L., Yan, D., Williamson, D. (2015). Bayesian networks in educational assessment. Springer.

Aston, C. E., Wilson, S. R. (1984). Comment on M.B. Brown, C. Fuchs, "On maximum likelihood estimation in sparse contingency table". *Computational Statistics and Data Analysis*, **2**, 71–77.

Baker. R.J., Clarke. M. R. B., Lane. P. W. (1985). Zero entries in contingency tables. *Computational Statistics and Data Analysis*, **3**, 33–45.

Bartlett, M. S. (1935). Contingency table interactions. *Supplement to the Journal of the Royal Statistical Society*, **2 (2)**, 248–252.

Birch, M. W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society*, Series B, **25 (1)**, 220–233.

Bishop, Y. M. M., Fienberg, S. E., Holland, P. W. (1975) Discrete multivariate analysis, theory and practice. MIT Press. Reprinted by Springer, 2007.

Brooks, S. P., Catchpole, E. A., Morgan, B. J.T. Barry, S. C. (2000). On the Bayesian analysis of ring-recovery data. *Biometrics*, **56**, 951–956.

Brown, M. B., Fuchs, C. (1983). On Maximum likelihood estimation in sparse contingency table. *Computational Statistics and Data Analysis*, **1**, 3–15.

Catchpole, E. A., Morgan, B. J. T. (1997). Detecting parameter redundancy. *Biometrika*, **84**, 187–196.

Catchpole, E. A., Morgan, B. J. T., Freeman, S. N. (1998). Estimation in parameter redundant models. *Biometrika*, **85**, 462–468.

Catchpole, E. A., Morgan, B. J. T. (2001). Deficiency of parameter redundant models. *Biometrika Trust*, **88 (2)**, 593–598.

Chappell, M. J., Gunn, R. N. (1998). A procedure for generating locally identifiable reparameterisations of unidentifiable non-linear systems by the similarity transformation approach. *Mathematical Biosciences*, **148**, 21–41.

Choquet, R., Cole, D. J. (2012). A hybrid symbolic-numerical method for determining model structure. *Mathematical Biosciences*, **236**, 117–125.

Cole, D. J., Morgan, B. J. T., Titterington, D. M. (2010). Detecting the parametric structure of models. *Mathematical Biosciences*, **228**, 16–30.

Cole, D. J., Morgan, B. J. T., Catchpole, E. A., Hubbard, B. A. (2012). Parameter redundancy in mark-recovery models. *Biometrical Journal*, **54**, 507–523.

Delforge, J. (1989). Relations between the main approaches to linear system identifiability: application to calculation of jacobian matrices determinants. *International Journal of Systems Science*, **20**, 1079–1097.

Dutour, M. (2008). *Computational methods for cones and polytopes with symmetry*. arXiv:math/0201110v1.

Earl, R. (2003). Induction. *Lecture notes*, Mathematical Institute, Oxford.

Eriksson, N., Fienberg, S. E., Rinaldo, A., Sullivant, S. (2006) Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models. *Journal of Symbolic Computation*, **41**, 222–233.

Evans, N. D., Chappell, M. J. (2000). Extensions to a procedure for generating locally identifiable reparametrisations of unidentifiable systems. *Mathematical Biosciences*, **168**, 137–159.

Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete 2k contingency table *Biometrika*, **59 (3)**, 591–603.

Fienberg, S. E., Rinaldo, A. (2005). Computing maximum likelihood estimation in log-linear models. Carnegie Mellon University. http://www.stat.cmu.edu/tr/tr835/tr835.pdf.

Fienberg, S. E., Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, **137**, 3430–3445.

Fienberg, S. E., Rinaldo, A. (2012a). Maximum likelihood estimation in log-linear models. *The Annals of Statistics*, **40 (2)**, 996–1023.

Fienberg, S. E., Rinaldo, A. (2012b). Maximum likelihood estimation in log-linear models, Supplementary material: Algorithms. http://www.stat.cmu.edu/~arinaldo/Fienberg_Rinaldo_Supplementary_Material.pdf.

Gentle, J. E. (2007). Matrix algebra, theory, computations, and applications in statistics. Springer.

Gimenez, O., Viallefont, A. , Catchpole, E. A., Choquet, R., Morgan, B. J. T. (2004). Methods for investigating parameter redundancy. *Animal Biodiversity and Conservation*, **27**, 1–12.

Green, P. J., Hjort, N. L., Richardson, S. (2003). Highly structured stochastic systems. Volume 27 of Oxford Statistical Science Series. Oxford University Press.

Goodman, L.A. (1970). The multivariate analysis of qualitative data: interactions among multiple classifications. *Journal of the American Statistical Association*, **65 (329)**, 226–256.

Goodman, L.A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables. *Journal of the American Statistical Association*, **66 (334)**, 339–344.

Goodman, L. A. (1974). Exploratory latent structure analysis using bothidentifiable and unidentifiable models. *Biometrika*, **61**, 215–23.

Haberman, S. J. (1973). Log-linear models for frequency data: Sufficient statistics and likelihood equations. *The Annals of Statistics*, **1 (4)**, 617–632.

Haberman, S. J. (1974). The analysis of frequency data. University of Chicago press.

Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Chen, C., Goodman, G., Field, J.K., Liloglou, T., Xinarianos, G., Cassidy, A., McLaughlin, J., Liu, G., Narod, S., Krokan, H.E., Skorpen, F., Elvestad, M.B., Hveem, K., Vatten, L., Linseisen, J., Clavel-Chapelon, F., Vineis, P., Bueno-de-Mesquita, H.B., Lund, E., Martinez, C., Bingham, S., Rasmuson, T., Hainaut, P., Riboli, E., Ahrens, W., Benhamou, S., Lagiou, P., Trichopoulos, D., Holc´atov´a, I., Merletti, F., Kjaerheim, K., Agudo,

A., Macfarlane, G., Talamini, R., Simonato, L., Lowry, R., Conway, D.I., Znaor, A., Healy, C., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M., Brennan P. (1973). A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*, **452**, 633–637.

Johndrow, J. E., Bhattacharya, A.l Dunson, D. (2014). Tensor decompositions and sparse log-linear models. arXiv:1404.0396v1.

Lauritzen, S. L. (1996). Graphical models. Oxford Statistical Science Series 17, Oxford University Press.

Lee, S. (2011). Handbook of latent variable and related models. Volume 1 of Handbook of Computing and Statistics with Applications. Elsevier.

Massam, H., Wang, N. (2015). A local approach to estimation in discrete loglinear models. arXiv:1504.05434.

McCullagh, P., Nelder, J. A. (1989). Generalized linear models, Second Edition. Chapman and Hall/CRC.

Overstall, A. M., King, R. (2014). conting: An R package for Bayesian analysis of complete and incomplete contingency tables. *Journal of Statistical Software*, **58 (7)**, 1–26.

Overstall, A. M., King, R., Bird, S. M., Hutchinson, S. J., Hayf, G. (2014). Incomplete contingency tables with censored cells with application to estimating the number of people who inject drugs in Scotland. *Statistics in Medicine*, **33 (9)**, 1564–1579.

Papathomas, M., Molitor, J., Hoggart, C., Hastie, D., Richardson, S. (2012). Exploring data from genetic association studies using Bayesian variable selection and the Dirichlet process: Application to searching for gene × gene patterns. *Genetic Epidemiology*, **36**, 663–674.

Papathomas, M., Richardson, S. (2016). Exploring dependence between categorical variables: Benefits and limitations of using variable selection within Bayesian clustering in relation to log-linear modelling with interaction terms. *Journal of Statistical Planning and Inference*, **173**, 47–63.

Pohjanpalo, H. (1982). Identifiability of deterministic differential models in state space. *Technical report, Research Centre of Finland Report, No. 56*.

Rao, C. R., Dey, D. K. (2005). Bayesian thinking, modeling and computation. Volume 25 of Handbook of Statistics. Gulf Professional Publishing.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, **39 (3)**, 577–591.

Silverman, B. (2014). Modern Slavery: an application of multiple systems estimation. https://www.gov.uk/government/publications/modern-slavery-strategy.

Silvey, S. D. (1975). Statistical inference. Chapman and Hall/CRC.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association*, **81**, 142–149.

Thowsen, A. (1978). Identifiability of dynamic systems. *International Journal of Systems Science*, **9**, 813–825.

Wang, N., Rauhyand, J., Massam, H. (2016). Approximating faces of marginal polytopes in discrete hierarchical models. arXiv:1603.04843.

Ziegler, G. M. (1995). Lectures on polytopes. Graduate Texts in Mathematics. Springer-Verlag.

Appendix A

Computer code

Chapter 2

• R code to fit model (2.5) to the data in Table 2.2:

```
y <- c(456,538,44,911)
A <- matrix(c(1,0,0,0,1,1,0,0,1,0,1,0,1,1,1,1),nrow=4,ncol=4,byrow=T)
model <- glm(formula=y~A-1, family=poisson)
summary(model)
predict(model,type="response")
```

• WinBUGS code to fit model (2.5) to the data in Table 2.2:

```
model{
# Model's likelihood
for (i in 1:4){
         n[i] ~ dpois( mu[i])
         mu[i] <- exp(logmu[i])
          }
         logmu[1] <- theta[1]
         logmu[2] <- theta[1]+ theta[2]
         logmu[3] <- theta[1]+ theta[3]
         logmu[4] <- theta[1]+ theta[2]+theta[3]+theta[4]
# Prior specification
   for (i in 1:4){
       theta[i]~dnorm(0,0.0001)
  }
}
# Data
list(n=c(456,538,44,911))
# Initial parameters values
list(theta=c(0,0,0,0))
list(theta=c(10,10,10,10))
list(theta=c(-10,-10,-10,-10))
```

• R code to fit model (2.7) to the data in Table 2.2:

```
y <- c(456,538,44,911)
A <- matrix(c(1,1,1,1,1,-1,1,-1,1,1,-1,-1,1,-1,-1,-1),nrow=4,
            byrow=T)
model <- glm(formula=y~A-1, family=poisson)
summary(model)
predict(model,type="response")
```

• The R function to make the design matrix and the derivative matrix for a specified model with m variables and l levels for each of them:

```
Dmatfor <- function(m,l,formula){
  le <- l-1
  li <- rep(list(0:le), m)
  vars <- expand.grid(li)
  # Making a dataframe
  d.f <- model.frame(vars)
  # Changing the names of columns to A,B,C...
  lett <- LETTERS[1:m]
  colnames(d.f) <- lett
  # Adding the cell counts vector (y) with length of l^m to the
    dataframe
  y.data <- paste0("y", 1:l^m)
  d.f$new.col <- y.data
  colnames(d.f)[m+1] <- "y"
  # Factorizing the levels
  nc <- ncol(d.f)-1
  for (i in 1:nc){
  d.f[,i] <- factor(d.f[,i])
  }
  attach(d.f)
  # Making the model of data frame vectors
  mf <- model.frame(formula=formula)
  # Making the design matrix
  X <- model.matrix(attr(mf, "terms"), data=mf)
  # Make the derivative matrix as the transpose of the design
    matrix
  Dm <- t(X)
  return(Dm)
}
```

• Three Maple procedures are provided here. DmatY takes the derivative matrix for a model with m variables and l levels for each variable and transforms it to the derivative matrix with y_is and zero columns corresponding to zero cell observations.

```
DmatY := proc (m, l, DR, a)
 local p, Y, i, j, ans;
```

```
 description "Converting the derivative matrix from function
              Dmatfor in R to matrix DR with ys and 0.
              a is the vector of zero cells, 1<a_i<n=l^m";
 p := RowDimension(DR);
 Y := Vector(l^m, symbol = y);
 Y[a] := 0;
 for j to l^m do
  for i to p do
    DR[i, j] := Y[j].DR[i, j]:
  end do
 end do;
 DR;
end proc:
```

• The `Estpars` procedure, which is mostly the work of Cole et al. [2010], takes the derivative matrix from `DmatY`, produces the $\boldsymbol{\alpha}$ and corresponding partial differential equations and then solves them. The procedure's output is the rank of the model, the deficiency of the model, the vector of estimable parameters and $\boldsymbol{\alpha}$ vectors.

```
Estpars := proc (DD1, pars)
 local r, d, alphapre, alpha, PDE, FF, i, ans, x, j;
 description "Finding the estimable set of parameters";
 with(LinearAlgebra);
 r := Rank(DD1);
 d := Dimension(pars)-r;
 alphapre := NullSpace(Transpose(DD1));
 if NullSpace(Transpose(DD1)) = {} then print('Model is full rank')
   else
   alpha := Matrix(d, Dimension(pars));
   PDE := Vector(d);
   FF := f(seq(pars[i], i = 1 .. Dimension(pars)));
   for i to d do
     alpha[i, 1 .. Dimension(pars)] := alphapre[i];
     PDE[i] := add((diff(FF, pars[j]))*alpha[i,j],
                    j=1..Dimension(pars)):
   end do;
   ans := pdsolve({seq(PDE[i] = 0, i = 1 .. d)});
   <r, d, ans, {alpha}>:
 end if;
end proc:
```

• The `EsoCon` procedure takes the vector of cell counts, the model's design matrix, the vector of parameters and the $\boldsymbol{\alpha}$ vectors. The output is $\boldsymbol{\alpha}^T U(\boldsymbol{\theta})$ and the esoteric constraints are determined by setting it to zero.

```
EsoCon := proc (y, X, pars, alpha)
```

```
local tmar, XP, n, XPE, p, U, ltheta, i;
description "Give alpha*U to specify if there exists any esoteric
             constraints";
tmar := Transpose(X).y;
XP := X.pars;
n := Dimension(y);
XPE := Vector(n, 0);
 for i to n do
   XPE[i] := exp(XP[i]):
 end do;
ltheta := Transpose(tmar).pars-Transpose(Vector(n, 1)).XPE;
p := RowDimension(pars);
U := Vector(p, 0);
 for i to p do
   U[i] := diff(ltheta, pars[i]):
 end do;
Transpose(alpha).U;
end proc:
```

Appendix B

Plots and Data

Chapter 2

• Trace and density plots derived by `WinBUGS` to fit model (2.5) to the data in Table 2.2 in section 2.4.1.

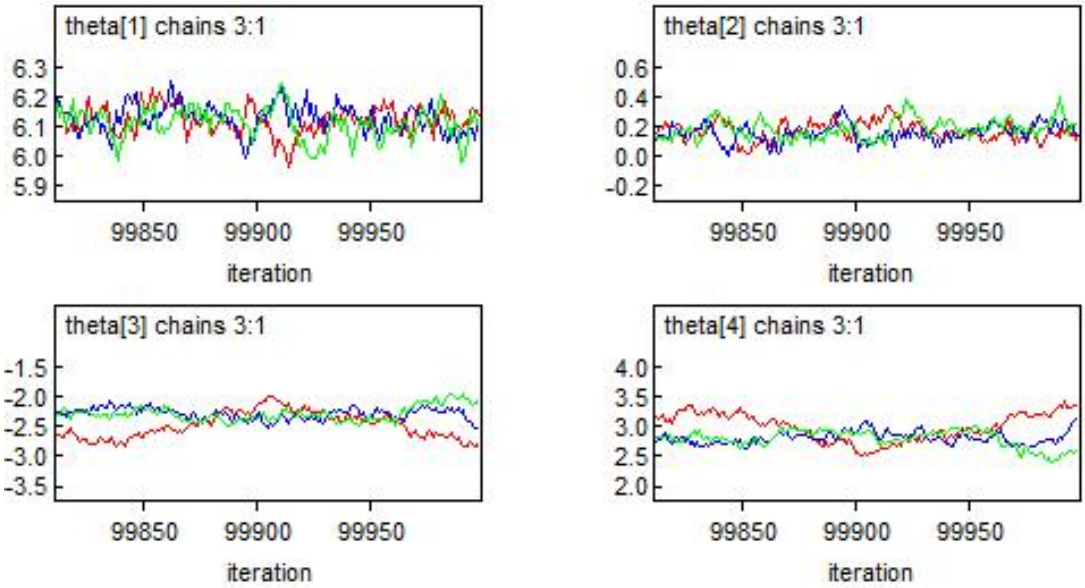

Fig. B.1 Trace plots for the parameters $\boldsymbol{\theta}$, if all observations are positive

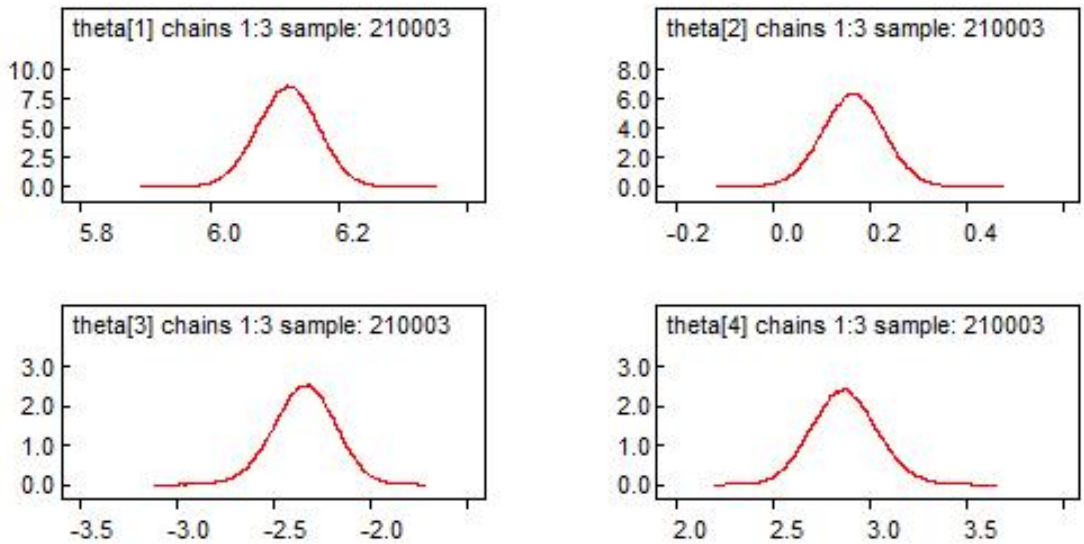

Fig. B.2 Marginal density plots for the parameters $\boldsymbol{\theta}$, if all observations are positive

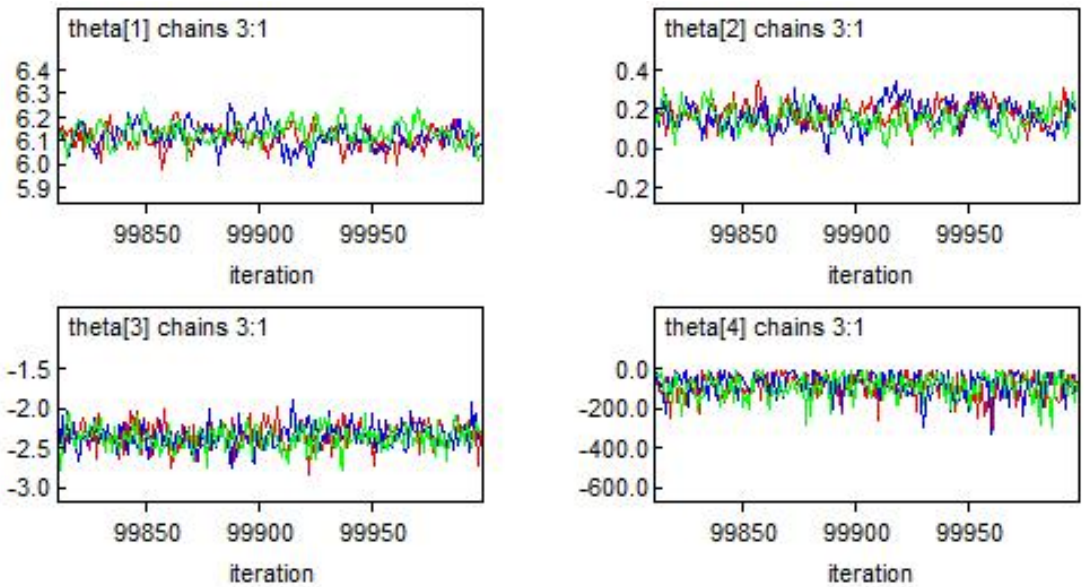

Fig. B.3 Trace plots for the parameters $\boldsymbol{\theta}$, if $y_4 = 0$

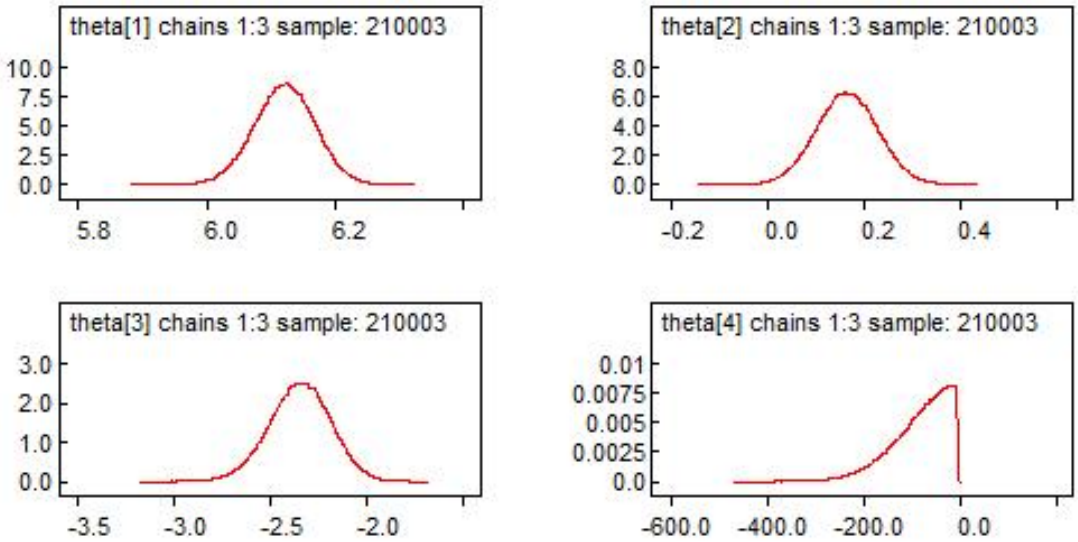

Fig. B.4 Marginal density plots for the parameters $\boldsymbol{\theta}$, if $y_4 = 0$

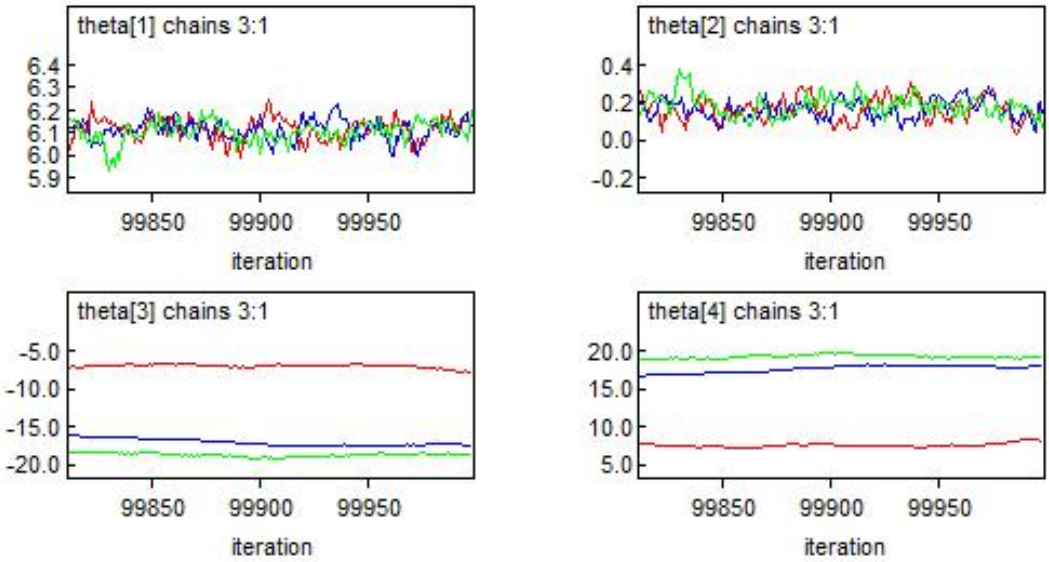

Fig. B.5 Trace plots for the parameters $\boldsymbol{\theta}$, if $y_3 = 0$

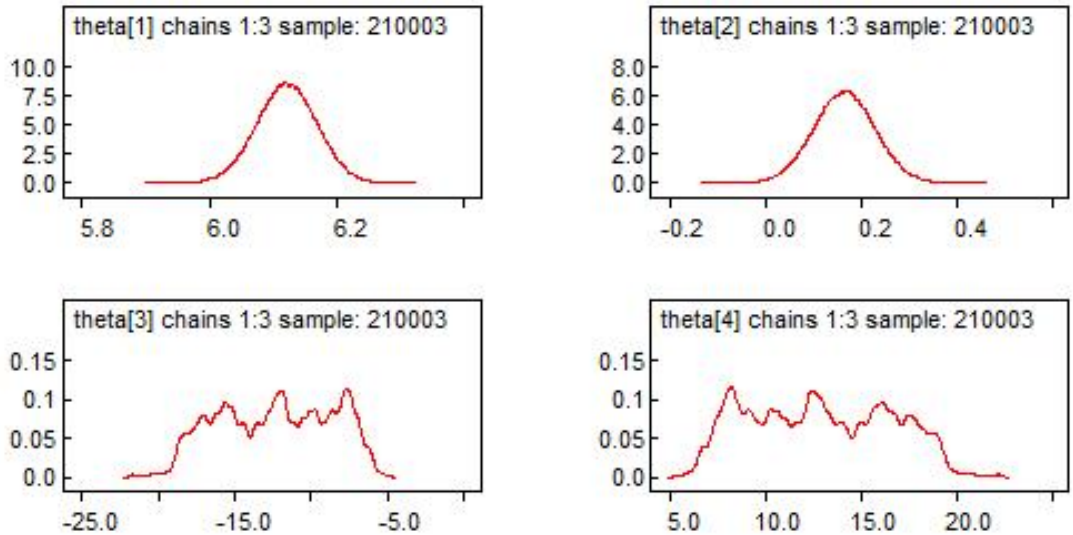

Fig. B.6 Marginal density plots for the parameters $\boldsymbol{\theta}$, if $y_3 = 0$

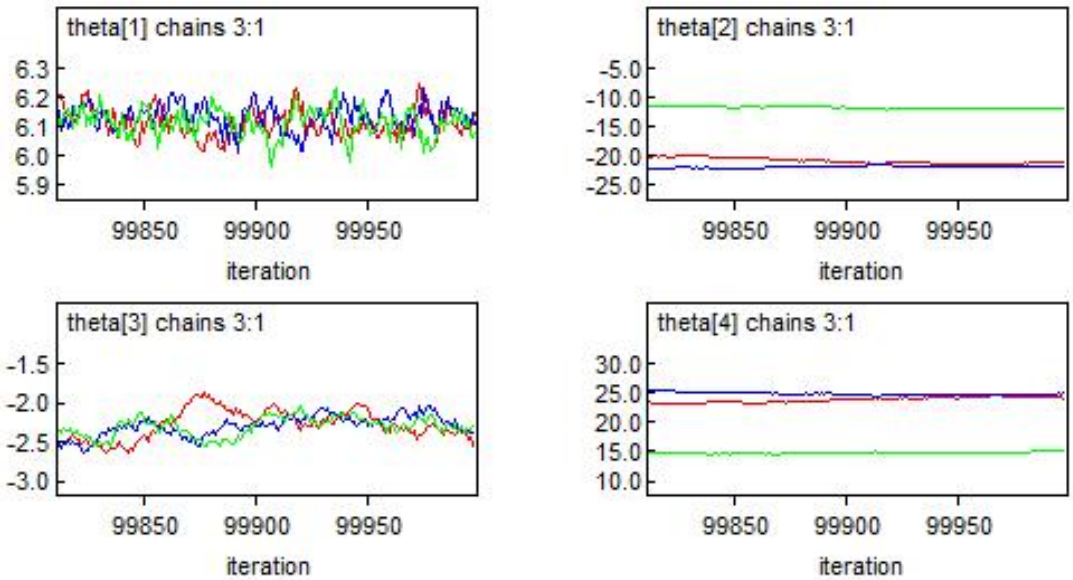

Fig. B.7 Trace plots for the parameters $\boldsymbol{\theta}$, if $y_2 = 0$

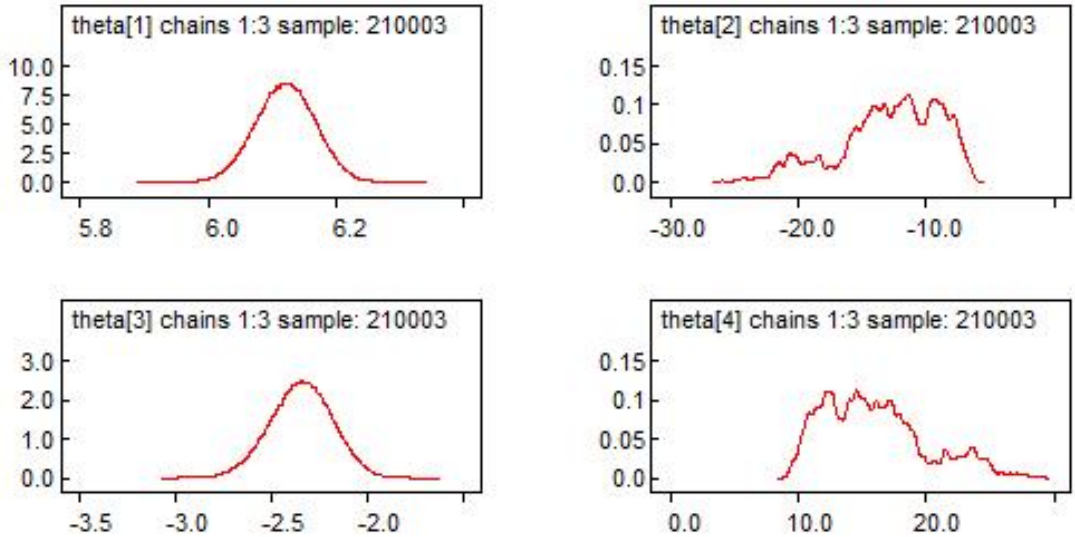

Fig. B.8 Marginal density plots for the parameters $\boldsymbol{\theta}$, if $y_2 = 0$

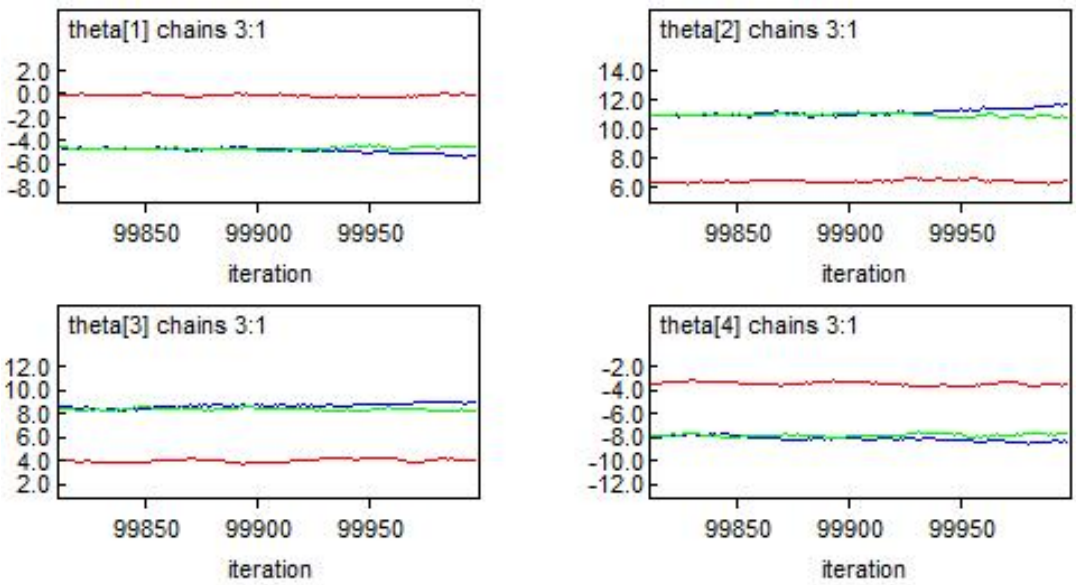

Fig. B.9 Trace plots for the parameters $\boldsymbol{\theta}$, if $y_1 = 0$

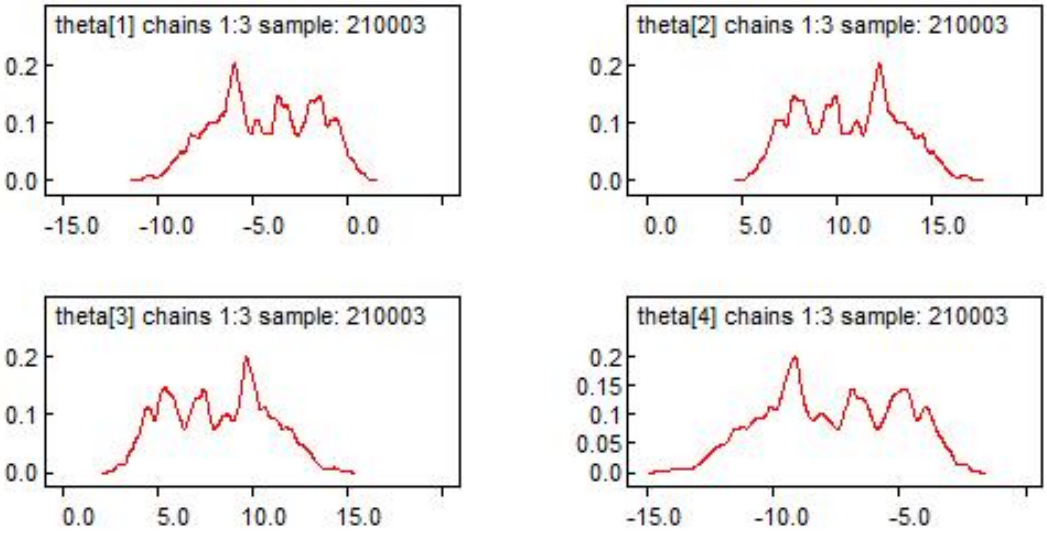

Fig. B.10 Marginal density plots for the parameters $\boldsymbol{\theta}$, if $y_1 = 0$

Chapter 5

- The 13 $\boldsymbol{\alpha}$ vectors with 22 parameters for the parameter redundant model in Section 5.2.

```
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,-1,0,0,0,0,0,0,0,1,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0)
```

- The 36 $\boldsymbol{\alpha}$ vectors with 131 parameters for the parameter redundant model in Section 5.4. Each column represents a vector.

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0-1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0-1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
```

• The 3 $\boldsymbol{\alpha}$ vectors with 62 parameters for the parameter redundant model in Section 5.4.4.

```
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,
 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,
 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,
 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
```

• For the example presented in Section 5.4, observations (y_i) of 243 possible cross-classification of the five variables are as follows:

rs9306859_A	rs11128775_G	rs6803988_T	rs4975616_G	rs7748167_C	y_i
0	0	0	0	0	436
				1	69
				2	5
			1	0	511
				1	85
				2	7
			2	0	133
				1	25
				2	1
		1	0	0	300
				1	51
				2	5
			1	0	374
				1	64
				2	2
			2	0	115
				1	17
				2	0
		2	0	0	74
				1	10

			2	1
		1	0	64
			1	13
			2	0
		2	0	20
			1	3
			2	1
1	0	0	0	87
			1	24
			2	2
		1	0	146
			1	15
			2	2
		2	0	25
			1	9
			2	1
	1	0	0	84
			1	14
			2	0
		1	0	94
			1	22
			2	0
		2	0	25
			1	3
			2	0
	2	0	0	14
			1	5
			2	1
		1	0	17
			1	3
			2	1
		2	0	6
			1	0
			2	0
2	0	0	0	4
			1	2
			2	0
		1	0	7
			1	2
			2	0
		2	0	0
			1	0
			2	0
	1	0	0	4
			1	2
			2	0
		1	0	5
			1	0
			2	0
		2	0	4
			1	0
			2	0
	2	0	0	0
			1	0
			2	0
		1	0	0

				1	0
				2	0
			2	0	0
				1	0
				2	0
1	0	0	0	0	117
				1	22
				2	0
			1	0	124
				1	19
				2	1
			2	0	28
				1	11
				2	0
		1	0	0	92
				1	16
				2	1
			1	0	79
				1	16
				2	2
			2	0	40
				1	4
				2	0
		2	0	0	10
				1	1
				2	0
			1	0	17
				1	4
				2	0
			2	0	7
				1	1
				2	0
	1	0	0	0	27
				1	8
				2	0
			1	0	31
				1	12
				2	0
			2	0	5
				1	2
				2	1
		1	0	0	26
				1	8
				2	1
			1	0	20
				1	2
				2	0
			2	0	7
				1	0
				2	0
		2	0	0	5
				1	0
				2	0
			1	0	2
				1	3
				2	1

			2	0	3
				1	0
				2	0
	2	0	0	0	2
				1	0
				2	0
			1	0	0
				1	0
				2	0
			2	0	0
				1	0
				2	0
		1	0	0	2
				1	0
				2	0
			1	0	0
				1	0
				2	0
			2	0	0
				1	0
				2	0
		2	0	0	0
				1	0
				2	0
			1	0	2
				1	0
				2	0
			2	0	0
				1	0
				2	0
2	0	0	0	0	7
				1	0
				2	0
			1	0	11
				1	3
				2	0
			2	0	1
				1	2
				2	0
		1	0	0	2
				1	0
				2	0
			1	0	5
				1	0
				2	0
			2	0	1
				1	2
				2	0
		2	0	0	0
				1	0
				2	0
			1	0	1
				1	0
				2	0
			2	0	1
				1	0

			2	0
1	0	0	0	1
			1	0
			2	0
		1	0	3
			1	0
			2	0
		2	0	0
			1	0
			2	0
	1	0	0	0
			1	0
			2	0
		1	0	3
			1	0
			2	0
		2	0	0
			1	0
			2	0
	2	0	0	0
			1	0
			2	0
		1	0	0
			1	0
			2	0
		2	0	0
			1	0
			2	0
2	0	0	0	0
			1	0
			2	0
		1	0	0
			1	0
			2	0
		2	0	0
			1	0
			2	0
	1	0	0	0
			1	0
			2	0
		1	0	0
			1	0
			2	0
		2	0	0
			1	0
			2	0
	2	0	0	0
			1	0
			2	0
		1	0	0
			1	0
			2	0
		2	0	0
			1	0
			2	0