UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Loglinear Model Selection and Inference for Contingency Tables

Permalink

https://escholarship.org/uc/item/8tx3d4w4

Author

Chowdhury, Arnab

Publication Date

2018

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Loglinear Model Selection and Inference for Contingency Tables

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Arnab Chowdhury

September 2018

Dissertation Committee:

Professor Subir Ghosh, Chairperson
Professor Barry C. Arnold
Professor Gregory J. Palardy

Copyright by
Arnab Chowdhury
2018

The Dissertation of Arnab Chowdhury is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to extend my sincere thanks to my advisor Professor Subir Ghosh who trusted in me, stayed by my side, and helped me to develop statistical intuition and to become a statistician. He guided me through the whole journey, and influenced the way I think about statistics. Not only has he been a great mentor in my research, but he has always been there to support and help in every difficulties during these years. I also want to thank Professor Barry C. Arnold and Professor Gregory J. Palardy for their time and interest in my research. I want to express my deepest thanks to the other professors in UCR and my colleagues. A special word of thanks goes out to all my friends in UCR. You guys have been pillars of strength and perseverance through out my stay in Riverside.

My parents, Mr. Baladeb Chowdhury and Mrs. Swati Chowdhury, deserve special mention for their endless support and prayers. I would like to thank Ahindra for being supportive and caring sibling. Every bit of success that I have been able to achieve, I owe it to my lovely family. I would also like to thank my friends Vaibhav, Shaocheng, Supawadee, Ying, Sakar, Mi, Matthew and Luke who encouraged me a lot during this journey of my PhD.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Loglinear Model Selection and Inference for Contingency Tables

by

Arnab Chowdhury

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, September 2018
Professor Subir Ghosh, Chairperson

We propose a class of multiplicative models to describe the dependence of the response count on the effects of the explanatory variables and their interactions in a contingency table. The proposed models are motivated by loglinear models for contingency tables. Under these models, every cell count and/ or probability is the product of effects of the categorical variables used as explanatory variables. Such models are useful in analyzing complete as well as incomplete tables. We present a connection between these models and graphical probability models in describing conditional independence structures among the explanatory variables.

We extend the standard unsaturated loglinear models to a complete model retaining the conditional independence structure. We characterize the extension of the unsaturated model to the standard saturated model. We also compare the estimates of the unknown parameters in these two saturated representations. Necessary and sufficient conditions for the equivalence of the estimates of unknown parameters in these two models are given. We propose a criterion function for model selection based on the extensions of the unsaturated models. We discuss uniqueness and sum to zero proper-

ties of these extensions. We illustrate model building, estimation and selection methods with a auto-accident data set (Agresti, 2013). We also study the performance comparison of the newly proposed method with the standard methods that are commonly used in practice.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main objective of this dissertation is to build a new class of multiplicative models to describe the dependence of the response count on the effects of the categorical variables in a contingency table. Under these models, the cell counts and/ or probabilities are the products of the effects of the categorical variables. These models can be used to analyze complete as well as incomplete contingency tables. The proposed models are the generalizations of the standard loglinear models and are motivated by Relational models (Klimova et al., 2011).

1.1 Thesis Description

In Chapter 2, we introduce contingency tables and some standard models to describe the data in the contingency tables. We also introduce the standard loglinear models and the extension of these models which will be used for describing the effects of one or more categorical variables on the response count. In this chapter we discuss different unsaturated models and characterize the interaction parameters in different conditional

independence models. We also introduce undirected graphs to define conditional independence and establish the connection with interaction parameters. In Chapter 3, we discuss Maximum Likelihood Estimation and Minimum Discrimination Information Estimation as the two estimation methods of unknown parameters in loglinear models. We also check the goodness-of-fit of these unsaturated models in this chapter. In Chapter 4, we characterize the extension of the unsaturated model to the standard saturated model. We also investigate the relation between the estimates of the unknown parameters between these two saturated models in this chapter. In Chapter 5, we discuss different choices of the extension and some important properties of this extension. We also introduce our criterion function for selecting the best fitted model based on D in this chapter. The results and comparison of different criterion functions from the simulation study are presented in Chapter 6. Chapter 7 presents the conclusion of this dissertation.

1.2 Thesis Contribution

In this dissertation, we discuss a method of loglinear model building by adding one or more new terms orthogonally to the terms of the existing model. This way we can extend any lower order model to a higher order model. When we add all the new non-existing terms to the existing model this way, we obtain a new saturated model (Saturated 2). We compare this new saturated model with the standard saturated loglinear model (Saturated 1). The orthogonal extension provide some conditional independence structures among the categorical variables in the unsaturated lower order models. This

way of extending the loglinear models allows us to relate the parameters that are present in the model to the parameters that are assumed to be zero in an orthogonal way. We compare the Saturated 1 model with the Saturated 2 model. We establish the necessary and sufficient conditions for the equivalence of the parameter estimates between these two models. We demonstrate the uniqueness and sum to zero properties of this orthogonal extension. We also propose a new criterion function for selecting the best fitted unsaturated model based on the extensions of the unsaturated models.

Chapter 2

Contingency Tables and Models

2.1 Summary

In this chapter, we introduce contingency tables and some standard models to describe the data in the contingency tables. We define the basic notations and terminologies in a contingency table (section 2.2). We present two contingency tables with no missing observations and one contingency table with a missing observation (section 2.3). We describe the Poisson Probabililty Models (PPM), Multinomial Probability Models (MPM), the Loglinear Models (LLM) and the Multiplicative Models (MM) (section 2.4) which will be used in fitting the data (Chapter 5) in a contingency table. These models are generally used for describing the effects of one or more categorical variables on the response. We discuss different unsaturated LLMs (section 2.5). We characterize the interaction parameters in conditional independence LLMs (section 2.6). We use undirected graphs to define conditional independence between two categorical variables and establish the connection with interaction parameters (section 2.7).

2.2 Contingency Tables

In many observational studies individuals are classified to different levels of one or more categorical variables. The categorical variables are often called the explanatory variables or the factors. The response is the number of observations or the counts belonging to the different level combinations of the factors. An m-way contingency table is formed if the classification is done on m categorical variables $X_1, X_2, .., X_m$. The cells in a multidimensional contingency table represent the level combinations of the categorical variables under consideration. The individuals with the same characteristics are identified and the count is placed in a cell of a contingency table. In a contingency table we are normally interested in the relations between one classification with respect to a factor and one or more other classifications with respect to the other factors. In general, we wish to determine the effects of the factors on the data by fitting a suitable model.

A two-way contingency table is the simplest form of classification where an individual is classified by two categorical variables. Let $Y_{i_1 i_2}$ be the random variable representing the cell counts for i_1-th level of X_1 and i_2-th level of X_2 ($i_1 = 1, 2, .., I_1; i_2 = 1, 2, .., I_2$). We define $p_{i_1 i_2}$ as the cell probability of an individual belonging to the i_1-th category of X_1 and i_2-th category of X_2, satisfying $0 < p_{i_1 i_2} < 1, \forall\ i_1, i_2$ and $\sum_{i_1, i_2} p_{i_1 i_2} = 1$. If we consider 0 and 1 be the two levels of a factor then for a 2×2 contingency table, all possible level combinations of (X_1, X_2) are $S = \{(i_1, i_2); i_1, i_2 = 0, 1\}$. We also define $N = I_1 I_2$ as the total number of cells in a $I_1 \times I_2$ contingency table. The subscript i is used to denote different level combinations of the explanatory variables. We define $n_i = n_{i_1 i_2} = y_{i_1 i_2} = y_i$ = number of observations belonging to the i_1-th category of X_1 and i_2-th category of X_2, where $i = 1, 2, .., N$ and $\sum_{i_1, i_2} n_{i_1 i_2} = \sum_i n_i = n$ = total

counts = total number of observations in the study. The rows represent the categories of a variable X_1 and the columns represent the categories of another variable X_2. A blank 2×2 contingency table is presented in Table 2.1.

Table 2.1: A blank 2×2 contingency table with cell counts, Row and Column totals

		X_2		Row Total
		(1)	(0)	
X_1	(1)	n_{10}	n_{10}	n_{1+}
	(0)	n_{01}	n_{00}	n_{0+}
Column Total		n_{+1}	n_{+0}	n

These notations can be generalized to an m-way contingency table. In an m-way contingency table, we have $N = I_1 \times I_2 \times \ldots I_m$ cells and $n_{i_1 i_2 \ldots i_m}$, $p_{i_1 i_2 \ldots i_m}$ and $\mu_{i_1 i_2 \ldots i_m}$ represent the counts, probabilities and means of the cell $(i_1, i_2, \ldots, i_m)$, respectively. For simplicity, we consider m categorical variables each with 2 levels 0 and 1. The set $S = \{(i_1, i_2, \ldots, i_m);\ i_1, \ldots i_m = 0, 1\}$ is the collection of all possible level combinations of $(X_1, X_2, \ldots, X_m)$. We also define $\underline{\lambda} = (\lambda, \lambda_1, \lambda_2, \lambda_3, \lambda_{12}, ..., \lambda_{123}, ..., \lambda_{12\ldots m})'$, the unknown parameters, where $\lambda_{j_1 \ldots j_m}$ is the joint effect of $X_1 \ldots X_m$. The cells represent the presence or absence of the unknown parameters in the response Y. The elements x_{ij} in the design matrix $\mathbf{X}$ (except the first column) are 1 if the λ_j is present in the expression of $\ln \mu_i$ and 0, otherwise. We define a design matrix X as

$$\mathbf{X} = \left[\begin{array}{c|ccc|cccc} 1 & x_{11} & x_{12} & .. & x_{11}x_{12} & .. & x_{11}x_{12}x_{13} & .. & x_{11}x_{12}..x_{1m} \\ 1 & x_{21} & x_{22} & .. & x_{21}x_{22} & .. & x_{21}x_{22}x_{23} & .. & x_{21}x_{22}..x_{2m} \\ 1 & x_{31} & x_{32} & .. & x_{31}x_{32} & .. & x_{31}x_{32}x_{33} & .. & x_{31}x_{32}..x_{3m} \\ .. & .. & .. & .. & .. & .. & .. & .. & .. \\ 1 & x_{N1} & x_{N2} & .. & x_{N1}x_{N2} & .. & x_{N1}x_{N2}x_{N3} & .. & x_{N1}x_{N2}..x_{Nm} \end{array}\right].$$

2.3 Examples

The following are some examples of contingency tables. Example 2.1 and 2.2 represent complete contingency tables and Example 2.3 represents an incomplete contingency table.

Example 2.1: Does seat-belt use in automobiles reduce fatal injuries? To answer this question a study was performed to observe the injury outcomes of 68694 passengers in autos and light trucks involved in accidents for one year in the state of Maine. Table 2.2 presents the injury outcomes of 68694(=n) passengers (Agresti, 2013). In this example, we have 3 factors: Location (X_1), Seat-belt use (X_2) and Injury (X_3), each at 2 levels. For Location: Urban (0), Rural (1); Seat-belt use: No (0), Yes (1) and Injury: Yes (1), No (0). The total number of all possible cells/ level-combinations are $2\times2\times2 = 8$. The counts are placed in each of the 8 cells.

Example 2.2: In this example we consider three categorical variables X_1 = alcohol use, X_2 = cigarette use and X_3 = marijuana use (Agresti, 2013), each having two categories 0 and 1. We denote the 8 category combinations by (i_1, i_2, i_3), where $i_u = 0, 1$; $u = 1, 2, 3$

Table 2.2: The response counts for different locations (X_1), seat-belt use (X_2) and injuries (X_3)

X_1	X_2	X_3	
		(0)	(1)
(0)	(0)	17668 (n_{000})	1808 (n_{001})
	(1)	22556 (n_{010})	1139 (n_{011})
(1)	(0)	9369 (n_{100})	2057 (n_{101})
	(1)	12827 (n_{110})	1270 (n_{111})

and $\sum n_{i_1 i_2 i_3} = n$. The data are presented in Table 2.3.

The above two are the examples of contingency tables with no missing cell counts. The following (Example 2.3) is an example of the contingency tables with missing observation.

Example 2.3: A sample of 156 dairy calves born in Okeechobee County, Florida, were classified according to whether they caught pneumonia within 60 days of birth (Agresti, 2013). Calves that got a pneumonia infection were also classified according to whether they got a secondary infection within 2 weeks after the first infection cleared up. Since the calves cannot have secondary infection without having the primary infection, a

Table 2.3: The response counts of High School Seniors for Alcohol, Cigarette and Marijuana Use

X_1	Cigarette Use (X_2)	Marijuana Use (X_3)	
		Yes (1)	No (0)
Yes (1)	Yes (1)	911 (n_{111})	538 (n_{110})
	No (0)	44 (n_{101})	456 (n_{100})
No (0)	Yes (1)	3 (n_{011})	43 (n_{010})
	No (0)	2 (n_{001})	279 (n_{000})

structural zero cell (a cell is known a priori to have a zero value) occurs in a 2×2 table corresponding to the cell (no primary infection, secondary infection). Table 2.4 shows the data.

2.4 Probability models for Contingency table

In this section we define different probability models for a contingency table. The complete data provides only the counts at all level combinations of factors. These counts immediately provide the estimates of probabilities or means at all level combinations of factors but do not immediately provide the dependence of means or proba-

Table 2.4: Pneumonia infection in calves (Agresti, 2013)

		Secondary infection		Row Total
		Yes (1)	No (2)	
Primary infection	Yes (1)	30	63	93
	No (2)	–	63	63
Column Total		30	126	156

bilities on the individual factors and their interactions. To describe the dependence we consider the response variable as the count in each cell and the categorical variables used to define the cells or level combinations are treated as explanatory variables. For simplicity, we consider a contingency table with $N = 2^m$ cells.

2.4.1 Relational Models (RM)

For the Relational Models (Klimova, Rudas, & Dobra, 2011), the cell means or the cell probabilities are expressed as the product of some functions of the unknown parameters involved in the model. In RMs, the cell means $\underline{\mu}$ of the true distribution are expressed as

$$\ln \underline{\mu} = X\underline{\lambda},$$

that is,

$$\ln \mu_i = \lambda + \sum_{j=1}^{m} x_{ij}\lambda_j + \sum_{j_1<j_2=1}^{m} x_{ij_1}x_{ij_2}\lambda_{j_1j_2} + ... + x_{i1}x_{i2}...x_{im}\lambda_{12..m}. \tag{2.1}$$

From (2.1), μ_i can be expressed as

$$\mu_i = \exp[\lambda + \sum_j x_{ij}\lambda_j + \dots] = \prod_j \exp[\lambda + x_{i1}\lambda_1 + \dots], \tag{2.2}$$

the product of the unknown parameters ($\underline{\lambda}$) in the model, where the elements of $\underline{\mu}$ correspond to the elements of S, the set of all possible cells. The μ_i is the element of $\underline{\mu}$ corresponding to the i-th element of S. The subscript i is used to denote different level combinations of the explanatory variables. We discuss the commonly used loglinear models for count data (section 2.4.2) which is a special case of RMs.

2.4.1.1 Dual and Non-dual representation of a Relational model

A relational model can be represented in either a dual (Klimova, Rudas, & Dobra, 2011) or a non-dual way.

Definition 2.1: For a non-null matrix $\mathbf{D}$ with rank($\mathbf{D}$) = N − rank($\mathbf{X}$) and $\mathbf{D}'\mathbf{X} = \mathbf{0}$, a dual representation of a relational model is defined as

$$\mathbf{D}' \ln \underline{\mu} = \underline{0}. \tag{2.3}$$

Definition 2.2: A non-dual representation of a relational model is defined as

$$\mathbf{D}' \ln \underline{\mu} = \mathbf{X}^* \underline{\lambda} \neq \underline{0}, \tag{2.4}$$

where $\mathbf{X}^* = \mathbf{D}'\mathbf{X} \neq \mathbf{0}$.

If $\mathbf{D} = \mathbf{0}$, then $\mathbf{D}'\mathbf{X} = \mathbf{0}\mathbf{X} = \mathbf{0}$, for any design matrix $\mathbf{X}$. We consider the situation where $\mathbf{D} \neq 0$. For saturated models, there always exists a matrix $\mathbf{D}$, such that $\mathbf{D}'\mathbf{X} \neq \mathbf{0}$, i.e., if $\mathbf{D} \neq \mathbf{0}$, saturated models can also be expressed in non-dual form. We consider the unsaturated models where $1 \leq rank(\mathbf{D}) < N$. Unsaturated models can be expressed

in both dual and non-dual form when $\mathbf{D} \neq \mathbf{0}$. For any unsaturated model, we can find matrices $\mathbf{D}_i \neq \mathbf{0}, i = 1, 2$ with $rank(\mathbf{D}_i) = N - rank(\mathbf{X})$ such that $\mathbf{D}_1^{'} \ln \underline{\mu} = \underline{0}$, where $\mathbf{D}_1^{'}\mathbf{X} = 0$ and $\mathbf{D}_2^{'} \ln \underline{p} = \mathbf{D}_2^{'}\mathbf{X}\underline{\lambda} = \mathbf{X}^{*}\underline{\lambda} \neq \underline{0}$, where $\mathbf{D}_2^{'}\mathbf{X} = \mathbf{X}^{*} \neq \mathbf{0}$.

We consider the relational model for independence between X_1 and X_2 in a 2×2 contingency table. The relational model for cell means ($\underline{\mu}$) is defined as

$$\ln \underline{\mu} = \mathbf{X}\underline{\lambda}, \tag{2.5}$$

where $\underline{\mu} = (\mu_{00}, \mu_{01}, \mu_{10}, \mu_{11})'$, $\underline{\lambda} = (\lambda, \lambda_1, \lambda_2)'$and $\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, $rank(\mathbf{X}) = 3$.

If we take $\mathbf{D}_1 = (1, -1, -1, 1)^{'}$, we get

$$\mathbf{D}_1^{'}\mathbf{X} = \underline{0},$$

and hence the dual representation of the relational model (2.5) is

$$\mathbf{D}_1^{'} \ln \underline{\mu} = \underline{0},$$

i.e.,

$$\ln \frac{\mu_{00}\mu_{11}}{\mu_{01}\mu_{10}} = 0,$$

i.e., X_1 is independent of X_2. If we take $\mathbf{D}_2 = (1, 1, 1, 1)^{'}$, we get

$$\mathbf{D}_2^{'}X = (4, 2, 2) \neq \underline{0},$$

and hence a non-dual representation of the relational model (2.5) is

$$\mathbf{D}_2^{'} \ln \underline{\mu} = 4\lambda + 2\lambda_1^{X_1} + 2\lambda_1^{X_2} \neq 0.$$

2.4.2 Loglinear Models (LLM)

We define the expected count for a cell by μ_i, i.e., $E[Y_i] = \mu_i = np_i$. In all the models introduced in this section, the effects of the explanatory variables on the response variables Y_i are modeled through the parameters μ_i. We propose a general class of models for a contingency table where we express the logarithm of the $E[Y]$ as a linear function of the unknown parameters $\underline{\lambda}$. In an LLM, we describe the dependencies of the responses on the explanatory variables as

$$\ln E[Y_i] = \lambda + \sum_{j=1}^{m} x_{ij}\lambda_j + \sum_{j_1<j_2=1}^{m} x_{ij_1}x_{ij_2}\lambda_{j_1j_2} + ... + x_{i1}x_{i2}...x_{im}\lambda_{12..m}, \tag{2.6}$$

where $\lambda_0 = \lambda$ and $x_{i0} = 1 \; \forall \; i = 1, 2, \ldots, N$. We define the following

$$\begin{aligned}
(\underline{x}_i^{(1)})' \underline{\lambda}^{(1)} &= \sum_{j=1}^{m} x_{ij}\lambda_j, \\
(\underline{x}_i^{(2)})' \underline{\lambda}^{(2)} &= \sum_{j_1<j_2=1}^{m} x_{ij_1}x_{ij_2}\lambda_{j1j2}, \\
&\ldots, \\
(\underline{x}_i^{(u)})' \underline{\lambda}^{(u)} &= \sum_{j_1<j_2\cdots<j_u} x_{ij_1}x_{ij_2}\ldots x_{ij_u}\lambda_{j_1j_2\ldots j_u}, \\
&\ldots,
\end{aligned}$$

in lexicographic order. The LLMs are special cases of MMs since we can express E$[Y_i]$ as

$$E[Y_i] = \exp[\sum_{u=1}^{m} (\underline{x}_i^{(u)})' \underline{\lambda}^{(u)}] = \prod_{u=1}^{m} \exp[(\underline{x}_i^{(u)})' \underline{\lambda}^{(u)}]. \tag{2.7}$$

If Y_i's are assumed to be independent Poisson distributions with parameters μ_i,

$$f(y_i, \mu_i) = \frac{\mu_i e^{-\mu_i}}{y_i!}, y_i = 0, 1, ..; \mu_i > 0, \tag{2.8}$$

where μ_i is the expected count for the i-th cell, the model is called Poisson Loglinear Model (PLLM). For Poisson model $E[Y_i] = \mu_i$ and $Var[Y_i] = \mu_i$ (Bishop et al., 2007). In PLLM, the dependencies of the response on explanatory variables can be expressed as

$$\ln \mu_i = \sum_j x_{ij}\lambda_j. \tag{2.9}$$

In addition to the Poisson model in (2.9), if the total sample size $\sum_i Y_i$ is fixed by the design of the study then the probability distribution of Y_i's, conditional on the sum $\sum_i Y_i = n$, is the multinomial distribution with probabilitiy density (Bishop et al., 2007)

$$f(\underline{y}|n) = \frac{n!}{\prod_i y_i!} \prod_i p_i^{y_i}, \tag{2.10}$$

where $p_i = \frac{\mu_i}{\sum_i \mu_i}$ and $\sum_i p_i = 1$. Under Multinomial Loglinear Model (MLLM), the expected value of Y_i is

$$E[Y_i] = np_i.$$

In MLLM, we can model p_i since the total sample size is fixed. For fixed sample size n, the effects of the explanatory variables on the responses Y_i is modeled as

$$\ln p_i = \sum_{j=0}^{m} x_{ij}\lambda_j, \tag{2.11}$$

where $0 < p_i < 1 \; \forall \; i$ and the condition $\sum_i p_i = 1$ implies

$$\exp[\lambda_0] = \frac{1}{\sum_{i=1}^{N} \exp[\sum_{j=1}^{m} x_{ij}\lambda_j]}. \tag{2.12}$$

The vector of unknown parameters $\underline{\lambda}$ is consists of main effects and interaction terms. The LLM in (2.6) is called a saturated model if all the $N(= 2^m)$ parameters, assuming each factor has 2 levels, are present in the model. Unsaturated models are the

models where one or more components of $\underline{\lambda}$ are zero. In practice, unsaturated models are preferred since they have less number of parameters to interpret the association. We consider the class of models where some interaction parameters are not present (Section 2.5). We define $\underline{\lambda} = (\lambda, \lambda_1, \lambda_2, \lambda_3, \lambda_{12}, \ldots, \lambda_{123}, \ldots, \lambda_{12\ldots m})' = (\lambda, \underline{\lambda}_1', \underline{\lambda}_2')'$. The vector $\underline{\lambda}_1'$ contains all the main effect parameters and the vector $\underline{\lambda}_2'$ contains all the interaction parameters. The design matrix X can be written as

$$\mathbf{X} = \left[\begin{array}{c|ccc|ccccc} 1 & x_{11} & x_{12} & .. & x_{11}x_{12} & .. & x_{11}x_{12}x_{13} & .. & x_{11}x_{12}..x_{1m} \\ 1 & x_{21} & x_{22} & .. & x_{21}x_{22} & .. & x_{21}x_{22}x_{23} & .. & x_{21}x_{22}..x_{2m} \\ 1 & x_{31} & x_{32} & .. & x_{31}x_{32} & .. & x_{31}x_{32}x_{33} & .. & x_{31}x_{32}..x_{3m} \\ .. & .. & .. & .. & .. & .. & .. & .. & .. \\ 1 & x_{N1} & x_{N2} & .. & x_{N1}x_{N2} & .. & x_{N1}x_{N2}x_{N3} & .. & x_{N1}x_{N2}..x_{Nm} \end{array}\right]$$

$$= [\underline{j}|\mathbf{X}_1^*|\mathbf{X}_2]$$

$$= [\mathbf{X}_1|\mathbf{X}_2].$$

2.5 Different Unsaturated Loglinear Models

We are interested in the class of unsaturated models where two or higher order interactions are not present. We define some classes of unsaturated loglinear models for a contingency table with m categorical variables each at 2 levels in the following way:

C1 : 1 component of $\underline{\lambda}_2$ is zero, number of models in C1 = $\binom{N-m-1}{1}$;

C2 : 2 components of $\underline{\lambda}_2$ are zero, number of models in C2 = $\binom{N-m-1}{2}$;

...,

C($N-m-1$) : all components of $\underline{\lambda}_2$ are zero, number of models in C($N-m-1$) = 1.

We also define $\underline{\lambda}_2$ as $\underline{\lambda}_2 = \begin{pmatrix} \underline{\lambda}_2^{(1)} \\ \\ \underline{\lambda}_2^{(0)} \end{pmatrix}$, where $\underline{\lambda}_2^{(1)}$ = non-zero components of $\underline{\lambda}_2$ and $\underline{\lambda}_2^{(0)}$ consists of the components of $\underline{\lambda}_2$ that are zero and $X_1^{(1)} = [\underline{j}|X_1|X_2^{(1)}]$ as the design matrix for an unsaturated model, where $X_2 = [X_2^{(1)}|X_2^{(0)}]$. We want to find the best fitted unsaturated model among all these models. We choose the best fitted unsaturated model using the criterion functions (Chapter 3).

For example, we consider a 2×2×2 contingency table and all the unsaturated models associated with it. Let $Y_{i_1 i_2 i_3}$ be the random variable representing the cell counts for i_1-th level of X_1, i_2-th level of X_2 and i_3-th level of X_3 ($i_1 = 0, 1; i_2 = 0, 1; i_3 = 0, 1$), $p_{i_1 i_2 i_3}$ be the cell probability of an individual belonging to i_1-th category of X_1, i_2-th category of X_2 and i_3-th category of X_3, satisfying $0 < p_{i_1 i_2 i_3} < 1, \forall i_1, i_2, i_3$ and $\sum_{i_1,i_2,i_3} p_{i_1 i_2 i_3} = 1$.We have $N = 2 * 2 * 2 = 8$ as the total number of cells. If we consider 0 and 1 be the two levels of a factor, then for Table 2.2, all possible level combinations of (X_1, X_2, X_3) are

$S = \{(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)\}$

$= \{(i_1, i_2, i_3); i_1, i_2, i_3 = 0, 1\}$

$= \{(i); i = 1, 2, .., 8\}$.

We have, from Table 2.2, $n_i = n_{i_1 i_2 i_3} = y_{i_1 i_2 i_3} = y_i$ = number of observations belonging to the i_1-th category of X_1, i_2-th category of X_2 and i_3-th category of X_3, where $i = 1, 2, .., N(= 8)$ and $\sum_{i_1,i_2,i_3} n_{i_1 i_2, i_3} = \sum_i n_i = n$, $\underline{\lambda} = (\lambda, \lambda_1, \lambda_2, \lambda_3, \lambda_{12}, \lambda_{13}, \lambda_{23}, \lambda_{123})' = (\lambda, \underline{\lambda}_1', \underline{\lambda}_2')'$ are the unknown parameters, where $\underline{\lambda}_1' = (\lambda_1, \lambda_2, \lambda_3)'$, $\underline{\lambda}_2' = (\lambda_{12}, \lambda_{13}, \lambda_{23}, \lambda_{123})'$, and the design matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \left[\begin{array}{c|ccc|ccccc} 1 & x_{11} & x_{12} & x_{13} & x_{11}x_{12} & x_{11}x_{13} & x_{12}x_{13} & x_{11}x_{12}x_{13} \\ 1 & x_{21} & x_{22} & x_{23} & x_{21}x_{22} & x_{21}x_{23} & x_{22}x_{23} & x_{21}x_{22}x_{23} \\ 1 & x_{31} & x_{32} & x_{33} & x_{31}x_{32} & x_{31}x_{33} & x_{32}x_{33} & x_{31}x_{32}x_{33} \\ 1 & x_{41} & x_{42} & x_{43} & x_{41}x_{42} & x_{41}x_{43} & x_{42}x_{43} & x_{41}x_{42}x_{43} \\ 1 & x_{51} & x_{52} & x_{53} & x_{51}x_{52} & x_{51}x_{53} & x_{52}x_{53} & x_{51}x_{52}x_{53} \\ 1 & x_{61} & x_{62} & x_{63} & x_{61}x_{62} & x_{61}x_{63} & x_{62}x_{63} & x_{61}x_{62}x_{63} \\ 1 & x_{71} & x_{72} & x_{73} & x_{71}x_{72} & x_{71}x_{73} & x_{72}x_{73} & x_{71}x_{72}x_{73} \\ 1 & x_{81} & x_{82} & x_{83} & x_{81}x_{82} & x_{81}x_{83} & x_{82}x_{83} & x_{81}x_{82}x_{83} \end{array}\right]$$

$$= \left[\begin{array}{c|ccc|cccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$$

We are interested in the class of unsaturated models where two or higher order interactions are not present. The classes of unsaturated loglinear models are

C1 : 1 component of $\underline{\lambda}_2$ is zero, number of models in C1 = $\binom{8-3-1}{1}$= 4;

C2 : 2 components of $\underline{\lambda}_2$ are zero, number of models in C2 = $\binom{8-3-1}{2}$= 6;

C3 : 3 components of $\underline{\lambda}_2$ are zero, number of models in C3 = $\binom{8-3-1}{3}$= 4;

C4 : all 4 components of $\underline{\lambda}_2$ are zero, number of models in C4 = $\binom{8-3-1}{4}$= 1.

For the data in Table 2.2, we consider the k-th unsaturaed loglinear model from one of the classes C1, C2, C3 or C4.

$$\ln \underline{\mu} = \lambda . \underline{j} + \mathbf{X}_1 \underline{\lambda}_1 + \mathbf{X}_2^{(k)} \underline{\lambda}_2^{(k)} = \mathbf{X}_1^{(k)} \begin{pmatrix} \lambda \\ \underline{\lambda}_1 \\ \underline{\lambda}_2^{(i)} \end{pmatrix}. \tag{2.13}$$

For example, we consider the unsaturated model from the class C2 with $\lambda_{12}, \lambda_{123} = 0$, i.e., $\underline{\lambda}_1 = (\lambda_1, \lambda_2, \lambda_3)', \underline{\lambda}_2^{(k)} = (\lambda_{13}, \lambda_{23})'$ and

$$\mathbf{X}_1^{(k)} = \left[\begin{array}{c|ccc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]. \tag{2.14}$$

We present some unsaturated models in Tables 2.5 - 2.9 in a $2 \times 2 \times 2$ contingency table with the corresponding design matrices.

2.6 Characterization of the interaction parameters in LLMs

LLM describing the data well is used to determine the dependence of means or probabilities on the individual factors and their interactions. LLMs describing the data well provide the quantitative measures of direct association as well as conditional association between factors. We present two definitions of conditional independence of X_1 and X_2 given X_3 and establish the equivalence between them (Theorem 2.1). We present the conditional association between factors using the unknown λ parameters (Section 2.6.2) or the cell probabilities in both complete and incomplete contingency tables.

2.6.1 Conditional Independence

The main objective is to establish the conditions on the unknown parameters when there exists a conditional independence between two variables. First, we define the conditional independence in two ways and show their equivalence in Theorem 2.1. Theorem 2.2 demonstrates the constraints on λ parameters when X_1 is conditionally independent of X_2 given X_3. We consider three categorical variables X_1, X_2 and X_3, each having two levels 0 and 1.

Definition 2.1. For a fixed category i_3 of X_3, the ratio

$$\theta_{I_1 I_2(i_3)} = \frac{p_{i_1 i_2 i_3} p_{(1-i_1)(1-i_2)i_3}}{p_{i_1(1-i_2)i_3} p_{(1-i_1)i_2 i_3}},\ i_3 = 0, 1, \tag{2.15}$$

is called the conditional odds ratio.

Definition 2.2. X_1 and X_2 are conditionally independent given X_3 if they are condition-

ally independent at every level of X_3 (Agresti, 2013), i.e.,

$$\theta_{I_1 I_2(i_3)} = 1, \ \forall \ i_1, i_2, i_3 = 0, 1. \tag{2.16}$$

If X_1 and X_2 are conditionally independent given X_3, we have

$$\theta_{I_1 I_2(i_3)} = 1 \Longleftrightarrow p_{i_1 i_2 i_3} = \frac{p_{i_1(1-i_2)i_3} p_{(1-i_1)i_2 i_3}}{p_{(1-i_1)(1-i_2)i_3}}, \ \forall \ i_1, i_2, i_3 = 0, 1.$$

Definition 2.3. X_1 and X_2 are said to be conditionally independent given X_3 (Agresti, 2013) if

$$p_{i_1 i_2 i_3} = \frac{p_{i_1+i_3} p_{+i_2 i_3}}{p_{++i_3}}, \ \forall \ i_1, i_2, i_3 = 0, 1. \tag{2.17}$$

Theorem 2.1. Definition 2.2 is equivalent to Definition 2.3.

Proof. **(A)** Definition 2.2 $\Rightarrow$ Definition 2.3.

From Definition 2.2, we have, for $i_3 = 0, 1$,

$p_{i_1 i_2 i_3} = \frac{p_{i_1(1-i_2)i_3} p_{(1-i_1)i_2 i_3}}{p_{(1-i_1)(1-i_2)i_3}}$.

$\Leftrightarrow p_{i_1 i_2 i_3} = \frac{(p_{+i_2 i_3} - p_{i_1 i_2 i_3})(p_{i_1+i_3} - p_{i_1 i_2 i_3})}{(p_{++i_3} - p_{+i_2 i_3} - p_{i_1+i_3} + p_{i_1 i_2 i_3})}$.

$\Leftrightarrow p_{i_1 i_2 i_3} = \frac{p_{i_1+i_3} p_{+i_2 i_3}}{p_{++i_3}}$.

(B) Definition 2.2 $\Leftarrow$ Definition 2.3.

Again, we have, $\forall \ i_1, i_2, i_3 = 0, 1$,

$p_{i_1+i_3} = p_{i_1 i_2 i_3}(1 + \frac{p_{i_1(1-i_2)i_3}}{p_{i_1 i_2 i_3}})$,

$p_{+i_2 i_3} = p_{i_1 i_2 i_3}(1 + \frac{p_{(1-i_1)i_2 i_3}}{p_{i_1 i_2 i_3}})$,

$p_{++i_3} = p_{i_1 i_2 i_3}(1 + \frac{p_{i_1(1-i_2)i_3}}{p_{i_1 i_2 i_3}} + \frac{p_{(1-i_1)i_2 i_3}}{p_{i_1 i_2 i_3}} + \frac{p_{(1-i_1)(1-i_2)i_3}}{p_{i_1 i_2 i_3}})$.

If X_1 and X_2 are conditionally independent given X_3, from Definition 2.3, we get

$\forall \ i_1, i_2, i_3 = 0, 1$,

$(1 + \frac{p_{i_1(1-i_2)i_3}}{p_{i_1 i_2 i_3}})(1 + \frac{p_{(1-i_1)i_2 i_3}}{p_{i_1 i_2 i_3}}) = (1 + \frac{p_{i_1(1-i_2)i_3}}{p_{i_1 i_2 i_3}} + \frac{p_{(1-i_1)i_2 i_3}}{p_{i_1 i_2 i_3}} + \frac{p_{(1-i_1)(1-i_2)i_3}}{p_{i_1 i_2 i_3}})$.

$\Leftrightarrow p_{i_1 i_2 i_3} = \frac{p_{i_1(1-i_2)i_3} p_{(1-i_1)i_2 i_3}}{p_{(1-i_1)(1-i_2)i_3}}$.

This completes the proof.

2.6.2 Unknown parameters and Conditional Independence Model

In this section we establish a relation between the unknown parameters in a conditional independence model. For example, we consider a 3-way contingency table each having 2 levels: 0 and 1. We consider the general saturated model (2.6) in a contingency table for $\mu_{i_1 i_2 i_3}$'s, describing the dependence of $\mu_{i_1 i_2 i_3}$'s on (i_1, i_2, i_3), as

$$\mu_{i_1 i_2 i_3} = exp\{\lambda + \sum_{j=1}^{3} i_j \lambda_j + \sum_{j<k=1}^{3} i_j i_k \lambda_{jk} + i_1 i_2 i_3 \lambda_{123}\},\ \forall\ i_1, i_2, i_3 = 0, 1. \tag{2.18}$$

where $\lambda, \lambda_1, ..., \lambda_{12}, ..., \lambda_{123}$ are unknown parameters. The equation (2.18) can also be written as

$$\mu_{i_1 i_2 i_3} = exp\{\lambda + \sum_{j=1}^{3} i_j \lambda_j + \sum_{j=1}^{2} i_j i_3 \lambda_{j3} + i_1 i_2 (\lambda_{12} + i_3 \lambda_{123}\},\ \forall\ i_1, i_2, i_3 = 0, 1. \tag{2.19}$$

A necessary and sufficient condition for the conditional independence of X_1 and X_2 given X_3 is $(\lambda_{12} + i_3\lambda_{123}) = 0, \forall\ i_3 = 0, 1$, shown in the following theorem. For $i_3 = 0$ and $i_3 = 1$, we have $\lambda_{12} = 0$ and $\lambda_{12} + \lambda_{123} = 0$, respectively. Hence the condition $(\lambda_{12} + i_3\lambda_{123}) = 0, \forall\ i_3 = 0, 1$ can be written equivalently as $\lambda_{12} = 0$ and $\lambda_{123} = 0$.

Theorem 2.2. X_1 and X_2 are conditionally independent given X_3 if and only if $\lambda_{12} = 0$ and $\lambda_{123} = 0$.

Proof. From Definition 2.2, we have, $\forall\ i_1, i_2, i_3 = 0, 1$,

$$p_{i_1 i_2 i_3} = \frac{p_{i_1(1-i_2)i_3} p_{(1-i_1)i_2 i_3}}{p_{(1-i_1)(1-i_2)i_3}}. \tag{2.20}$$

Under model (2.18) and using $\mu_i = np_i$, we can write the RHS of (2.20) as

$$RHS = exp\{\lambda + \sum_{j=1}^{3} i_j\lambda_j + i_1i_3\lambda_{13} + i_2i_3\lambda_{23} + (-1 + 2i_1 + 2i_2 - 3i_1i_2)(\lambda_{12} + i_3\lambda_{123})\}. \tag{2.21}$$

From equation (2.19), we have

$$\mu_{i_1i_2i_3} = exp\{\lambda + \sum_{j=1}^{3} i_j\lambda_1^{X_j} + \sum_{j=1}^{2} i_ji_3\lambda_{11}^{X_jX_3} + i_1i_2(\lambda_{11}^{X_1X_2} + i_3\lambda_{111}^{X_1X_2X_3})\},\ \forall\ i_1, i_2, i_3 = 0, 1. \tag{2.22}$$

Equation (2.20) must hold if

$$exp\{i_1i_2(\lambda_{12} + i_3\lambda_{123})\} = exp\{(-1 + 2i_1 + 2i_2 - 3i_1i_2)(\lambda_{12} + i_3\lambda_{123})\}, \tag{2.23}$$

i.e., when

$$exp\{(-1 + 2i_1 + 2i_2 - 4i_1i_2)(\lambda_{11}^{X_1X_2} + i_3\lambda_{111}^{X_1X_2X_3})\} = 1. \tag{2.24}$$

(A) If part: If the condition $\lambda_{12} = 0$ and $\lambda_{123} = 0$, i.e., $(\lambda_{12} + i_3\lambda_{123}) = 0$ is true, then

$$exp\{(-1 + 2i_1 + 2i_2 - 4i_1i_2)(\lambda_{12} + i_3\lambda_{123})\} = exp\{0\} = 1, \tag{2.25}$$

and hence the conditional independence between X_1 and X_2 given X_3 holds.

(B) Only if part: We assume that X_1 and X_2 are conditionally independent given X_3 and we must have

$$exp\{(-1 + 2i_1 + 2i_2 - 4i_1i_2)(\lambda_{12} + i_3\lambda_{123})\} = 1, \tag{2.26}$$

i.e.,

$$(-1 + 2i_1 + 2i_2 - 4i_1i_2)(\lambda_{12} + i_3\lambda_{123}) = 0. \tag{2.27}$$

But we have no guarantee that $(-1 + 2i_1 + 2i_2 - 4i_1i_2) = 0,\ \forall\ i_1, i_2$. Hence

$$(\lambda_{12} + i_3\lambda_{123}) = 0. \tag{2.28}$$

This completes the proof.

Example 2.3 (Continued). We consider the incomplete contingency table in Example 2.3 and discuss the independence between Primary Infection (**PI**) and Secondary Infection (**SI**). From Table 2.4, we have

$P(\mathbf{PI}) = p_{11} + p_{12}, P(\mathbf{PI} \text{ and } \mathbf{SI}) = p_{11}, P(\mathbf{PI} \text{ and } \mathbf{SI}^{\mathbf{C}}) = p_{12}, P(\mathbf{PI}^{\mathbf{C}} \text{ and } \mathbf{SI}^{\mathbf{C}}) = p_{22} = 1 - p_{11} - p_{12}, P(\mathbf{SI} \mid \mathbf{PI}) = \frac{p_{11}}{p_{11}+p_{12}}, P(\mathbf{PI} \mid \mathbf{SI}^{\mathbf{C}}) = \frac{p_{12}}{p_{12}+p_{22}}$ and

$$P(\mathbf{PI}^{\mathbf{C}} \mid \mathbf{SI}^{\mathbf{C}}) = \frac{p_{22}}{p_{12} + p_{22}}. \tag{2.29}$$

We are interested in evaluating the effect of immunization. We want to compare the models having no immunization effect with the models having an immunization effect.

Definition 2.4. The effect of immunization on the primary infection is defined by the difference

$$P(\mathbf{PI}) - P(\mathbf{SI} \mid \mathbf{PI}). \tag{2.30}$$

Definition 2.5. The effect of immunization is said to be zero, i.e., the vaccination has no effect on the primary infection if

$$P(\mathbf{PI}) - P(\mathbf{SI} \mid \mathbf{PI}) = 0.$$

Definition 2.6. The effect of immunization is said to be positive, i.e., the vaccination has a positive effect on the primary infection if

$$P(\mathbf{PI}) - P(\mathbf{SI} \mid \mathbf{PI}) > 0.$$

Definition 2.7. The effect of immunization is said to be negative, i.e., the vaccination has a negative effect on the primary infection if

$$P(\mathbf{PI}) - P(\mathbf{SI} \mid \mathbf{PI}) < 0.$$

If there is no immunization effect on the primary infection, from Definition 2.5, we have $P(\mathbf{PI}) = P(\mathbf{SI} \mid \mathbf{PI}) = \frac{P(\mathbf{PI} \text{ and } \mathbf{SI})}{P(\mathbf{PI})}$, i.e.,

$$P(\mathbf{PI} \cap \mathbf{SI}) = P^2(\mathbf{PI}). \tag{2.31}$$

Using (2.29), the equation (2.31) can be expressed as

$$p_{11} = (p_{11} + p_{12})^2. \tag{2.32}$$

The probability of getting the primary infection can be expressed as

$$P(\mathbf{PI}) = P(\mathbf{PI} \cap \mathbf{SI}) + P(\mathbf{PI} \cap \mathbf{SI}^{\mathbf{C}}) = P(\mathbf{SI} \mid \mathbf{PI})P(\mathbf{PI}) + P(\mathbf{SI}^{\mathbf{C}} \mid \mathbf{PI})P(\mathbf{PI}). \tag{2.33}$$

Using (2.30), the effect of immunization can be written as

$$P(\mathbf{PI}) - P(\mathbf{SI} \mid \mathbf{PI}) = P(\mathbf{SI} \mid \mathbf{PI})[P(\mathbf{PI}) - 1] + P(\mathbf{SI}^{\mathbf{C}} \mid \mathbf{PI})P(\mathbf{PI}). \tag{2.34}$$

From the definitions stated above, for positive/ zero/ negative effect of immunization, we consider the following situations:

$$P(\mathbf{PI}) - P(\mathbf{SI} \mid \mathbf{PI}) \gtreqless 0. \tag{2.35}$$

Using (2.34), we can express (2.35) as

$$\frac{P(\mathbf{PI})}{1 - P(\mathbf{PI})} \gtreqless \frac{P(\mathbf{SI} \mid \mathbf{PI})}{P(\mathbf{SI}^{\mathbf{C}} \mid \mathbf{PI})}. \tag{2.36}$$

The LHS of (2.36) represents the odds of the primary infection and the RHS of (2.36) represents the conditional odds of getting the secondary infection given that the calf had

the primary infection. If there is no immunization effect, these two odds will be equal.

From equations in (2.29), we have

$$\frac{P(\mathbf{PI})}{1-P(\mathbf{PI})} = \frac{p_{11}+p_{12}}{p_{22}} \text{ and } \frac{P(\mathbf{SI} \mid \mathbf{PI})}{P(\mathbf{SI^C} \mid \mathbf{PI})} = \frac{\frac{p_{11}}{p_{11}+p_{12}}}{\frac{p_{12}}{p_{11}+p_{12}}} = \frac{p_{11}}{p_{12}}. \tag{2.37}$$

If there is no immunization effect of the vaccination, from (2.36) and (2.37), we can write

$$\frac{p_{11}+p_{12}}{p_{22}} = \frac{p_{11}}{p_{12}}. \tag{2.38}$$

Theorem 2.3: Definition 2.5 $\equiv P(\mathbf{SI} \mid \mathbf{PI}) = \frac{P(\mathbf{PI} \mid \mathbf{SI^C})}{P(\mathbf{PI^C} \mid \mathbf{SI^C})}$

Proof. **(A)** Definition 2.5 $\Rightarrow P(\mathbf{SI} \mid \mathbf{PI}) = \frac{P(\mathbf{PI} \mid \mathbf{SI^C})}{P(\mathbf{PI^C} \mid \mathbf{SI^C})}$

If there is no immunization effect on the primary infection, we have

$$\frac{p_{11}+p_{12}}{p_{22}} = \frac{p_{11}}{p_{12}}, \text{ i.e.,}$$

$$\frac{p_{11}}{p_{11}+p_{12}} = \frac{p_{12}}{p_{22}}. \tag{2.39}$$

From the equations in (2.29), the LHS of (2.39) is the probability of secondary infection given that the calf has the primary infection, i.e., $P(\mathbf{SI} \mid \mathbf{PI})$ and the RHS of (2.39) can be expressed as

$$\frac{p_{12}}{p_{22}} = \frac{\frac{p_{12}}{p_{12}+p_{22}}}{\frac{p_{22}}{p_{12}+p_{22}}} = \frac{P(\mathbf{PI} \mid \mathbf{SI^C})}{P(\mathbf{PI^C} \mid \mathbf{SI^C})}.$$

(B) $P(\mathbf{SI} \mid \mathbf{PI}) = \frac{P(\mathbf{PI} \mid \mathbf{SI^C})}{P(\mathbf{PI^C} \mid \mathbf{SI^C})} \Rightarrow$ Definition 2.5

If $P(\mathbf{SI} \mid \mathbf{PI}) = \frac{P(\mathbf{PI} \mid \mathbf{SI^C})}{P(\mathbf{PI^C} \mid \mathbf{SI^C})}$ is true, we have

$$\frac{P(\mathbf{PI} \cap \mathbf{SI^C})}{P(\mathbf{PI^C} \cap \mathbf{SI^C})} = \frac{P(\mathbf{PI} \cap \mathbf{SI})}{P(\mathbf{PI})}. \tag{2.40}$$

From (2.29), (2.40) can be expressed as

$$\frac{p_{12}}{p_{22}} = \frac{p_{11}}{p_{11}+p_{12}}.$$

This completes the proof.

In the next section, we discuss the connection between undirected graphs and conditional independence. We present the graph theoretical interpretation of conditional independence.

2.7 Undirected Graphs and Conditional Independence

A graph comprises *nodes* (also called vertices) connected by *links* (also known as edges or arcs). Each node represents a random variable (or group of random variables), and the links express the association between these variables. In *directed* graphs (Koller, 2009) the links have arrows representing the direction or path of information flow. In these graphs, we can only move from one node to other node following the direction of the arrow. The other major class of graphs, which will be discussed in next subsection, is *undirected graphs* (Bishop, 2006; Koller, 2009), in which the links do not carry arrows and have no directional significance. In these graphs, we can move from a node A to a node B and come back to node A following the path if there is ann edge between node A and node B. *Directed* graphs are useful for expressing the causal relationships between random variables, i.e., if the occurrence of one variable causes the other variable(s), whereas *undirected* graphs are better suited in expressing connections between random variables, i.e., whether they are related to each other. These graphs are called *probabilistic graphical models* when we express these associations using probability distributions. These graphs provide a way to visualize the joint probability distribution of these random variables including conditional independence structure among these

variables. The graph captures the way in which the joint distribution over all of the random variables can be decomposed into a product of marginal and/ or conditional probabilities each depending only on a subset of the variables. Undirected graphical models exhibit the pairwise Markov property that any two non-adjacent random variables are conditionally independent given all other variables. In the next subsection, we discuss Undirected graphs and define Markov Blanket of a node in these graphs. We discuss conditional independence properties and show conditional independence structures by undirected graphs in subsection 2.7.2. We define cliques and potential functions related to cliques to describe the factorization property of undirected graphs in subsection 2.7.3.

2.7.1 Markov Random Fields or Undirected Graphs

Our goal is to establish connections between Markov Random Fields and conditional independence among the random variables. A Markov random field (Koller, 2009), also known as a Markov network or an undirected graphical model, has a set of nodes each of which corresponds to a variable or group of variables, as well as a set of links each of which connects a pair of nodes. The links are undirected, that is they do not carry arrows. In the case of undirected graphs, it is convenient to begin with a discussion of conditional independence properties. To discuss the conditional independence structure, we need to consider all the adjacent nodes connected by links from a particular node. In order to do that, we define Markov blanket before discussing Conditional Independence properties. A Markov blanket (Bishop, 2006) of a node, as the name suggests, includes all the nodes connected to the node by edges.

Definition 2.4. For an undirected graph, the Markov blanket of a node X_i consists of the set of neighboring nodes. It has the property that the conditional distribution of X_i, conditioned on all the remaining variables in the graph, is dependent only on the variables in the Markov blanket. The Markov blanket for an undirected graph takes a

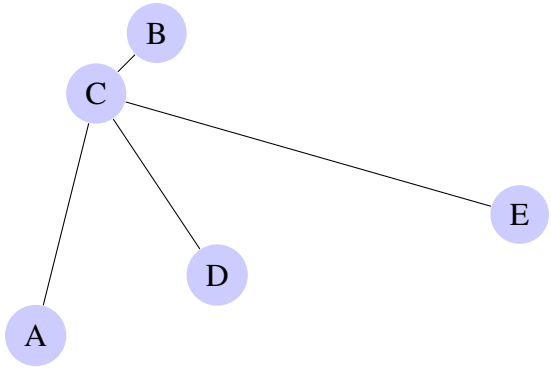

Figure 2.1: Markov Blanket

particularly simple form, because a node will be conditionally independent of all other nodes conditioned only on the neighboring nodes. In Figure 2.1, nodes {A, B, D, E} creates a Markov Blanket of node C. In the next subsection, we discuss the conditional independence properties from undirected graphical models.

2.7.2 Conditional Independence Properties

The graph structure in a Markov network can be viewed as displaying a set of independence assumptions among the random variables. In Markov networks, probabilistic influence flows along the undirected paths in the graph, but it is "blocked" if we observe the intervening nodes. For example, suppose A, B and C all three are preparing for an exam. A and C study together whereas, B and C study together. But A and B do not study together. A and B is connected only through C. If C takes the exam before A and

B, there is no way A and C can influence each other and/ or B and C can influence each other, i.e., A and B will be disconnected if C is observed. Suppose that in an undirected graph we identify four sets of nodes, denoted A, B, C and D, and that we consider the conditional independence property that A is conditionally independent of B given C, i.e., $A \perp\!\!\!\perp B \mid C$. Note that each of A, B, C and D could be a single node or a collection of nodes.

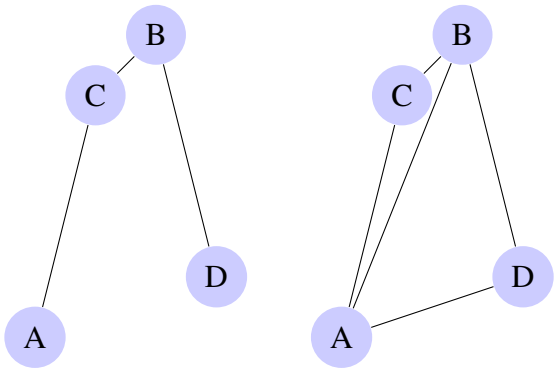

Figure 2.2: Graphical representation of conditional independence and dependence. **Left**: $A \perp\!\!\!\perp B \mid C$. **Right**: $A \not\perp\!\!\!\perp B \mid C$

To test whether this property is satisfied by a probability distribution which is defined by a graph, we consider all possible paths that connect nodes in set A to nodes in set B. If all such paths pass through one or more nodes in set C, then all such paths are "blocked" (Koller, 2009) and so the conditional independence property holds. However, if there is at least one such path that is not blocked, then the property does not necessarily hold, or more precisely there will exist at least some distributions corresponding to the graph that do not satisfy this conditional independence relation. This is illustrated with an example in Figure 2.2. An alternative way to view the conditional independence test is to imagine removing all nodes in set C from the graph together with any links that

connect to those nodes. We then ask if there exists a path that connects any node in A to any node in B. If there are no such paths, then the conditional independence property must hold (Koller, 2009). In Figure 2.2 left, there is no direct path from A to B and the path from A to B is blocked by C. We can say that A is conditionally independent of B given C. But in Figure 2.2 right, A is not conditionally independent of B given C and D since there exists a direct path between A and B. Using this concept, different conditional independences are exhibited in Figure 2.3 using undirected graphs.

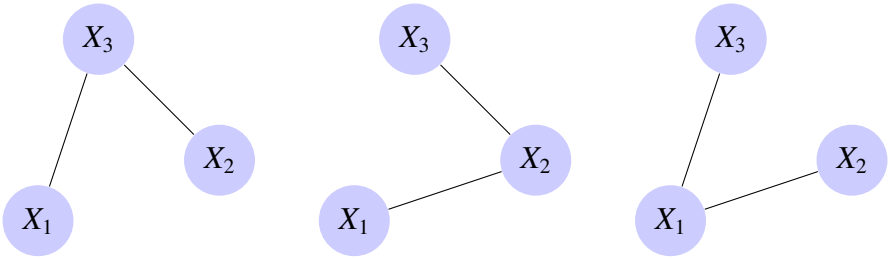

Figure 2.3: All possible conditional independences among X_1, X_2 and X_3. **Left:** $X_1 \perp\!\!\!\perp X_2 \mid X_3$ **Center:** $X_1 \perp\!\!\!\perp X_3 \mid X_2$ **Right:** $X_2 \perp\!\!\!\perp X_3 \mid X_1$

In Figure 2.3, three types of conditional independences are shown when we have three variables X_1, X_2 and X_3. In the Left diagram, there is no direct path between X_1 and X_2. There exists a path between X_1 and X_2 blocked by node X_3. Hence, $X_1 \perp\!\!\!\perp X_2 \mid X_3$. Similarly, in the other two diagrams the paths between X_1 and X_3 and X_2 and X_3 are blocked by X_2 and X_1, respectively. We have presented conditional independence structures using undirected graphs. In the next subsection, we will discuss how these structures can be represented by probability distributions using potential functions.

2.7.3 Factorization properties

In this subsection, we discuss how we can express the joint probability distribution of the random variables from an undirected graph. The conditional independence of two random variables X_i and X_j given the others represented by the undirected graph implies that if the nodes X_i and X_j are not connected by a direct link then the corresponding variables are conditionally independent given all other variables representing the remaining nodes (Bishop, 2006). This follows from the fact that there is no direct path between these two nodes, and all other paths pass through nodes that are observed, and hence those paths are blocked. This conditional independence property can be expressed as

$$p(X_i, X_j \mid X_{\{i,j\}}) = p(X_i \mid X_{\{i,j\}})p(X_j \mid X_{\{i,j\}}),$$

where $X_{i,j}$ denotes the set of nodes of all variables with X_i and X_j removed. The factorization of the joint distribution must therefore be such that X_i and X_j do not appear in the same factor in order for the conditional independence property to hold for all possible distributions belonging to the graph. This leads us to consider a graphical concept called a clique, which is defined below.

Definition 2.5 A clique (Bishop, 2006) is defined as a subset of the nodes in a graph such that there exists a link between all pairs of nodes in the subset.

In other words, the set of nodes in a clique is fully connected. Furthermore, a maximal clique is a clique such that it is not possible to include any other nodes from the graph in the set without it remaining a clique. These concepts are illustrated by the undirected graph over four variables shown in Figure 2.4.

This graph has five cliques of two nodes each given by {A, C}, {C, B}, {B, E}, {E, C}

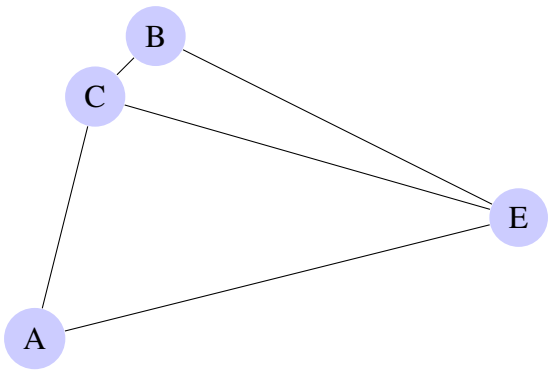

Figure 2.4: Cliques

and {E, A} as well as two maximal cliques of three nodes each given by {A, C, E} and {B, C, E}. The set {A, C, B, E} is not a clique since we are missing and edge between A and B. We can therefore define the factors in the decomposition of the joint distribution to be functions of the variables in the cliques. In fact, we can consider functions of the maximal cliques, without loss of generality, because other cliques must be subsets of maximal cliques.

Let us denote $\mathbf{C}$ as the collection of all maximal cliques and $\mathbf{X}_c$ be the nodes in c $\in$ $\mathbf{C}$. Then the joint distribution is written as a product of *potential functions* $\psi_c(\mathbf{x}_c)$ over the maximal cliques of the graph

$$p(\underline{X}) = \frac{1}{Z} \prod_{c \in \mathbf{C}} \psi_c(\mathbf{X}_c),$$

where the quantity Z is called the *partition function* and is given by

$$Z = \sum_{x} \prod_{c \in \mathbf{C}} \psi_c(\mathbf{X}_c).$$

These potential functions represent the local associations among the random variables involved in a maximal clique. These functions may also be represented by probability distributions. Z is the standardized constant such that $\sum_{\underline{x}} p(\underline{X}) = 1$. For example,

consider a simplest case with three nodes X_1, X_2 and X_3 defining the conditional independence between X_1, X_2 given X_3, as illustrated in Figure 2.5. To reach X_2 from X_1,

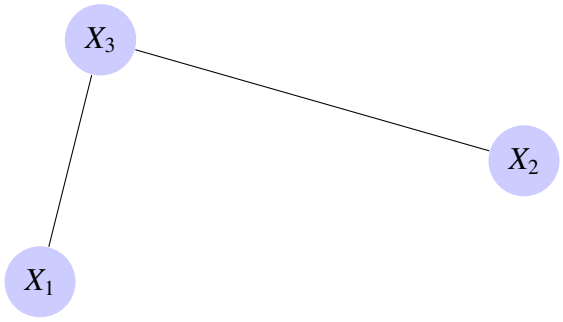

Figure 2.5: Conditional Independence of X_1 and X_2 given X_3. $X_1 \perp\!\!\!\perp X_2 \mid X_3$

we need to go through X_3. In this case, we have two maximal cliques given by $\{X_1, X_3\}$ and $\{X_2, X_3\}$. We define two potential functions given by $\psi(X_1, X_3)$ and $\psi(X_2, X_3)$ based on the maximal cliques $\{X_1, X_3\}$ and $\{X_2, X_3\}$, respectively. The joint distribution can be written as

$$p(X_1, X_2, X_3) \propto \psi(X_1, X_3)\, \psi(X_2, X_3). \tag{2.41}$$

If we choose $\psi(X_1, X_3) = p(X_3)\, p(X_1 \mid X_3)$ and $\psi(X_2, X_3) = p(X_2 \mid X_3)$ and $Z = 1$, we have

$$p(X_1, X_2, X_3) = p(X_3)\; p(X_1 \mid X_3) p(X_2 \mid X_3). \tag{2.42}$$

Dividing both sides of (2.30) by $p(X_3)$ (assuming $p(X_3) > 0$), we have

$$\frac{p(X_1, X_2, X_3)}{p(X_3)} = p(X_1, X_2 \mid X_3) = \; p(X_1 \mid X_3) p(X_2 \mid X_3), \tag{2.43}$$

the condition for conditional independence between X_1 and X_2 given X_3. This condition is equivalent to the condition defined in Definition 2.2. This can be generalized when to m categorical variables. We have shown how the undirected graph in Figure 2.5

implies the conditional independence between X_1 and X_2 given X_3. Using theorem 2.2 and the undirected graphs to describe the conditional association, we have established the connection between the undirected graphs and the unknown λ parameters.

Table 2.5: Models with Design matrices

Models	Design matrices	Parameters assumed zero
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{123}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|c} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{12} \\ \lambda_{13} \\ \lambda_{23} \end{pmatrix}$
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|c} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{13} \\ \lambda_{23} \\ \lambda_{123} \end{pmatrix}$

Table 2.6: Models with Design matrices

Models	Design matrices	Parameters assumed zero
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{13}$	$\mathbf{X}_1 = \left[\begin{array}{cccc\|c} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{12} \\ \lambda_{23} \\ \lambda_{123} \end{pmatrix}$
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{23}$	$\mathbf{X}_1 = \left[\begin{array}{cccc\|c} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{12} \\ \lambda_{13} \\ \lambda_{123} \end{pmatrix}$

Table 2.7: Models with Design matrices

Models	Design matrices	Parameters assumed zero
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{123}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{13} \\ \lambda_{23} \end{pmatrix}$
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{13} + \lambda_{123}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{12} \\ \lambda_{23} \end{pmatrix}$

Table 2.8: Models with Design matrices

Models	Design matrices	Parameters assumed zero
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{23} + \lambda_{123}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{12} \\ \lambda_{13} \end{pmatrix}$
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{13}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{23} \\ \lambda_{123} \end{pmatrix}$

Table 2.9: Models with Design matrices

Models	Design matrices	Parameters assumed zero
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{12} + \lambda_{23}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{13} \\ \lambda_{123} \end{pmatrix}$
$\lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{13} + \lambda_{23}$	$\mathbf{X}_1 = \left[\begin{array}{c\|ccc\|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right]$	$\begin{pmatrix} \lambda_{12} \\ \lambda_{123} \end{pmatrix}$

Chapter 3

Estimation of Unknown Parameters and Goodness of Fit Tests

3.1 Summary

In this chapter, we discuss the estimation of the unknown parameters and/ or cell probabilities in an LLM. We discuss Maximum Likelihood Estimation (section 3.2) and Minimum Discrimination Information Estimation (section 3.3) as the two estimation procedures. In section 3.4 we investigate the goodness-of-fit of the unsaturated models using different criterion functions.

3.2 Maximum Likelihood Estimation

We consider the case of single multinomial sampling where the total sample size n is fixed. The subscript i is used represent the cells. In Example 2.1, we define $p_i =$

proportion of the observations belonging to the i-th cell, $p_{i(o)}=\frac{n_i}{n}$ = observed proportion of observations belonging to the i-th cell. For multinomial sampling (Bishop et al., 2007) with cell count $\underline{n}$ and the corresponding probabilities $\underline{p}$, the complete likelihood function can be written as

$$f(\underline{n}) = \frac{n!}{\prod_i n_i!} \prod_i (p_i)^{n_i}. \tag{3.1}$$

To avoid notational complications, we assume the design matrix $\mathbf{X} = \mathbf{M}$, where m_{ij} = the (i,j)-th element of the design matrix $\mathbf{X}$. The exponent (k) represent the k-th loglinear model in the above expressions. The kernel of the log-likelihood under model (3.1) can be written as

$$l = \sum_{i=1}^{N} n_i \ln p_i = \sum_{i=1}^{N} n_i \sum_{j=1}^{s} m_{ij}^{(k)} \lambda_j^{(k)}, \tag{3.2}$$

given $\sum_i p_i = \sum_i \exp(\sum_j m_{ij}^{(k)} \lambda_j^{(k)}) = 1$. Maximizing the log-likelihood in (3.2) with the constraint $\sum_i p_i = 1$ is equivalent to maximizing the objective function (Klimova et al., 2012)

$$l^* = \sum_{i=1}^{N} n_i \sum_{j=1}^{s} m_{ij}^{(k)} \lambda_j^{(k)} - \alpha(\sum_i \exp(\sum_j m_{ij}^{(k)} \lambda_j^{(k)}) - 1), \tag{3.3}$$

where α is the Lagrange multiplier.

Differentiating (3.3) with respect to $\lambda_1^{(k)}, \lambda_2^{(k)}, ..., \lambda_s^{(k)}$ and α, we get

$$\begin{aligned}
\frac{dl^*}{d\lambda_1^{(k)}} = 0 \Rightarrow \sum_i n_i m_{i1}^{(k)} = \alpha(\sum_i \hat{p}_i m_{i1}^{(k)}), \\
\frac{dl^*}{d\lambda_2^{(k)}} = 0 \Rightarrow \sum_i n_i m_{i2}^{(k)} = \alpha(\sum_i \hat{p}_i m_{i2}^{(k)}), \ldots, \\
\frac{dl^*}{d\lambda_s^{(k)}} = 0 \Rightarrow \sum_i n_i m_{i8}^{(k)} = \alpha(\sum_i \hat{p}_i m_{i8}^{(k)}), \text{ and} \\
\frac{dl^*}{d\alpha} = 0 \Rightarrow \sum_i \hat{p}_i = 1.
\end{aligned} \tag{3.4}$$

Since $m_{i1}^{(k)} = 1, \ \forall i$, i.e., all the elements in 1-st column of $M^{(k)}$ are equal to 1, the first estimating equation can be written as $\sum_i n_i = \sum_i \hat{p}_i$. From the first and last estimating equations in (3.4), we get

$$\alpha = \sum_i n_i = n. \tag{3.5}$$

Using (3.5), the estimating equations can be written as

$$\sum_{i=1}^{N} m_{ij}^{(k)} n.(p_{i(o)} - \hat{p}_i) = 0, \ \forall j = 1, 2, ..., s, \tag{3.6}$$

where $n = \sum_i n_i = \alpha$.

We also define $\underline{p}_0 = (p_{i(o)})_{i=1,2,...,8}$; $\underline{\hat{p}} = (\hat{p}_i)_{i=1,2,...,8}$. The estimating equations for the Maximum Likelihood Estimation (MLE) in (3.6) can be expressed in matrix notation as

$$\mathbf{M}^{(k)'} \underline{\hat{p}} = \mathbf{M}^{(k)'} \underline{p}_0 \tag{3.7}$$

where rank $(\mathbf{M}^{(k)}) = s \leq N$.

If we consider the saturated model, then rank $(\mathbf{M}^{(k)}) = N$. Pre-multilying both sides of (3.7) by $(\mathbf{M}^{(k)})^{-1}$, we get

$$\underline{\hat{p}} = \underline{p}_0.$$

In Table 3.1, we show the counts for each cell in Example 2.1 for some fitted loglinear models using maximum likelihood estimation.

Example 2.1 (Revisited). Table 3.1 displays the observed cell probabilities and the estimated cell probabilities under some fitted unsaturated models. The column $\underline{p}_o$ represents the observed cell probabilities (4 decimal places), the columns with heading $\underline{\hat{p}}$

Table 3.1: ML estimates for different unsaturated models

Cell	$\underline{p}_o$	$\hat{\underline{p}}$ $\lambda_{12}, \lambda_{123} = 0$	$\hat{\underline{p}}$ $\lambda_{13}, \lambda_{123} = 0$	$\hat{\underline{p}}$ $\lambda_{23}, \lambda_{123} = 0$	$\hat{\underline{p}}$ $\lambda_{123} = 0$
(0,0,0)	0.2572	0.2536	0.2481	0.2642	0.2568
(0,0,1)	0.0263	0.0264	0.0355	0.0194	0.0268
(0,1,0)	0.3284	0.3319	0.3229	0.3214	0.3288
(0,1,1)	0.0166	0.0165	0.0220	0.0235	0.0161
(1,0,0)	0.1364	0.1400	0.1455	0.1446	0.1368
(1,0,1)	0.0299	0.0298	0.0208	0.0217	0.0295
(1,1,0)	0.1867	0.1832	0.1921	0.1785	0.1863
(1,1,1)	0.0185	0.0186	0.0131	0.0268	0.0189

show the estimated cell probabilities with a description of the fitted model. We will now consider the MLE of the cell probabilities in the incomplete contingency table described in Example 2.3. We consider two different models as H_0 : no immunization effect on the primary infection and H_1 : existence of an immunization effect. Before we start writing the likelihood under these two models, it is convenient to define regular and curved exponential families of distributions. We will be using sufficiency principle in exponential families.

Definition 2.8. A family $\{P_{\underline{\eta}}\}$ of distributions is said to be a regular exponential family of order s if the distributions $P_{\underline{\eta}}$ have densities of the form (Lehmann & Casella, 1998)

$$p(\underline{x}, \underline{\eta}) = h(\underline{x}) \exp\left[\sum_{i=1}^{s} \eta_i T_i(\underline{x}) - A(\underline{\eta})\right], \tag{3.8}$$

where η_i's and A are real-valued functions of the parameters and T_i's are real-valued statistics, $\underline{x}$ is a point in the sample space $\varkappa$, the support of the density and $\underline{\eta} = (\eta_1, \eta_2, ..., \eta_s)'$: natural parameters.

Definition 2.9. When the natural parameters of an exponential family are related in a nonlinear way, (3.8) is said to form a curved exponential family of order q (Lehmann & Casella, 1998). The densities in a curved exponential family can be expressed as

$$p(\underline{x}, \underline{\eta}(\underline{\theta})) = h(\underline{x}) \exp\left[\sum_{i=1}^{s} \eta_i(\underline{\theta}) T_i(\underline{x}) - B(\underline{\eta}(\underline{\theta}))\right], \tag{3.9}$$

where $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_q)'$ is a vector of parameters in $\Re^q, q < s$.

Example 2.3 (Revisited). Under non immunization effect, let $P(\mathbf{PI}) = p_{11} + p_{12} = \pi$. Then we have $P(\mathbf{PI^c}) = p_{22} = (1-\pi)$ and $P(\mathbf{SI} \cap \mathbf{PI}) = P^2(\mathbf{PI}) = p_{11} = \pi^2$, from (2.29). $\Rightarrow P(\mathbf{SI}^c \cap \mathbf{PI}) = P(\mathbf{PI}) - P(\mathbf{PI} \cap \mathbf{SI}) = p_{12} = \pi - \pi^2 = \pi(1-\pi)$ and $P(\mathbf{SI^c} \cap \mathbf{PI}^c) = p_{22} = 1 - \pi$. If n_{11}, n_{12} and n_{22} are the sample counts and if the number of infected calves is

Table 3.2: Probabilities under non immunization effect of primary infection

		Secondary infection		Row Total
		Yes (1)	No (2)	
Primary infection	Yes (1)	π^2	$\pi(1-\pi)$	π
	No (2)	–	$1-\pi$	$1-\pi$
Column Total		π^2	$1-\pi^2$	1

a multinomial random variable with parameters N and $(p_{11}, p_{12}, p_{22})'$, the kernel of the likelihood under $H_0 \cup H_1$, is

$$L_{H_0 \cup H_1} = \exp[n_{11} \ln p_{11} + n_{12} \ln p_{12} + n_{22} \ln p_{22}]. \tag{3.10}$$

Since $n_{22} = n - n_{11} - n_{12}$ and $p_{11} + p_{12} + p_{22} = 1$, equation (3.10) can be written as

$$L_{H_0 \cup H_1} = \exp[n_{11} \ln \frac{p_{11}}{p_{22}} + n_{12} \ln \frac{p_{12}}{p_{22}} - n \ln(1 + \frac{p_{11}}{p_{22}} + \frac{p_{12}}{p_{22}})]. \tag{3.11}$$

The canonical parameters are $\eta_1 = \ln \frac{p_{11}}{p_{22}}$ and $\eta_2 = \ln \frac{p_{12}}{p_{22}}$. From definition 2.8, this is a regular exponential family of order 2 $(=s)$ with sufficient statistic (n_{11}, n_{12}) and $A(\underline{\eta}) = n \ln(1 + \frac{p_{11}}{p_{22}} + \frac{p_{12}}{p_{22}})$. Under $H_0 \cup H_1$, the kernel of the log-likelihood is

$$l_{H_0 \cup H_1} = n_{11} \ln \frac{p_{11}}{p_{22}} + n_{12} \ln \frac{p_{12}}{p_{22}} - n \ln(1 + \frac{p_{11}}{p_{22}} + \frac{p_{12}}{p_{22}}). \tag{3.12}$$

Differentiating (3.12) with respect to the parameters p_{11} and p_{12} respectively and equating them to zero, subject to $p_{11} + p_{12} + p_{22} = 1$, we get

$$\hat{p}_{11} = \frac{n_{11}}{n}, \hat{p}_{12} = \frac{n_{12}}{n}, \hat{p}_{22} = \frac{n_{22}}{n}.$$

Under H_0, the kernel of the likelihood is

$$L_{H_0} = \exp[(2n_{11} + n_{12}) \ln \pi + (n_{12} n_{22}) \ln(1 - \pi)]. \tag{3.13}$$

Equation (3.13) can be expressed in matrix notation as

$$L_{H_0} = \exp[(n_{11}, n_{12}, n_{22}) \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \ln \pi \\ \ln(1 - \pi) \end{pmatrix}].$$

The likelihood in (3.13) has two canonical parameters $\eta_1(\pi) = \ln \pi$ and $\eta_2(\pi) = \ln(1 - \pi)$ with only a single parameter $\pi \in (0, 1)$. From definition 2.9, the model under H_0 is a

curved exponential family (Casella et al., 2007) of order 1(=q). The sufficient statistic is $\underline{T} = (T_1, T_2) = (2n_{11} + n_{12}, n_{12} + n_{22})$. Under H_0, the kernel of the log-likelihood is

$$l_{H_0} = (2n_{11} + n_{12}) \ln \pi + (n_{12} n_{22}) \ln(1 - \pi). \tag{3.14}$$

Differentiating equation (3.14) with respect to π and equating it to zero, we get

$$\hat{\pi} = \frac{2n_{11} + n_{12}}{2n_{11} + 2n_{12} + n_{22}} = \frac{T_1}{T_1 + T_2} = 0.4940.$$

Table 3.3 shows the estimated cell counts (2 decimal places) under no immunization effect.

Table 3.3: Cell counts under non immunization effect of primary infection

		Secondary infection		Row Total
		Yes (1)	No (2)	
Primary infection	Yes (1)	38.07	38.99	π
	No (2)	–	78.94	78.94
Column Total		38.07	117.93	156

3.3 MDI Estimation

The aim of Minimum Discrimination Information (MDI) estimation (Gokhale et al., 1978) is to get a good fit to the observed contingency table. This is achieved by fitting loglinear models to the data using a minimal number of parameters satisfying some constraints on the linear combination of observed cell entries. The constraints could be given externally by the experimenter (*External Constraint Problems*) or could be

matching marginals to the observed contingency table (*Internal Constraint Problems*), which will be discussed later in this section. After fitting a loglinear model we want to test how good is the fit and compare it with other unsaturated loglinear models using the MDI statistic (Section 3.4.1) which can be used as a criterion function for selecting the best fitted model. We will use the notations $p_i,\ \hat{p}_i,\ p_{io}$ as the unknown probability, estimated probability and the observed probability of the w-th cell in a contingency table, respectively.

We use the Cross-Entropy or MDI (Kullback, 1959) as a measure of "closeness" between two probability distributions. Suppose there are two probability distributions or contingency tables defined over the same set of cells, say p_i and π_i, satisfying $\sum_i p_i = \sum_i \pi_i = 1$.

Definition 3.1. The MDI is defined by

$$I(\underline{p} : \underline{\pi}) = \sum_i p_i \ln(\frac{p_i}{\pi_i}). \tag{3.15}$$

We note that the distribution π_i is arbitrary. A suitable choice of π_i will be made according to the problem of interest, discussed later. $I(\underline{p} : \underline{\pi})$ is considered as a measure of the deviation from $\underline{p}$ to $\underline{\pi}$.

Many problems in the analysis of contingency tables may be characterized as estimating a distribution or contingency table subject to certain constraints and then comparing the estimated table with an observed table to determine whether the observed table satisfies a null hypothesis implied by a set of constraints. In accordance with the principle of MDI estimation we select that member of the family of probability distri-

butions satisfying the constraints

$$X^{'}\underline{p} = \underline{\theta}, \tag{3.16}$$

which minimizes the MDI $I(\underline{p} : \underline{\pi})$ over the family of suitable probability distributions. The design matrix X is $N \times s_1$ with $\text{rank}(X) = s_1 \leq N$ and the vector of constraining values $\underline{\theta}$ is $s_1 \times 1$. We denote the MDI estimate of $\underline{p}$ by $\underline{\hat{p}}$ so that

$$I(\underline{\hat{p}} : \underline{\pi}) = \sum_i \hat{p}_i \ln(\frac{\hat{p}_i}{\pi_i}) = min\ I(\underline{p} : \underline{\pi}). \tag{3.17}$$

The constraints in (3.16) specify some linear functions of cell counts in the estimated contingency table are equal to the same linear functions of cell counts in the observed contingency table for internal constraint problems or to some externally given values for external constraint problems. The equation (3.16) can also be written as

$$\sum_i x_{ij} p_i = \theta_j, \forall\, j = 1, 2, \ldots, s_1. \tag{3.18}$$

To satisfy $\sum_i p_i = 1$, we take $x_{i1} = 1, \forall i = 1, 2, \ldots, N$ and $\theta_1 = 1$. We find the value of $\underline{p}$ that minimizes (3.15) subject to the constraints (3.18) using Lagrange multipliers. The objective function to minimize can be written as

$$\sum_i p_i \ln(\frac{p_i}{\pi_i}) - \sum_{j=1}^{s_1} \lambda_j (\sum_i x_{ij} p_i - \theta_j), \tag{3.19}$$

where λ_j's are Lagrange multipliers. Differentiating (3.19) with respect to p_i's and equating to zero, we get

$$\ln(\frac{p_i}{\pi_i}) = \sum_{j=1}^{s_1} x_{ij}\lambda_j = \lambda_1 + \sum_{j=2}^{s_1} x_{ij}\lambda_j - 1, \forall i = 1, 2, \ldots, N. \tag{3.20}$$

If π_i's are known, the equations in (3.20) can be written as

$$\ln(p_i) = \lambda_1^* + \sum_{j=2}^{s_1} x_{ij}\lambda_j, \forall i = 1, 2, \ldots, N, \tag{3.21}$$

λ_j's are also called the natural parameters. This also explains why we consider loglinear form for fitting models in contingency table. In case of ICP, the constraints can be written in matrix notation as

$$X' \underline{\hat{p}} = X' \underline{p}_0, \tag{3.22}$$

where $\underline{p}_0$ is the vector of observed probabilities and rank(X)= $s_1 \leq N$. Equation (3.22) is equivalent to the Maximum Likelihood Estimating Equations in (3.3).

We now discuss two different classes of MDI estimation in a contingency tables with examples. In one class of problems the constraints specify some requirement external to the observed values, for example, that a set of marginals, have specified values as determined by genetic or other theory or that marginals be homogeneous or that the distribution satisfy certain symmetry conditions. These are called *External Constraints Problems* (ECP). In such problems $\underline{\pi}$ is taken to be an observed contingency table, that is, $n\underline{\pi} = \underline{n} = n\underline{p}_o$, where $n = \sum_i n_i$. We discuss this in the following illustrative example.

Example 3.1. A 2×2 contingency table with some arbitrary counts is shown in Table 0.1. The hypothesis of interest is $H_0 : P(Y = 0) = P(Z = 0)$. The hypothesis H_0 is equivalent to $p_{00} + p_{01} = p_{00} + p_{10}$, i.e., $p_{01} = p_{10}$. This is a linear constraint on the underlying probabilities. We choose $\underline{\pi} = \underline{p}_o$ = vector of observed probabilities. $\underline{p}$ does not satisfy the constraint given by H_0. We find the estimate $\underline{\hat{p}}$ so as to minimize $I(\underline{p} : \underline{p}_o)$ subject to the constraint $p_{01} = p_{10}$. We are choosing $\underline{\hat{p}}$ that satisfy $p_{01} = p_{10}$ and is as close as possible to the observed data in the sense of minimizing (3.15). The constraints can be written as

$$X' \underline{p} = \underline{\theta}, \tag{3.23}$$

Table 3.4: An illustrative example

		Z		Row Total
		No (0)	Yes (1)	
Y	No (0)	35	15	50
	Yes (1)	21	29	50
Column Total		56	44	100

where $\underline{p} = (p_{00}, p_{01}, p_{10}, p_{11})'$, $\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix}$ and $\underline{\theta} = (1, 0)'$. The conditions in (3.23) are

$$p_{00} + p_{01} + p_{10} + p_{11} = 1,$$

$$p_{01} - p_{10} = 0. \tag{3.24}$$

We get the MDI estimates, by solving (3.24), as $\hat{\underline{p}} = (0.352, 0.178, 0.178, 0.292)'$.

In another class of problems the constraints specify that the estimated distribution or contingency table have some set of marginals which are the same as those of an observed contingency table. These are called *Internal Constraints Problems* (ICP). In such cases $\underline{\pi}$ is taken to be either the uniform distribution $\pi_i = \frac{1}{N}$, where N is the number of cells, or a distribution already estimated subject to constraints contained in and implied by the constraints under examination. The latter case includes the classical hypotheses of independence, conditional independence, homogeneity, conditional homogeneity and interaction, all of which can be considered as instances of generalized independence. In case of ICP, the cell probability estimates are equal to the ML estimates, shown in (3.22).

We consider Example 2.1. We want to fit the model to the data where X_1 is conditionally independent of X_2 given X_3. We specify the marginals of X_1X_3 and X_2X_3 table in our fitted model. These marginals are equal to the marginals of X_1X_3 and X_2X_3 in the observed table.

If we consider the transpose of the design matrix X, a 6×8 matrix, as

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and from the observed table, $\underline{\theta}$ as

$\underline{\theta}' = (1, 0.3715, 0.5502, 0.0913, 0.0484, 0.0351)$ and

$\underline{p} = (p_{000}, p_{001}, p_{010}, p_{011}, p_{000}, p_{001}, p_{010}, p_{011})'$,

we have the constraints as

$$\begin{aligned} p_{+++} &= 1 \\ p_{1++} &= 0.3715 \\ p_{+1+} &= 0.5502 \\ p_{++1} &= 0.0913 \\ p_{1+1} &= 0.0484 \\ p_{+11} &= 0.0351 \end{aligned} \tag{3.25}$$

The first constraint in (3.25) satisfy the condition $\sum_i = p_i = 1$. The second constraint

with the first one specify the marginals for X_1 since $p_{+++} - p_{1++} = 1 - 0.3715 = 0.6285$. Similarly marginals for X_2 and X_3 are fixed by the next two constraints. All these four constraints along with the last two, we specify the marginals of X_1X_3 and X_2X_3 table. Solving the six equations with eight unknowns in (3.25), we get the cell probability estimates as $\underline{\hat{p}} = (0.2536, 0.0264, 0.3319, 0.0165, 0.1400, 0.0298, 0.1832, 0.0186)'$. The following table shows the cell probability estimates for different unsaturated models in Example 2.1. The second column ($\underline{p}_o$) represents the observed cell probabilities of the

Table 3.5: Cell probability estimates (MDI) for different unsaturated models

Cell	$\underline{p}_o$	$\underline{\hat{p}}$ $\lambda_{12}, \lambda_{123} = 0$	$\underline{\hat{p}}$ $\lambda_{13}, \lambda_{123} = 0$	$\underline{\hat{p}}$ $\lambda_{23}, \lambda_{123} = 0$	$\underline{\hat{p}}$ $\lambda_{123} = 0$
(0,0,0)	0.2572	0.2536	0.2481	0.2642	0.2568
(0,0,1)	0.0263	0.0264	0.0355	0.0194	0.0268
(0,1,0)	0.3284	0.3319	0.3229	0.3214	0.3288
(0,1,1)	0.0166	0.0165	0.0220	0.0235	0.0161
(1,0,0)	0.1364	0.1400	0.1455	0.1446	0.1368
(1,0,1)	0.0299	0.0298	0.0208	0.0217	0.0295
(1,1,0)	0.1867	0.1832	0.1921	0.1785	0.1863
(1,1,1)	0.0185	0.0186	0.0131	0.0268	0.0189

$2 \times 2 \times 2$ contingency table described in Example 2.1. The third, fourth and the fifth columns represent the MDI estimates of the cell probabilities using MDI estimation when we fit the models $\lambda_{12} = \lambda_{123} = 0$, $\lambda_{13} = \lambda_{123} = 0$ and $\lambda_{23} = \lambda_{123} = 0$, respectively.

After careful consideration, we can see that the estimates from the model $\lambda_{12} = \lambda_{123} = 0$ and $\lambda_{123} = 0$ are closer to the observed cell probabilities compared to other models. In the next section we discuss a few criterion functions for comparing different models.

3.4 Goodness-of-Fit Tests

A good-fitting loglinear model provides a basis for describing the relations among the response and the categorical variables. Standard methods apply for checking the model fit and making inferences about the model parameters. In this section, we discuss three standard criterion functions and a proposed criterion function for selecting the best fitted model. In our thesis, we considered different possible unsaturated models and extended these models to a complete form, i.e., the design matrices are full row rank matrices, i.e., rank(X)= N. We want to test whether the fitted unsaturated model is a good fit to the observed data. The null and the alternative hypotheses can be written as $H_0 :\ \mathbf{D}_0\underline{\lambda}_2 = \underline{0}$ and $H_a :\ \mathbf{D}_0\underline{\lambda}_2 \neq \underline{0}$.

3.4.1 MDI Statistic

To test whether an observed contingency table satisfies the null hypothesis as represented by the MDI estimate we compute a measure of the deviation between the observed distribution and the appropriate estimate of the MDI statistic. The MDI statistic turns out to be

$$2nI(\underline{\hat{p}} : \underline{p}_o) = 2n \sum_w \hat{p}_i \ln(\frac{\hat{p}_i}{p_{io}}), \tag{3.26}$$

which is asymptotically distributed as a chi-square with appropriate degrees of freedom under the null hypothesis. For the ICP, that is, with the constraints implied by a set of observed marginals, or those of a generalized independence hypothesis, the MDI statistic is

$$2nI(\underline{p}_o : \underline{\hat{p}}) = 2n \sum_i \hat{p}_{io} \ln(\frac{p_{io}}{\hat{p}_i}). \tag{3.27}$$

which is also asymptotically distributed as chi-square with appropriate degrees of freedom. In case of ICP, the MDI estimation procedure yields the same value as MLE. The statistics is also equivalent to the Deviance for testing the model under null.

3.4.2 AIC

The Akaike Information Criterion (AIC) judges a model by how close its fitted values tend to be to the true values. The saturated model is the most complex model and gives the best fit to the data. In practice, simple models with less parameters are preferred since their fit smooths the sample data and have simpler interpretations. Thus, the optimal model is the one that tends to have fit to the true values. The criterion function AIC selects the model that minimizes the quantity

-2*maximized log-likelihood + 2*number of parameters in the model.

This penalizes the model for having too many parameters. We can use AIC to aid in variable selection when we have many potential predictors. From a set of candidate models, we pick the one with smallest AIC. In table (0.3), we show the AIC values for some unsaturated models fitted to the data in Example 2.1.

3.4.3 BIC

An alternative criterion function Bayesian Information Criterion (BIC) penalizes more severely for the number of parameters present in the model. The BIC is defined as

-2*maximized log-likelihood + log(n)*number of parameters in the model.

It is derived based on a Bayesian argument for determining which of a set of models has highest posterior probability.

3.4.4 SAC

We can always create a $\mathbf{D}$ matrix with the design matrix $\mathbf{X}_1$ of any unsaturated model by using any one of the ways discussed in section 6.2. We create a $\mathbf{D}$ matrix for any unsaturated model using the following

$$\mathbf{D} = \mathbf{X}_2 - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2 = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2.$$

We can create the $\mathbf{D}$ matrices for all possible unsaturated models under consideration. For each $\mathbf{D}$ under the extended complete models, using the sample proportions $\underline{p}_0$ from the generated sample, we can calculate $\mathbf{D}\hat{\lambda}_2$ (Section 5.3.1) by

$$\mathbf{D}\hat{\lambda}_2 = (\mathbf{I}_8 - (\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)\mathbf{X}_1'))\ln \underline{p}_0.$$

$\mathbf{D}\hat{\lambda}_2$ is a $N \times 1$ vector. We propose a criterion function based on $\mathbf{D}\hat{\lambda}_2$. If the i-th unsaturated model is the best fit to the data, we would expect $\mathbf{D}\hat{\lambda}_2$ to be very close to $\underline{0}$. We define a measure by adding all the absolute values in $\mathbf{D}\hat{\lambda}_2$. We call it Sum of Absolute Components of $\mathbf{D}\hat{\lambda}_2$ (SAC($\mathbf{D}\hat{\lambda}_2$)). We calculate the SAC($\mathbf{D}\hat{\lambda}_2$) values (section 5.4) for

different unsaturated models and choose the best model with minimum SAC($\mathbf{D}\hat{\lambda}_2$). The following table shows the AIC, BIC, MDI and SAC values for different unsaturated models in Example 2.1.

Table 3.6: AIC, BIC, MDI and SAC values for different unsaturated models

Unsaturated Models	AIC	BIC	MDI	SAC
$\boldsymbol{\lambda_{123} = 0}$	**99.14**	**99.69**	**85.14**	**0.09**[1]
$\lambda_{23}, \lambda_{123} = 0$	878.96	879.43	866.96	1.50
$\lambda_{13}, \lambda_{123} = 0$	830.53	831.01	818.53	1.44
$\boldsymbol{\lambda_{12}, \lambda_{123} = 0}$	**111.70**	**112.18**	**99.70**	**0.09**[2]
$\lambda_{13}, \lambda_{23}, \lambda_{123} = 0$	1596.57	1596.96	1586.57	1.51
$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$	877.73	878.13	867.73	1.50
$\boldsymbol{\lambda_{12}, \lambda_{13}, \lambda_{123} = 0}$	**829.31**	**829.71**	**819.31**	**1.47**

The criterion function values are rounded to two decimal places. According to the AIC, BIC and MDI statistic values the best fitted model is $\lambda_{123} = 0$, if we consider all unsaturated models listed in Table (0.3). The second best fitted model is the model with $\lambda_{12}, \lambda_{123} = 0$, which implies conditional independence between X_1 and X_2 given X_3. These values are closer to the corresponding values of the best fitted model compared to other saturated models. Our proposed criterion function, SAC, also suggests the fact that the best fitted model is $\lambda_{123} = 0$, if rounded to sixteen decimal places. If we round to two decimal places, the SAC values are same for both the models $\lambda_{123} = 0$ and $\lambda_{12}, \lambda_{123} = 0$. Although the unsaturated model satisfying $\lambda_{12} = \lambda_{123} = 0$ giving the

Table 3.7: SAC values rounded to 16 decimal places

Unsaturated Models	SAC
$\lambda_{123} = 0$	0.0900610778333877
$\lambda_{12}, \lambda_{123} = 0$	0.0900610778333870

smallest value of SAC in Table 3.5, this value is very close to the SAC value for the unsaturated model $\lambda_{123} = 0$. The values of AIC, BIC and MDI are also close to each other for these two unsaturated models but to a greater extent.

Chapter 4

Characterization of Relations between Two Models : Saturated 1 and Saturated 2

4.1 Summary

In previous chapters, we extended an unsaturated loglinear model to a complete model in a $2 \times 2 \times 2$ contingency table and investigated the validity of an unsaturated model. In this chapter, we want to characterize the extension of the unsaturated model to the saturated model. In section 4.2, we define the saturated model, an unsaturated model and its extension to a complete model. We name these saturated models as "Saturated 1" and "Saturated Representation 2". We investigate the relation between the "Saturated 1" and "Saturated Representation 2" in section 4.3. In section 4.4, we investigate the estimates of the unknown parameters from each saturated representation

of the unsaturated models. We demonstrate some conditions on the design matrices when some parameter estimates are equal in these two representations.

4.2 Models

In this section we introduce the two saturated representations from an unsaturated model. We can model p_i since the total sample size is fixed. We will use the notations $\underline{p}$ for cell probabilities in a contingency table, $\mathbf{X}_1$ for the design matrix of an unsaturated model, $\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2]$ for the design matrix of "Saturated 1" model, $\mathbf{X}^* = [\mathbf{X}_1|\mathbf{D}_0]$ for the design matrix of "Saturated 2" model, and $\underline{\lambda}_1 = (\lambda_1, \lambda_2, ..., \lambda_{s_1})'$, $\underline{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)' = (\underline{\lambda}_1', \underline{\lambda}_2')$, $\underline{\lambda}^* = (\underline{\lambda}_1', \underline{\lambda}_2^{*\prime})$ for the unknown parameters of the models, respectively, $s_1 < N$.

4.2.1 Unsaturated Model

An unsaturated loglinear model in a contingency table with cell probabilities $\underline{p}$ is described as

$$\ln \underline{p} = \mathbf{X}_1 \underline{\lambda}_1, \tag{4.1}$$

where $\underline{\lambda}_1 = (\lambda_1, \lambda_2, ..., \lambda s_1)'$ are the unknown parameters, the matrix $\mathbf{X}_1$ with dimension $N \times s_1$ is the design matrix of the unsaturated model and $rank(\mathbf{X}_1) = s_1 < N$.

4.2.2 Saturated 1

We consider the saturated model for a contingency table as

$$\ln \underline{p} = \mathbf{X}\underline{\lambda} = \mathbf{X}_1\underline{\lambda}_1 + \mathbf{X}_2\underline{\lambda}_2 = [\mathbf{X}_1 \mid \mathbf{X}_2] \begin{pmatrix} \underline{\lambda}_1 \\ \\ \underline{\lambda}_2 \end{pmatrix}, \tag{4.2}$$

where $\ln \underline{p} = (\ln p_1, \ln p_2, ..., \ln p_N)^{'}$, $\underline{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)^{'} = (\underline{\lambda}_1^{'}, \underline{\lambda}_2^{'})$ are the unknown parameters, $\mathbf{X}$ is the design matrix of the saturated model with dimension $N \times N$, rank($\mathbf{X}$)=N, rank($\mathbf{X}_1$) = $s_1 < N$, rank($\mathbf{X}_2$) = $N - s_1 = s_2$, say and $\mathbf{X}_2^{'}\mathbf{X}_1 \neq \mathbf{0}$. The i-th row of (4.2) gives

$$\ln p_i = \sum_{j=1}^{N} x_{ij}\lambda_j. \tag{4.3}$$

4.2.3 Saturated 2

We extend an unsaturated model to a complete model by constructing a $\mathbf{D}_0$ matrix with dimension $N \times (N - s_1)$ in the following way

$$\ln \underline{p} = \mathbf{X}_1\underline{\lambda}_1 + \mathbf{D}_0\underline{\lambda}_2^* = [\mathbf{X}_1 \mid \mathbf{D}_0]\begin{pmatrix} \underline{\lambda}_1 \\ \\ \underline{\lambda}_2^* \end{pmatrix}, \tag{4.4}$$

where $rank(\mathbf{D}_0) = N - rank(X_1) = N - s_1 = s_2$, say, $rank(\mathbf{X}_1) = s_1$ and $\mathbf{X}_1^{'}\mathbf{D}_0 = \mathbf{0}$. The i-th row of (4.4) gives

$$\ln p_i = \sum_{j=1}^{s_1+N-s_1} x_{ij}^*\lambda_j^*. \tag{4.5}$$

In the above two saturated representations (4.2 and 4.4), we have considered two different design matrices with full rank as

$$\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2] \text{ and } \mathbf{X}^* = [\mathbf{X}_1|\mathbf{D}_0]\,, \tag{4.6}$$

where $\mathbf{X}_2^{'}\mathbf{X}_1 \neq \mathbf{0}$ and $\mathbf{D}_0^{'}\mathbf{X}_1 = \mathbf{0}$. If we pre-multiply both sides of (4.4) by $\mathbf{D}_0^{'}$, we get, by using $\mathbf{D}_0^{'}\mathbf{X}_1 = \mathbf{0}$,

$$\mathbf{D}_0^{'} \ln \underline{p} = 0 + \mathbf{D}_0^{'}\mathbf{D}_0\underline{\lambda}_2^*, \tag{4.7}$$

and if we pre-multiply both sides of (4.4) by $\mathbf{X}_1^{'}$, we get, by using $\mathbf{X}_1^{'}\mathbf{D}_0 = \mathbf{0}$,

$$\mathbf{X}_1^{'} \ln \underline{p} = \mathbf{X}_1^{'}\mathbf{X}_1\underline{\lambda}_1 + 0. \tag{4.8}$$

4.3 Saturated 1 versus Saturated 2

In the previous section, we have defined an unsaturated model and its extension to a complete model. Both the saturated models (4.2 and 4.4) represent the same set of cell probabilities. In this section we establish the relations between saturated model and extended complete model, i.e., among $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{D}_0$. In subsection 4.3.1, we create a $\mathbf{D}_0$ using $\mathbf{X}_1$ and $\mathbf{X}_2$ and in subsection 4.3.2, we present the "Saturated 1" (S1) from a "Saturated 2" (S2).

4.3.1 Relation Between X and $\mathbf{D}_0$

We consider two full rank matrices $[\mathbf{X}_1 \mid \mathbf{X}_2]$ and $[\mathbf{X}_1 \mid \mathbf{D}_0]$ where $\mathbf{X}_2^{'}\mathbf{X}_1 \neq \mathbf{0}$ and $\mathbf{D}_0^{'}\mathbf{X}_1 = \mathbf{0}$ and express $\mathbf{D}_0$ from $\mathbf{X}_1$ and $\mathbf{X}_2$ using. **Lemma** 1 shows that there always exists a matrix $\mathbf{Q}_1$ such that $\mathbf{D}_0 = [\mathbf{X}_1 \mid \mathbf{X}_2]\, \mathbf{Q}_1$.

Lemma 1 *Let* $[\boldsymbol{X}_1 \mid \boldsymbol{X}_2]$ *and* $[\boldsymbol{X}_1 \mid \boldsymbol{D}_0]$ *be two matrices. There exists a matrix* $\boldsymbol{Q}_1 = \begin{pmatrix} -(\boldsymbol{X}_1^{'}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^{'}\boldsymbol{X}_2 \\ \boldsymbol{I} \end{pmatrix}$ *of rank* $N - s_1$ *such that* $\boldsymbol{D}_0 = [\boldsymbol{X}_1|\boldsymbol{X}_2]\boldsymbol{Q}_1$ *and* $\boldsymbol{X}_1^{'}\boldsymbol{D}_0 = [\boldsymbol{X}_1^{'}\boldsymbol{X}_1|\boldsymbol{X}_1^{'}\boldsymbol{X}_2]\boldsymbol{Q}_1 = \boldsymbol{0}$.

Proof. For any two given matrices $[\mathbf{X}_1 \mid \mathbf{X}_2]$ and $[\mathbf{X}_1 \mid \mathbf{D}_0]$, we can write

$$[\mathbf{X}_1 \mid \mathbf{X}_2]\, [\mathbf{X}_1 \mid \mathbf{Q}_1] = [\mathbf{X}_1 \mid \mathbf{D}_0], \tag{4.9}$$

where $\text{rank}(\mathbf{Q}_1) = N - s_1$.

From (4.9), we can express $\mathbf{D}_0$ as

$$\mathbf{D}_0 = [\mathbf{X}_1 \mid \mathbf{X}_2]\, \mathbf{Q}_1. \tag{4.10}$$

We can choose $\mathbf{Q}_1$ such that $\mathbf{X}_1^{'}\mathbf{D}_0 = [\mathbf{X}_1^{'}\mathbf{X}_1 \mid \mathbf{X}_1^{'}\mathbf{X}_2]\,\mathbf{Q}_1 = \mathbf{0}$. We define

$$\mathbf{Q}_1 = \begin{pmatrix} -(\mathbf{X}_1^{'}\mathbf{X}_1)^{-1}\mathbf{X}_1^{'}\mathbf{X}_2 \\ \mathbf{I} \end{pmatrix}. \tag{4.11}$$

Using (4.11) in (4.10), we get

$$\mathbf{D}_0 = \mathbf{X}_2 - \mathbf{X}_1(\mathbf{X}_1^{'}\mathbf{X}_1)^{-1}\mathbf{X}_1^{'}\mathbf{X}_2 = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^{'}\mathbf{X}_1)^{-1}\mathbf{X}_1^{'})\mathbf{X}_2. \tag{4.12}$$

Pre-multiplying both sides of (4.12) by $\mathbf{X}_1^{'}$, we get

$$\mathbf{X}_1^{'}\mathbf{D}_0 = \mathbf{X}_1^{'}(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1^{'}\mathbf{X}_1)^{-1}\mathbf{X}_1^{'})\mathbf{X}_2 = \mathbf{0}. \tag{4.13}$$

This completes the proof. ■

Using Lemma 1, we can always create a $\mathbf{D}_0$ using a known $\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2]$. In Example 2.1, we consider the design matrix and the unknown parameters of the saturated model in the following way

$$\mathbf{X} = [\mathbf{X}_1 \mid \mathbf{X}_2] = \left[\begin{array}{cccccc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right], \underline{\lambda}_1 = \begin{pmatrix} \lambda \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_{13} \\ \lambda_{23} \end{pmatrix}, \underline{\lambda}_2 = \begin{pmatrix} \lambda_{12} \\ \lambda_{123} \end{pmatrix},$$

where $\mathbf{X}_1$ is the design matrix of the unsaturated model with $\lambda_{12}, \lambda_{123} = 0$ from the class C2, i.e., the design matrix of the unsaturated model when X_1 is conditionally

independent of X_2 given X_3 in a $2 \times 2 \times 2$ contingency table. Using (4.12), we get a $\mathbf{D}_0$ as

$$\mathbf{D}_0 = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2 = \begin{bmatrix} 0.25 & 0.00 \\ 0.25 & 0.25 \\ -0.25 & 0.00 \\ -0.25 & -0.25 \\ -0.25 & 0.00 \\ -0.25 & -0.25 \\ 0.25 & 0.00 \\ 0.25 & 0.25 \end{bmatrix}, \tag{4.14}$$

satisfying $\mathbf{X}_1'\mathbf{D}_0 = \mathbf{0}$.

4.3.2 Non-uniqueness of $\mathbf{D}_0$

For any two matrices $\mathbf{D}_0^{(1)}$ and $\mathbf{D}_0^{(2)}$, satisfying $\mathbf{X}_1'\mathbf{D}_0^{(1)} = \mathbf{0}$ and $\mathbf{X}_1'\mathbf{D}_0^{(2)} = \mathbf{0}$, we can always find a full rank matrix $\mathbf{R}_{s_2 \times s_2}$ (C. R. Rao, 2006) such that

$$\mathbf{D}_0^{(1)}\mathbf{R} = \mathbf{D}_0^{(2)}, \tag{4.15}$$

where rank($\mathbf{R}$)=$N - s_1 = s_2$. For any given $\mathbf{D}_0^{(1)}$, we can write an extended complete model as

$$\ln \underline{p} = \mathbf{X}_1\underline{\lambda}_1^{(1)} + \mathbf{D}_0^{(1)}\underline{\lambda}_2^{(1)}. \tag{4.16}$$

Pre-multiplying both sides of (4.15) first by $\mathbf{D}_0^{(1)'}$ and then by $(\mathbf{D}_0^{(1)'}\mathbf{D}_0^{(1)})^{-1}$, we get

$$\mathbf{R} = (\mathbf{D}_0^{(1)'}\mathbf{D}_0^{(1)})^{-1}\mathbf{D}_0^{(1)'}\mathbf{D}_0^{(2)}, \tag{4.17}$$

with full rank. Using (4.17), we can express $\mathbf{D}_0^{(1)}\underline{\lambda}_2^{(1)}$ as

$$\mathbf{D}_0^{(1)}\underline{\lambda}_2^{(1)} = \mathbf{D}_0^{(1)}\mathbf{R}\mathbf{R}^{-1}\underline{\lambda}_2^{(1)} = \mathbf{D}_0^{(2)}\underline{\lambda}_2^{(2)}, \tag{4.18}$$

where $\mathbf{D}_0^{(2)} = \mathbf{D}_0^{(1)}\mathbf{R}$ and $\underline{\lambda}_2^{(2)} = \mathbf{R}^{-1}\underline{\lambda}_2^{(1)}$. If we consider $\mathbf{D}_0^{(2)}$ as in (4.12), constructed from $\mathbf{X}_1$ and $\mathbf{X}_2$, using (4.18) from (4.16), we can write

$$\ln \underline{p} = \mathbf{X}_1\underline{\lambda}_1^{(1)} + (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2\underline{\lambda}_2^{(2)}. \tag{4.19}$$

We define

$$\underline{\lambda}_1^{(2)} = (\underline{\lambda}_1^{(1)} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\underline{\lambda}_2^{(2)}). \tag{4.20}$$

Using (4.20), from (4.19) we can write $\ln \underline{p}$ as

$$\ln \underline{p} = \mathbf{X}_1\underline{\lambda}_1^{(2)} + \mathbf{X}_2\underline{\lambda}_2^{(2)}. \tag{4.21}$$

In Example 2.1, we consider the unsaturated model $\lambda_{12}, \lambda_{123} = 0$ from the class C2 with the design matrix $\mathbf{X}_1$ and a computer generated $\mathbf{D}_0^{(1)}$ as

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}; \ \mathbf{D}_0^{(1)} = \begin{bmatrix} 0.49 & -0.10 \\ 0.10 & 0.49 \\ -0.49 & 0.10 \\ -0.10 & -0.49 \\ -0.49 & 0.10 \\ -0.10 & -0.49 \\ 0.49 & -0.10 \\ 0.10 & 0.49 \end{bmatrix}. \tag{4.22}$$

If we consider a $\mathbf{D}_0^{(2)}$ in (4.14), we get $\mathbf{R} = \begin{bmatrix} 0.5886 & 0.0983 \\ 0.3919 & 0.4902 \end{bmatrix}$.

From a computer generated $\mathbf{D}_0^{(1)}$ in (4.22) and using $\mathbf{X}_1$, we can express $\ln \underline{p}$ as

$$\ln \underline{p} = \begin{pmatrix} \lambda + 0.49\lambda_{12} - 0.10\lambda_{123} \\ \lambda + \lambda_3 + 0.10\lambda_{12} + 0.49\lambda_{123} \\ \lambda + \lambda_2 - 0.49\lambda_{12} + 0.10\lambda_{123} \\ \lambda + \lambda_2 + \lambda_3 + \lambda_{23} - 0.10\lambda_{12} - 0.49\lambda_{123} \\ \lambda + \lambda_1 - 0.49\lambda_{12} + 0.10\lambda_{123} \\ \lambda + \lambda_1 + \lambda_3 + \lambda_{13} - 0.10\lambda_{12} - 0.49\lambda_{123} \\ \lambda + \lambda_1 + \lambda_2 + 0.49\lambda_{12} - 0.10\lambda_{123} \\ \lambda + \lambda_1 + \lambda_2 + \lambda_3 + \lambda_{13} + \lambda_{23} + 0.10\lambda_{12} + 0.49\lambda_{123} \end{pmatrix}. \tag{4.23}$$

We consider a new $\underline{\lambda}$ as $\underline{\lambda}^{(2)} = \begin{pmatrix} \lambda_1^{(2)} \\ \lambda_2^{(2)} \end{pmatrix}$, where $\lambda_1^{(2)}$, $\lambda_2^{(2)}$ are, from (4.18) and (4.20),

$$\underline{\lambda}_1^{(2)} = (\underline{\lambda}_1^{(1)} - (X_1'X_1)^{-1}X_1')X_2\underline{\lambda}_2^{(2)}) = \begin{pmatrix} \lambda + 0.49\lambda_{12} - 0.10\lambda_{123} \\ \lambda_1 - 0.98\lambda_{12} + 0.20\lambda_{123} \\ \lambda_2 - 0.98\lambda_{12} + 0.20\lambda_{123} \\ \lambda_3 - 0.39\lambda_{12} + 0.59\lambda_{123} \\ \lambda_{13} + 0.78\lambda_{12} - 1.18\lambda_{123} \\ \lambda_{23} + 0.78\lambda_{12} - 1.18\lambda_{123} \end{pmatrix},$$

$$\underline{\lambda}_2^{(2)} = \mathbf{R}^{-1}\underline{\lambda}_2^{(1)} = \begin{pmatrix} 1.96\lambda_{12} - 0.39\lambda_{123} \\ -1.57\lambda_{12} + 2.35\lambda_{123} \end{pmatrix}.$$

Using this new $\underline{\lambda}^{(2)}$, $\ln \underline{p}$ can be expressed as

$$\ln \underline{p} = \left[\begin{array}{cccccc|cc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right] \left(\begin{array}{c} \lambda + 0.49\lambda_{12} - 0.10\lambda_{123} \\ \lambda_1 - 0.98\lambda_{12} + 0.20\lambda_{123} \\ \lambda_2 - 0.98\lambda_{12} + 0.20\lambda_{123} \\ \lambda_3 - 0.39\lambda_{12} + 0.59\lambda_{123} \\ \lambda_{13} + 0.78\lambda_{12} - 1.18\lambda_{123} \\ \lambda_{23} + 0.78\lambda_{12} - 1.18\lambda_{123} \\ 1.96\lambda_{12} - 0.39\lambda_{123} \\ -1.57\lambda_{12} + 2.35\lambda_{123} \end{array}\right) = [\mathbf{X}_1|\mathbf{X}_2] \left(\begin{array}{c} \lambda_1^{(2)} \\ \lambda_2^{(2)} \end{array}\right). \tag{4.24}$$

4.4 Parameter Estimates for S1 and S2

In this section, we investigate if the estimates of the unknown parameters in these two representations are same or not. If they are not same, we are interested if they are partially same or not. **Lemma** 2 shows a necessary and sufficient condition when some row(s) of the inverses of two matrices of the form $\mathbf{X}$ and $\mathbf{X}^*$ are same. **Lemma** 3 and **Theorem** 3 show when the estimates of $\underline{\lambda}_2$ and $\underline{\lambda}_2^*$ are same from two different saturated representations. If the estimates of $\underline{\lambda}_2$ and $\underline{\lambda}_2^*$ are same then the estimates of $\underline{\lambda}_1$ are not same from both the models and is proved in **Corollary** 4.

We define the matrices $\mathbf{A}_{s_1\times s_1}$ with rank ($\mathbf{A}$) = s_1, $\mathbf{B}_{s_2\times s_1}$, $\mathbf{E}_{s_1\times s_2}$, $\mathbf{C}_{s_2\times s_2}$ with rank ($\mathbf{C}$) = s_2, $\mathbf{D}^{01}_{s_1\times s_2}$ and $\mathbf{D}^{02}_{s_2\times s_2}$ with rank ($\mathbf{D}^{02}$) = s_2 such that we can rewrite the design matrices

in (4.2) and (4.4) as

$$\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2] = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{E} \\ \hline \mathbf{B} & \mathbf{C} \end{array}\right] \text{ and } \mathbf{X}^* = [\mathbf{X}_1|\mathbf{D}_0] = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{D}^{01} \\ \hline \mathbf{B} & \mathbf{D}^{02} \end{array}\right], \tag{4.25}$$

respectively. We can always partition the design matrices in the above way for any unsaturated model from the classes C1, C2,...C(N-m-1). In order to achieve this partition of $\mathbf{X}$ and $\mathbf{X}^*$, we may need to interchange some rows in the design matrices (Section 5.4) , i.e., to change the order of the cells $1, 2, ..., N$.

Lemma 2 *Let* $\left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{E} \\ \hline \boldsymbol{B} & \boldsymbol{C} \end{array}\right]$ *and* $\left[\begin{array}{c|c} \boldsymbol{A} & \boldsymbol{D}^{01} \\ \hline \boldsymbol{B} & \boldsymbol{D}^{02} \end{array}\right]$ *be two matrices with rank(**A**) =* s_1 *and rank (**C**) = rank($\boldsymbol{D}^{02}$) =* s_2, $s_1 + s_2 = N$. *The following conditions are equivalent if **B** is of full row rank, i.e., rank(**B**) =* $s_2 \leq s_1$.

(i) $(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{E})^{-1} = (\boldsymbol{D}^{02} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{D}^{01})^{-1}$,

(ii) $(\boldsymbol{C} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{E})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} = (\boldsymbol{D}^{02} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{D}^{01})^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$.

Proof. $(i) \Longrightarrow (ii)$

If we post-multiply (i) by $\mathbf{B}\mathbf{A}^{-1}$, we get (ii). For this to be true, we don't need the matrix $\mathbf{B}$ to be full row rank. For any matrix $\mathbf{B}$ and $\mathbf{A}$, this is always true.

$(ii) \Longrightarrow (i)$

Let $\mathbf{B}\mathbf{A}^{-1} = \mathbf{B}^*$ with rank($\mathbf{B}^*$) = $s_2 \leq s_1$. If we post-multiply (ii) by $\mathbf{B}^{*'}$, we get

$$(\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{E})^{-1}\mathbf{B}^*\mathbf{B}^{*'} = (\mathbf{D}^{02} - \mathbf{B}\mathbf{A}^{-1}\mathbf{D}^{01})^{-1}\mathbf{B}^*\mathbf{B}^{*'}. \tag{4.26}$$

We can invert the matrix $\mathbf{B}^*\mathbf{B}^{*'}$ since it is a square matrix with full rank. If we post-multiply both sides of (4.26) by the inverse of the matrix $\mathbf{B}^*\mathbf{B}^{*'}$, we get

$$(\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{E})^{-1} = (\mathbf{D}^{02} - \mathbf{B}\mathbf{A}^{-1}\mathbf{D}^{01})^{-1}.$$

This completes the proof. ■

Lemma 3 *If the condition* (i) *in* ***Lemma*** 2 *is true then there always exists a non-singular matrix* $\boldsymbol{P} = \left[\begin{array}{c|c} \boldsymbol{I}_{s_1} & \boldsymbol{0} \\ \hline -\boldsymbol{BA}^{-1} & \boldsymbol{I}_{s_2} \end{array}\right]$ *with* $det(\boldsymbol{P}) = 1$ *such that* $\boldsymbol{PX} - \boldsymbol{PX}^* = \left[\begin{array}{c|c} \boldsymbol{A}^* & \boldsymbol{E}^* \\ \hline \boldsymbol{B}^* & \boldsymbol{0} \end{array}\right]$ *holds.*

Proof. If we pre-multiply both $\mathbf{X}$ and $\mathbf{X}^*$ by $\mathbf{P}$ and consider the difference, we have

$$\mathbf{PX} - \mathbf{PX}^* = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{E} \\ \hline \mathbf{0} & \mathbf{C} - \mathbf{BA}^{-1}\mathbf{E} \end{array}\right] - \left[\begin{array}{c|c} \mathbf{A} & \mathbf{D}^{01} \\ \hline \mathbf{0} & \mathbf{D}^{02} - \mathbf{BA}^{-1}\mathbf{D}^{01} \end{array}\right] \tag{4.27}$$

If the condition (i) in **Lemma** 2 is true then from (4.27), we get

$$\mathbf{PX} - \mathbf{PX}^* = \left[\begin{array}{c|c} \mathbf{A}^* & \mathbf{E}^* \\ \hline \mathbf{B}^* & \mathbf{0} \end{array}\right], \tag{4.28}$$

where $\mathbf{A}^* = \mathbf{0}$, $\mathbf{B}^* = \mathbf{0}$ and $\mathbf{E}^* = \mathbf{E} - \mathbf{D}^{01}$. ■

Theorem 4 *Let* $[\boldsymbol{X}_1|\boldsymbol{X}_2]$ *and* $[\boldsymbol{X}_1|\boldsymbol{D}_0]$ *be two matrices such that* $\boldsymbol{X}_2'\boldsymbol{X}_1 \neq \boldsymbol{0}$ *and* $\boldsymbol{D}_0'\boldsymbol{X}_1 = \boldsymbol{0}$. *If there exist square matrices* $\boldsymbol{A}$ *with rank* s_1, $\boldsymbol{C}$ *with rank* $N - s_1$ $(= s_2, say)$ *and* $\boldsymbol{D}^{02}$ *with rank* $N - s_1$ *such that the condition* $(\boldsymbol{C} - \boldsymbol{BA}^{-1}\boldsymbol{E})^{-1} = (\boldsymbol{D}^{02} - \boldsymbol{BA}^{-1}\boldsymbol{D}^{01})^{-1}$ *holds then the estimates of* $\underset{\sim}{\lambda}_2$ *and* $\underset{\sim}{\lambda}_2^*$ *are same.*

Proof. Let the condition $(\mathbf{C} - \mathbf{BA}^{-1}\mathbf{E})^{-1} = (\mathbf{D}^{02} - \mathbf{BA}^{-1}\mathbf{D}^{01})^{-1}$ holds for any representation of $\mathbf{X}$ and $\mathbf{X}^*$ in (4.25). We also consider $\ln \underset{\sim}{p}$ as $\ln \underset{\sim}{p} = \left(\begin{array}{c} \ln \underset{\sim}{p}^1_{s_1 \times 1} \\ \hline \ln \underset{\sim}{p}^2_{s_2 \times 1} \end{array}\right)$. Both the design matrices are from saturated models. With the observed data, we can consider $\ln \underset{\sim}{p}_0$ as an estimate of $\ln \underset{\sim}{p}$. Since both the matrices $\mathbf{X}$ and $\mathbf{X}^*$ are of full rank, we can invert

them and get the estimates of $\underline{\lambda}$ and $\underline{\lambda}^*$ using

$$\left(\begin{array}{c}\underline{\lambda}_1\\ \hline \underline{\lambda}_2\end{array}\right) = \left[\begin{array}{c|c}\mathbf{A} & \mathbf{E}\\ \hline \mathbf{B} & \mathbf{C}\end{array}\right]^{-1} \ln \underline{p}_0 = \left[\begin{array}{c|c}\cdots & \cdots\\ \hline -(\mathbf{C}-\mathbf{B}\mathbf{A}^{-1}\mathbf{E})^{-1}\mathbf{B}\mathbf{A}^{-1} & (\mathbf{C}-\mathbf{B}\mathbf{A}^{-1}\mathbf{E})^{-1}\end{array}\right] \ln \underline{p}_0 \quad \text{and}$$

$$\left(\begin{array}{c}\underline{\lambda}_1\\ \hline \underline{\lambda}_2^*\end{array}\right) = \left[\begin{array}{c|c}\mathbf{A} & \mathbf{D}^{01}\\ \hline \mathbf{B} & \mathbf{D}^{02}\end{array}\right]^{-1} \ln \underline{p}_0 = \left[\begin{array}{c|c}\cdots & \cdots\\ \hline -(\mathbf{D}^{02}-\mathbf{B}\mathbf{A}^{-1}\mathbf{D}^{01})^{-1}\mathbf{B}\mathbf{A}^{-1} & (\mathbf{D}^{02}-\mathbf{B}\mathbf{A}^{-1}\mathbf{D}^{01})^{-1}\end{array}\right] \ln \underline{p}_0. \tag{4.29}$$

Using **Lemma** 2, we can say that if $(\mathbf{C}-\mathbf{B}\mathbf{A}^{-1}\mathbf{E})^{-1} = (\mathbf{D}^{02}-\mathbf{B}\mathbf{A}^{-1}\mathbf{D}^{01})^{-1}$ is true then the last s_2 rows of both the inverses $\mathbf{X}$ and $\mathbf{X}^*$ are same and hence the estimates of $\underline{\lambda}_2$ and $\underline{\lambda}_2^*$ are same. This completes the proof. ■

Corollary 5 *If the estimates of $\underline{\lambda}_2$ and $\underline{\lambda}_2^*$ are same then the estimates of $\underline{\lambda}_1$ from saturated models are not same.*

Proof. From **Lemma** 1, we can always say that there exists a full rank matrix $\mathbf{Q}_{N\times N} = \left[\begin{array}{c|c}\mathbf{I}_{s_1} & -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\\ \hline \mathbf{0} & \mathbf{I}_{s_2}\end{array}\right]$ such that $[\mathbf{X}_1 \mid \mathbf{X}_2]\,\mathbf{Q} = [\mathbf{X}_1 \mid \mathbf{D}_0]$, where $\mathbf{D}_0'\mathbf{X}_1 = \mathbf{0}$.
Using the above and considering $\mathbf{D}_0$ as $(\mathbf{I}-\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2$, we can write $\ln \underline{p}$ as

$$\ln \underline{p} = \mathbf{X}_1\underline{\lambda}_1^{(1)} + (\mathbf{I}-\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2\underline{\lambda}_2^* = \mathbf{X}_1(\underline{\lambda}_1^{(1)} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\underline{\lambda}_2^*) + \mathbf{X}_2\underline{\lambda}_2^*. \tag{4.30}$$

From the saturated representation 1, $\ln \underline{p}$ can be expressed as

$$\ln \underline{p} = \mathbf{X}_1\underline{\lambda}_1^{(2)} + \mathbf{X}_2\underline{\lambda}_2^*. \tag{4.31}$$

Comparing (4.30) and (4.31), we have

$$\underline{\lambda}_1^{(2)} = \underline{\lambda}_1^{(1)} - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\underline{\lambda}_2^*.$$

■ This completes the proof.

The lemmas and the theorems always hold for the matrices

$$\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2] = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{E} \\ \hline \mathbf{B} & \mathbf{C} \end{array}\right] \text{ and } \mathbf{X}^* = [\mathbf{X}_1|\mathbf{D}_0] = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{D}^{01} \\ \hline \mathbf{B} & \mathbf{D}^{02} \end{array}\right], \tag{4.32}$$

where rank($\mathbf{A}$)= s_1, rank($\mathbf{C}$) = rank($\mathbf{D}_{02}$)= s_2. In the next subsections, we will show that we can always find the matrices of the form $\mathbf{X}$ and $\mathbf{X}^*$ in (4.32) for any unsaturated model under consideration. For example 2.1, since we are not considering the unsaturated models without the general effect λ and the main effects $\lambda_j, \forall j = 1, 2, \ldots, 3$, we are not changing the first 4 columns of matrix X. We are permuting the next 4 columns to construct X_2 for each unsaturated model. To achieve the rank conditions, permuting the columns is not enough. We need to permute the rows corresponding to the cells of the X matrix. All possible unsaturated models are grouped into two categories: highest order interaction is present in the unsaturated model (Section 4.4.1) and highest order interaction is not present in the unsaturated model (Section 4.4.2). We consider the $2 \times 2 \times 2$ contingency table described in Example 2.1. In the unsaturated models where the highest order interaction λ_{123} is not present the matrix $\mathbf{E}$ is always a null matrix and the partition is unique. If we decide the unsaturated model, i.e., fix the columns of X_2, we can always find this partition by creating $\mathbf{E} = \mathbf{0}$. In the unsaturated models where the highest order interaction λ_{123} is present the matrix $\mathbf{E} \neq \underline{0}$ since there is a row with all 1's (λ_{123} is present). To satisfy the rank conditions, we need to permute the rows of the design matrix and this partition is also unique.

Chapter 5

Orthogonal Extension of Unsaturated Models using D Matrices

5.1 Summary

In this chapter we discuss more on the extended complete model. We will be using the notation $\mathbf{D}$ instead of $\mathbf{D}_0$. In section 5.2, we present different choices of $\mathbf{D}$ matrices for an unsaturated model. We discuss some properties of $\mathbf{D}$ matrix and $\mathbf{D}\underline{\lambda}_2$ vector in section 5.3. We propose a criterion function based on $\mathbf{D}\underline{\lambda}_2$ for selection of best fitted models in section 5.4.

5.2 Different choices of D matrices for an unsaturated model

In this section we consider different choices of **D** matrices for an unsaturated model. For simplicity, we consider the auto-accidents example. In this example, we have 8(= N) cells and we try to fit different unsaturated models with the design matrix $\mathbf{X}_1$ with rank $(\mathbf{X}_1) = s_1 < N$. We can extend these unsaturated models to complete models by creating a matrix **D** with rank $N - s_1$ such that $\mathbf{D}'\mathbf{X}_1 = \mathbf{0}$ in the following ways.

1. Take any $N - s_1$ independent columns from the matrix $((\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')$ and create a **D**.

2. $\mathbf{D} = ((\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2$ (Lemma 1).

3. A computer generated **D** matrix such that $\mathbf{D}'\mathbf{X}_1 = \mathbf{0}$.

5.3 Properties of D

In this section, we discuss some properties of the matrix **D** (or the vector $\mathbf{D}\underline{\lambda}_2$). Theorem 6 shows the uniqueness property of $\mathbf{D}\underline{\lambda}_2$. The sum of all elements in $\mathbf{D}\underline{\lambda}_2$ is always zero and this is proved in Theorem 7.

5.3.1 Uniqueness property

For a given unsaturated model, i.e., for a given $\mathbf{X}_1$ design matrix, we can find different **D** matrices using the methods discussed in Section 5.2. In all different choices of **D**, the estimated $\mathbf{D}\underline{\lambda}_2$ is always the same. The proof is shown in the next theorem.

Theorem 6 *For any unsaturated model, $\boldsymbol{D}\underline{\hat{\lambda}}_2$ is unique.*

Proof. From any unsaturated model in (4.1), we can construct a saturated model given in (4.4). Using (4.4), we have

$$\mathbf{D}\underline{\hat{\lambda}}_2 = \ln \underline{p}_0 - \mathbf{X}_1\underline{\hat{\lambda}}_1. \tag{5.1}$$

From (4.7), the estimate of $\underline{\hat{\lambda}}_1$ does not depend on $\mathbf{D}$ and hence (5.1) can be written as

$$\mathbf{D}\underline{\hat{\lambda}}_2 = \ln \underline{p}_0 - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1' \ln \underline{p}_0. \tag{5.2}$$

The RHS of (5.2) is independent of $\mathbf{D}$. This completes the proof. ■

5.3.2 Sum of all elements is zero

We investigate the elements of $\mathbf{D}\underline{\lambda}_2$ closely. We propose a criterion function for model selection based on $\mathbf{D}$ in Section 5.4.

Theorem 7 *Sum of all elements in $\boldsymbol{D}\underline{\lambda}_2$ is zero.*

Proof. The design matrix of an unsaturated loglinear model can be written as

$$\mathbf{X}_1 = [\underline{j} \mid \mathbf{X}_1^1].$$

We create a matrix $\mathbf{D}$ such that

$$\mathbf{X}_1'\mathbf{D} = \mathbf{0}. \tag{5.3}$$

Let $\underline{d}_k$ be the k-th column of the $\mathbf{D}$ matrix, $k = 1, 2, ..., N - s_1$. From the first column of $\mathbf{X}_1$ and using (5.3), we get

$$\underline{j}'\underline{d}_k = 0, \ \forall\, k = 1, 2, ..., N - s_1. \tag{5.4}$$

From (5.4), we have

$$\sum_{i=1}^{N} d_{ik} = 0, \ \forall \ k = 1, 2, ..., N - s_1. \tag{5.5}$$

If $\underline{\lambda}_2 = (\lambda_{21}, \lambda_{22}, ..., \lambda_{2(N-s_1)})'$, the i-th element of $\mathbf{D}\underline{\lambda}_2$ is $\sum_{k=1}^{N-s_1} d_{ik}\lambda_{2k}$.

The sum of all elements in $\mathbf{D}\underline{\lambda}_2$ is

$$\text{Sum}\,(\mathbf{D}\underline{\lambda}_2) = \sum_{k=1}^{N-s_1} \lambda_{2k} \sum_{i=1}^{N} d_{ik} \tag{5.6}$$

Using (5.5), from (5.6), we have

$$\text{Sum}\,(\mathbf{D}\underline{\lambda}_2) = \sum_{k=1}^{N-s_1} \lambda_{2k}(0) = 0.$$

This completes the proof. ■

5.3.3 "Element Zero" property of D matrix

We are only interested in the unsaturated models where one or more interactions are absent. We divide the unsaturated models into three groups $k = 1, 2$ and 3, where k is the number of interactions added to the fitted model. We create $\mathbf{D}$ matrices for each of the unsaturated models in one of the ways described in section 5.2. We see some symmetries in the elements of $\mathbf{D}\hat{\underline{\lambda}}_2$. We have already shown that the sum of all the elements in $\mathbf{D}\hat{\underline{\lambda}}_2$ is zero. We also observe that some elements are zero and/or some values are positive and some values are negative with the same magnitude. In this section we investigate the "Element Zero" property of $\mathbf{D}\hat{\underline{\lambda}}_2$. The following theorem shows a necessary and sufficient condition for a diagonal element to be zero in a symmetric and idempotent matrix. This condition helps to explain "Element Zero" property of $\mathbf{D}\hat{\underline{\lambda}}_2$.

Theorem 8 *For a symmetric, idempotent matrix $\boldsymbol{A}_{n\times n}$, a diagonal element is zero if and only if all the elements in the row and column representing its position are zero.*

Proof. Let i-th row/column of a symmetric, idempotent matrix $\mathbf{A}$ be

$$a_i^{'} = (a_{i1}, a_{i2}, ..., a_{in}) = (0, 0, ..., 0). \tag{5.7}$$

The matrix $\mathbf{A}$ is symmetric and idempotent, i.e.,

$$a_{ij} = a_{ji},\ \forall i \neq j \text{ and}$$

$$\mathbf{A}^2 = \mathbf{A}. \tag{5.8}$$

Using (5.7-5.8) we get the (i, i)-th diagonal element of the matrix $\mathbf{A}$ as

$$a_{ii} = \underline{a}_i^{'}\underline{a} = 0.$$

Let the (i, i)-th diagonal element of a symmetric, idempotent matrix $\mathbf{A}$ be zero. From (5.8), we can write

$$a_{ii} = \underline{a}_i^{'}\underline{a} = \sum_{j=1}^{n} a_{ij}^2 = 0. \tag{5.9}$$

If (5.9) is true then we must have $a_{ij} = 0,\ \forall j$. This completes the proof. ■

Corollary 9 *For a symmetric, idempotent matrix* $\boldsymbol{D} = (\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1^{'}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^{'})$*, an element in* $\boldsymbol{D}\underline{\hat{\lambda}}_2$ *is zero if and only if all the elements in the row and column representing its position in* $\boldsymbol{D}$ *are zero.*

5.4 A proposed criterion function SAC($\mathbf{D}\hat{\lambda}_2$)

We now present a new criterion function for model selection based on $\mathbf{D}$. We can always create a $\mathbf{D}$ matrix with the design matrix $\mathbf{X}_1$ of any unsaturated model by using any one of the ways discussed in section 6.2. Let us assume that we create a $\mathbf{D}$ matrix

for any unsaturated model using the following

$$\mathbf{D} = \mathbf{X}_2 - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2 = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2.$$

We can create the $\mathbf{D}$ matrices for all possible unsaturated models under consideration. For each $\mathbf{D}$ under the extended complete models, using the sample proportions $\underline{p}_0$ from the generated sample, we can calculate $\mathbf{D}\hat{\underline{\lambda}}_2$ by

$$\mathbf{D}\hat{\underline{\lambda}}_2 = (\mathbf{I}_8 - (\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)\mathbf{X}_1')) \ln \underline{p}_0.$$

$\mathbf{D}\hat{\underline{\lambda}}_2$ is a $N \times 1$ vector. We propose a criterion function based on $\mathbf{D}\hat{\underline{\lambda}}_2$. If the i-th unsaturated model is the best fit to the data, we would expect $\mathbf{D}\hat{\underline{\lambda}}_2$ to be very close to $\underline{0}$. We define a measure by adding all the absolute values in $\mathbf{D}\hat{\underline{\lambda}}_2$. We call it Sum of Absolute Components of $\mathbf{D}\hat{\underline{\lambda}}_2$. We calculate the sum of the absolute values of $\mathbf{D}\hat{\underline{\lambda}}_2$ (SAC($\mathbf{D}\hat{\underline{\lambda}}_2$)) for different unsaturated models and choose the best model with minimum SAC($\mathbf{D}\hat{\underline{\lambda}}_2$). Table 3.6 represent the SAC values of different unsaturated models fitted to the data in Example 2.1.

Chapter 6

Simulation

6.1 Summary

In this chapter, we study the comparison between the proposed criterion function "SAC" and the standard method "-2logL" in finding the best fitted model through simulation. In section (6.2), we discuss the generation of random samples from a multinomial loglinear model. We generate samples from six different multinomial models. In the next two sections (6.3) and (6.4), we discuss the methods of calculating "-2logL" and "SAC", respectively, for the unsaturated models. In the section simulation and model comparisons, the results are shown from all the samples for each multinomial model. In section (6.6), we discuss a property of the proposed criterion function.

6.2 Generation of random samples

In this section, we discuss how we generate multinomial random samples for a given set of $\underline{\lambda}$ values in a contingency table. We consider six different models with six different sets of $\underline{\lambda}$ values. Since we are interested in the unsaturated models where 2 or higher order interactions are absent, we choose the sets of $\underline{\lambda}$ values by varying the interaction parameter values. In the saturated model for a $2 \times 2 \times 2$ contingency table, we have 8 unknown parameters, i.e. $\underline{\lambda} = (\lambda, \lambda_1, \lambda_2, \lambda_3, \lambda_{12}, \lambda_{13}, \lambda_{23}, \lambda_{123})'$ and 7 out of 8 are independent. We consider λ as the dependent parameter. We select 7 numerical values for $(\lambda_1, \lambda_2, \lambda_3, \lambda_{12}, \lambda_{13}, \lambda_{23}, \lambda_{123})'$. From these values, we find λ using the relation

$$\lambda = -\ln\left(1 + e^{\lambda_1} + e^{\lambda_2} + e^{\lambda_3} + e^{\lambda_1+\lambda_2} + e^{\lambda_1+\lambda_3+\lambda_{13}} + e^{\lambda_2+\lambda_3+\lambda_{23}} + e^{\lambda_1+\lambda_2+\lambda_3+\lambda_{13}+\lambda_{23}}\right),$$

satisfying the condition $\sum_w p_w = 1$. Table 6.1 shows the λ parameter values for six different models. With the design matrix X (saturated model) and the λ values from Table 6.1, we get the cell probabilities using the relation

$$p_w = \exp\left[\sum_j x_{wj}\lambda_j\right]. \tag{6.1}$$

Table 6.2 represents the corresponding cell probabilities for each set of $\underline{\lambda}$ values in Table 6.1. In Table 6.2, the last 4 rows under each model represent the values of the interaction parameters, $\lambda_{12}, \lambda_{13}, \lambda_{23}$ and λ_{123}, respectively. We want to generate samples from a loglinear model representing some parameters with small numerical values. We choose the values of λ_{12} and λ_{123} very close to zero compared to other interaction parameter values in models 1 - 6 (Table 6.1). In models 3 and 4, λ_{13} values is also close to zero but higher than λ_{12} and λ_{123} values. In model 5, all the interaction parameter values are close to zero with λ_{12} and λ_{123} being the smallest and the highest, respectively. In

model 6, we choose only one interaction λ_{12} value to be small. We generate $100,000$ multinomial samples for each set of $\underline{p}$ from the table Table 6.2 with fixed size $n = 68694$ in R.

Table 6.1: λ values for six simulated models

Parameters	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
λ	-2.4654	-4.3262	-1.3008	-2.0844	-0.7839	-3.9759
λ_1	-1.6094	0.5000	-1.6094	0.5000	-1.6094	-1.6094
λ_2	-0.9163	-0.9163	-0.9163	-0.9163	-0.9163	-0.9163
λ_3	-1.2040	-1.2040	-1.2040	-1.2040	-1.2040	-1.2040
λ_{12}	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100
λ_{13}	3.2834	3.2834	0.0500	0.0500	0.0150	3.2834
λ_{23}	2.3434	2.3434	2.3434	2.3434	0.0200	2.3434
λ_{123}	0.0300	0.0300	0.0300	0.0300	0.0300	1.9738

6.3 Method I

In this method, we partition the design matrix $\mathbf{X}'_{8\times 8}$ matrix (saturated model) into $\mathbf{X}'_1$ and $\mathbf{X}'_2$, where $\mathbf{X}'_1$ is 8×4 and $\mathbf{X}'_2$ is also 8×4. The columns of $\mathbf{X}'_1$ represent the parameters $\lambda, \lambda_1, \lambda_2$ and λ_3 and the columns of $\mathbf{X}'_2$ represent the parameters $\lambda_{12}, \lambda_{13}, \lambda_{23}$ and λ_{123}. We will take one or more columns from $\mathbf{X}'_2$ and add to $\mathbf{X}'_1$. We denote k as the number of columns added to $\mathbf{X}'_1$ from $\mathbf{X}'_2$. If $k = 4$, we get the saturated model. We only consider the unsaturated models with $k = 1, 2, 3$. In this way, we create design matrices

Table 6.2: Cell Probabilities for the six simulated models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
cell1 (0,0,0):	0.0850	0.0132	0.2803	0.1244	0.4566	0.0188
cell2 (0,0,1):	0.0255	0.0040	0.0841	0.0373	0.1370	0.0056
cell3 (0,1,0):	0.0340	0.0053	0.1121	0.0498	0.1826	0.0075
cell4 (0,1,1):	0.1062	0.0165	0.3504	0.1555	0.0559	0.0235
cell5 (1,0,0):	0.0170	0.0218	0.0561	0.2051	0.0913	0.0038
cell6 (1,0,1):	0.1360	0.1743	0.0177	0.0647	0.0278	0.0300
cell6 (1,1,0):	0.0069	0.0088	0.0227	0.0829	0.0369	0.0015
cell7 (1,1,1):	0.5895	0.7561	0.0767	0.2805	0.0118	0.9094

for different unsaturated models. With these design matrices for $k = 1, 2, 3$, we fit the unsaturated log-liner models to the generated data using "glm" function in R. For all possible unsaturated models, we calculate "$-2\ln(\hat{L})$" and we select the best model with minimum $-2\ln(\hat{L})$ value under different k. We repeat this process to all the rest of the samples from the same model.

6.4 Method II

In this method, we create a $\mathbf{D}$ matrix with the new design matrix $\mathbf{X}_1$ (after adding one or more columns from $\mathbf{X}_2$) and with the new $\mathbf{X}_2$ (after deleting one or more columns from $\mathbf{X}_2$) using the following

$$\mathbf{D} = \mathbf{X}_2 - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2 = (\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2.$$

For each **D** under the extended complete models, using the sample proportions $\underline{p}_0$ from the generated sample, we calculate $\mathbf{D}\hat{\lambda}_2$ by

$$\mathbf{D}\hat{\lambda}_2 = (\mathbf{I}_8 - (\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)\mathbf{X}_1')) \ln \underline{p}_0.$$

We calculate the sum of the absolute values of $\mathbf{D}\hat{\lambda}_2$ (SAC($\mathbf{D}\hat{\lambda}_2$)) for different unsaturated models and choose the best model with minimum SAC($\mathbf{D}\hat{\lambda}_2$).

6.5 Simulation and Model Comparisons

We generate 100,000 samples from a known distribution with $\underline{p}$. We consider the set of unsaturated models where at least one 2^{nd} or higher order interactions are present. For the LLMs describing the dependence structure among the categorical variables in Example 2.1, we have three 2-factor interactions $\lambda_{12}.\lambda_{13}$, λ_{23} and one 3-factor interaction λ_{123}. We define "k" as the number of interaction(s) present in the model. The possible values of "k" are 1,2 and 3 since we are only considering the unsaturated models. If we consider only one interaction present in the model, i.e., k=1, we get $\binom{4}{1} = 4$ unsaturated models. The number of possible unsaturated models when k=2 and k=3 are $\binom{4}{2} = 6$ and $\binom{4}{3} = 4$ respectively. The total number of unsaturated loglinear models is $4 + 6 + 4 = 14$. We fit these 14 unsaturated loglinear models to each sample data. We calculate the values of the criterion functions -2logL, SAVC($D\underline{\hat{\lambda}}_2$), AIC and BIC for all these fitted models under each sample. For a criterion function, we select the best fitted unsaturated models satisfying "the smaller the better" from each sample for different values of k. We find the best fitted model in each group of k=1,2,3 under each sample. This process is repeated for each of the six models described in Tables 6.2 and 6.3. The frequency

distribution of best fitted models are displayed for different values of k under each model. We also present the count of the interaction parameters that are present in the best fitted models for different values of k.

6.5.1 k=1

We compare all four unsaturated models when taking one column from X_2 to X_1. Table (6.3) to (6.8) show the frequency distribution of the best fitted models for all 6 models when only one interaction parameter is in the model. The values represent the number of times a specific model has been selected as the best model by each criterion function. The values in Tables (6.9) to (6.14) show the counts where parameters are present in the best model for each interaction parameter. In Model 1, Table (6.1), λ_{12} has the smallest value and λ_{123} has the second smallest value. The criterion function "-2logL" suggests the model "$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$" as the best fitted model every time whereas our proposed criterion function suggests two models: "$\lambda_{12}, \lambda_{13}, \lambda_{23} = 0$" with approximately 63% times and "$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$" with approximately 37% times. This result for Model 1 is suspicious since a higher percentage of best fitted models involve λ_{123} present in the model by both criterion functions. The SAC works better in this case since 37% times it chooses the model where λ_{13} (highest numerical value) is present. In Model 2, $-2logL$ works better than SAC since it selects the model where λ_{13} is present in all samples. In Models 3, 4 and 6, both the criterion functions select the same model all the time. In Models 5, the numbers are distributed in all the 4 possible models for k=1 by both criterion functions. It is needed to mention that the criterion function SAC has a tendency of selecting the best fitted model where λ_{123} (second smallest numerical

value) is present.

Table 6.3: Frequency distribution of best fitted models for k=1, Model 1

Fitted models where	parameters=0	$-2logL$	SAC($D\underline{\hat{\lambda}}$)
λ_{123} is present	$\lambda_{12}, \lambda_{13}, \lambda_{23} = 0$	100,000	63316
λ_{12} is present	$\lambda_{13}, \lambda_{23}, \lambda_{123} = 0$	0	0
λ_{23} is present	$\lambda_{12}, \lambda_{13}, \lambda_{123} = 0$	0	0
λ_{13} is present	$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$	0	36684
Total		100,000	100,000

Table 6.4: Frequency distribution of best fitted models for k=1, Model 2

Fitted models where	parameters=0	$-2logL$	SAC($D\underline{\hat{\lambda}}$)
λ_{123} is present	$\lambda_{12}, \lambda_{13}, \lambda_{23} = 0$	0	70847
λ_{12} is present	$\lambda_{13}, \lambda_{23}, \lambda_{123} = 0$	0	0
λ_{23} is present	$\lambda_{12}, \lambda_{13}, \lambda_{123} = 0$	0	0
λ_{13} is present	$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$	100,000	29153
Total		100,000	100,000

Table 6.5: Frequency distribution of best fitted models for k=1, Model 3

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
λ_{123} is present	$\lambda_{12}, \lambda_{13}, \lambda_{23} = 0$	0	0
λ_{12} is present	$\lambda_{13}, \lambda_{23}, \lambda_{123} = 0$	0	0
λ_{23} is present	$\lambda_{12}, \lambda_{13}, \lambda_{123} = 0$	100,000	100,000
λ_{13} is present	$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$	0	0
Total		100,000	100,000

Table 6.6: Frequency distribution of best fitted models for k=1, Model 4

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
λ_{123} is present	$\lambda_{12}, \lambda_{13}, \lambda_{23} = 0$	0	0
λ_{12} is present	$\lambda_{13}, \lambda_{23}, \lambda_{123} = 0$	0	0
λ_{23} is present	$\lambda_{12}, \lambda_{13}, \lambda_{123} = 0$	100,000	100,000
λ_{13} is present	$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$	0	0
Total		100,000	100,000

Table 6.7: Frequency distribution of best fitted models for k=1, Model 5

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
λ_{123} is present	$\lambda_{12}, \lambda_{13}, \lambda_{23} = 0$	37824	63545
λ_{12} is present	$\lambda_{13}, \lambda_{23}, \lambda_{123} = 0$	14311	8973
λ_{23} is present	$\lambda_{12}, \lambda_{13}, \lambda_{123} = 0$	29266	13806
λ_{13} is present	$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$	18599	13676
Total		100,000	100,000

Table 6.8: Frequency distribution of best fitted models for k=1, Model 6

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
λ_{123} is present	$\lambda_{12}, \lambda_{13}, \lambda_{23} = 0$	100,000	100,000
λ_{12} is present	$\lambda_{13}, \lambda_{23}, \lambda_{123} = 0$	0	0
λ_{23} is present	$\lambda_{12}, \lambda_{13}, \lambda_{123} = 0$	0	0
λ_{13} is present	$\lambda_{12}, \lambda_{23}, \lambda_{123} = 0$	0	0
Total		100,000	100,000

Table 6.9: Number of times parameters are present in the best fitted models for k=1, Model 1

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	0	0	100,000
absent($-2logL$)	100,000	100,000	100,000	0
present(SAC)	0	36684	0	63316
absent(SAC)	100,000	63316	100,000	36684

Table 6.10: Number of times parameters are present in the best fitted models for k=1, Model 2

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	100,000	0	0
absent($-2logL$)	100,000	0	100,000	100,000
present(SAC)	0	29153	0	70847
absent(SAC)	100,000	70847	100,000	29153

Table 6.11: Number of times parameters are present in the best fitted models for k=1, Model 3

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	0	100,000	0
absent($-2logL$)	100,000	100,000	0	100,000
present(SAC)	0	0	100,000	0
absent(SAC)	100,000	100,000	0	100,000

Table 6.12: Number of times parameters are present in the best fitted models for k=1, Model 4

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	0	100,000	0
absent($-2logL$)	100,000	100,000	0	100,000
present(SAC)	0	0	100,000	0
absent(SAC)	100,000	100,000	0	100,000

Table 6.13: Number of times parameters are present in the best fitted models for k=1, Model 5

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	0	100,000	0
absent($-2logL$)	100,000	100,000	0	100,000
present(SAC)	0	0	100,000	0
absent(SAC)	100,000	100,000	0	100,000

Table 6.14: Number of times parameters are present in the best fitted models for k=1, Model 6

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	0	0	100,000
absent($-2logL$)	100,000	100,000	100,000	0
present(SAC)	0	0	0	100,000
absent(SAC)	100,000	100,000	100,000	0

6.5.2 k=2

We compare all six unsaturated models when taking one column from X_2 to X_1. Table (6.15) to (6.20) show the frequency distribution of the best fitted models for all 6 models (Table 6.1) when two of the interaction parameters are present in the model. The values represent the number of times a specific model has been selected as the best model by each criterion function. The values in Tables (6.21) to (6.26) show the counts where parameters are present in the best model for each interaction parameter. When k=2, both the criterion functions selects the true model $\lambda_{12} = \lambda_{123} = 0$ for each sample. SAC is identifying the parameter with the smallest close-to-zero value more frequently than the deviance statistic in 100,000 realizations of the simulated data (Table 6.23). The deviance statistic is identifying the next close-to-zero value more frequently than the SAC (Table 6.10). In Models 3 and 4, these two select the true model more than 50% of the times. In Model 6, the deviance selects the model with λ_{12} present 13 times. But the SAC always selects the model with λ_{12} is not present.

6.5.3 k=3

For k=3, we have four possible unsaturated models. In Models 1 and 2, SAC selects the model $\lambda_{12} = 0$ most of the times than model $\lambda_{123} = 0$ compared to $-2logL$. It suggests that SAC detects the model smallest parameter value more often than $-2logL$. Similarly for Models 3-5, SAC performs better in selecting the model with $\lambda_{12} = 0$. In Models 5 and 6, SAC works better than the deviance in identifying the parameter with the smallest close-to-zero value.

Table 6.15: Frequency distribution of best fitted models for k=2, Model 1

Fitted models where	parameters=0	$-2logL$	SAC($D\underline{\hat{\lambda}}$)
$\lambda_{23}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{13} = 0$	0	0
$\lambda_{12}, \lambda_{123}$ are present	$\lambda_{13}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}$ are present	$\lambda_{23}, \lambda_{123} = 0$	0	0
$\lambda_{13}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{23}$ are present	$\lambda_{13}, \lambda_{123} = 0$	0	0
$\lambda_{13}, \lambda_{23}$ are present	$\lambda_{12}, \lambda_{123} = 0$	100, 000	100, 000
Total		100, 000	100, 000

Table 6.16: Frequency distribution of best fitted models for k=2, Model 2

Fitted models where	parameters=0	$-2logL$	SAC($D\underline{\hat{\lambda}}$)
$\lambda_{23}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{13} = 0$	0	0
$\lambda_{12}, \lambda_{123}$ are present	$\lambda_{13}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}$ are present	$\lambda_{23}, \lambda_{123} = 0$	0	0
$\lambda_{13}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{23}$ are present	$\lambda_{13}, \lambda_{123} = 0$	0	0
$\lambda_{13}, \lambda_{23}$ are present	$\lambda_{12}, \lambda_{123} = 0$	100, 000	100, 000
Total		100, 000	100, 000

Table 6.17: Frequency distribution of best fitted models for k=2, Model 3

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{23}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{13} = 0$	42081	47700
$\lambda_{12}, \lambda_{123}$ are present	$\lambda_{13}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}$ are present	$\lambda_{23}, \lambda_{123} = 0$	0	0
$\lambda_{13}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{23}$ are present	$\lambda_{13}, \lambda_{123} = 0$	2789	1342
$\lambda_{13}, \lambda_{23}$ are present	$\lambda_{12}, \lambda_{123} = 0$	55130	50958
Total		100, 000	100, 000

Table 6.18: Frequency distribution of best fitted models for k=2, Model 4

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{23}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{13} = 0$	41407	46860
$\lambda_{12}, \lambda_{123}$ are present	$\lambda_{13}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}$ are present	$\lambda_{23}, \lambda_{123} = 0$	0	0
$\lambda_{13}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{23}$ are present	$\lambda_{13}, \lambda_{123} = 0$	5777	3770
$\lambda_{13}, \lambda_{23}$ are present	$\lambda_{12}, \lambda_{123} = 0$	52816	49370
Total		100, 000	100, 000

Table 6.19: Frequency distribution of best fitted models for k=2, Model 5

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{23}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{13} = 0$	20836	24214
$\lambda_{12}, \lambda_{123}$ are present	$\lambda_{13}, \lambda_{23} = 0$	15039	22150
$\lambda_{12}, \lambda_{13}$ are present	$\lambda_{23}, \lambda_{123} = 0$	11205	7488
$\lambda_{13}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{23} = 0$	15118	27175
$\lambda_{12}, \lambda_{23}$ are present	$\lambda_{13}, \lambda_{123} = 0$	16763	7822
$\lambda_{13}, \lambda_{23}$ are present	$\lambda_{12}, \lambda_{123} = 0$	21039	11151
Total		100, 000	100, 000

Table 6.20: Frequency distribution of best fitted models for k=2, Model 6

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{23}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{13} = 0$	0	0
$\lambda_{12}, \lambda_{123}$ are present	$\lambda_{13}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}$ are present	$\lambda_{23}, \lambda_{123} = 0$	0	0
$\lambda_{13}, \lambda_{123}$ are present	$\lambda_{12}, \lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{23}$ are present	$\lambda_{13}, \lambda_{123} = 0$	13	0
$\lambda_{13}, \lambda_{23}$ are present	$\lambda_{12}, \lambda_{123} = 0$	99987	100, 000
Total		100, 000	100, 000

Table 6.21: Number of times parameters are present in the best fitted models for k=2, Model 1

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	100,000	100,000	0
absent($-2logL$)	100,000	0	0	100,000
present(SAC)	0	100,000	100,000	0
absent(SAC)	100,000	0	0	100,000

Table 6.22: Number of times parameters are present in the best fitted models for k=2, Model 2

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	100,000	100,000	0
absent($-2logL$)	100,000	0	0	100,000
present(SAC)	0	100,000	100,000	0
absent(SAC)	100,000	0	0	100,000

Table 6.23: Number of times parameters are present in the best fitted models for k=2, Model 3

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	2789	55130	100,000	42081
absent($-2logL$)	97211	44870	0	57919
present(SAC)	1342	50958	100,000	47700
absent(SAC)	98658	49042	0	52300

Table 6.24: Number of times parameters are present in the best fitted models for k=2, Model 4

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	5777	52816	100,000	41407
absent($-2logL$)	94223	47184	0	58593
present(SAC)	3770	49370	100,000	46860
absent(SAC)	96230	50630	0	53140

Table 6.25: Number of times parameters are present in the best fitted models for k=2, Model 5

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	5777	52816	100,000	41407
absent($-2logL$)	94223	47184	0	58593
present(SAC)	3770	49370	100,000	46860
absent(SAC)	96230	50630	0	53140

Table 6.26: Number of times parameters are present in the best fitted models for k=2, Model 6

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	13	99,987	100,000	0
absent($-2logL$)	99,987	13	0	100,000
present(SAC)	0	100,000	100,000	0
absent(SAC)	100,000	0	0	100,000

Table 6.27: Frequency distribution of best fitted models for k=3, Model 1

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{13}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{12} = 0$	51143	62773
$\lambda_{12}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{13} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{123}$ are present	$\lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{23}$ are present	$\lambda_{123} = 0$	48857	37227
Total		100,000	100,000

Table 6.28: Frequency distribution of best fitted models for k=3, Model 2

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{13}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{12} = 0$	55043	58533
$\lambda_{12}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{13} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{123}$ are present	$\lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{23}$ are present	$\lambda_{123} = 0$	44957	41467
Total		100,000	100,000

Table 6.29: Frequency distribution of best fitted models for k=3, Model 3

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{13}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{12} = 0$	46476	55344
$\lambda_{12}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{13} = 0$	12963	12594
$\lambda_{12}, \lambda_{13}, \lambda_{123}$ are present	$\lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{23}$ are present	$\lambda_{123} = 0$	40561	32062
Total		100,000	100,000

Table 6.30: Frequency distribution of best fitted models for k=3, Model 4

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{13}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{12} = 0$	40095	49677
$\lambda_{12}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{13} = 0$	18855	18633
$\lambda_{12}, \lambda_{13}, \lambda_{123}$ are present	$\lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{23}$ are present	$\lambda_{123} = 0$	41050	31690
Total		100,000	100,000

Table 6.31: Frequency distribution of best fitted models for k=3, Model 5

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{13}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{12} = 0$	27104	32107
$\lambda_{12}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{13} = 0$	25568	26008
$\lambda_{12}, \lambda_{13}, \lambda_{123}$ are present	$\lambda_{23} = 0$	19141	27228
$\lambda_{12}, \lambda_{13}, \lambda_{23}$ are present	$\lambda_{123} = 0$	28187	14657
Total		100,000	100,000

Table 6.32: Frequency distribution of best fitted models for k=3, Model 6

Fitted models where	parameters=0	$-2logL$	SAC($D\hat{\underline{\lambda}}$)
$\lambda_{13}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{12} = 0$	100,000	100,000
$\lambda_{12}, \lambda_{23}, \lambda_{123}$ are present	$\lambda_{13} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{123}$ are present	$\lambda_{23} = 0$	0	0
$\lambda_{12}, \lambda_{13}, \lambda_{23}$ are present	$\lambda_{123} = 0$	0	0
Total		100,000	100,000

Table 6.33: Number of times parameters are present in the best fitted models for k=3, Model 1

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	48857	100,000	100,000	51143
absent($-2logL$)	51143	0	0	48857
present(SAC)	37227	100,000	100,000	62773
absent(SAC)	62773	0	0	37227

Table 6.34: Number of times parameters are present in the best fitted models for k=3, Model 2

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	44957	100,000	100,000	55043
absent($-2logL$)	55043	0	0	44957
present(SAC)	41467	100,000	100,000	58533
absent(SAC)	58533	0	0	41467

Table 6.35: Number of times parameters are present in the best fitted models for k=3, Model 3

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	53524	87037	100,000	53439
absent($-2logL$)	46476	12963	0	40561
present(SAC)	44656	87406	100,000	67938
absent(SAC)	55344	12594	0	32062

Table 6.36: Number of times parameters are present in the best fitted models for k=3, Model 4

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	59905	81145	100,000	58950
absent($-2logL$)	40095	18855	0	41050
present(SAC)	50323	81367	100,000	68310
absent(SAC)	49677	18633	0	31690

6.6 Property of SAC

From all the tables above, we can conclude that "SAC" works better than the criterion function "$-2logL$" in detecting the smallest interaction. In all the models, λ_{12} has the smallest magnitude in comparison to other interaction parameters. When k=1, 3,

Table 6.37: Number of times parameters are present in the best fitted models for k=3, Model 5

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	59905	81145	100,000	58950
absent($-2logL$)	40095	18855	0	41050
present(SAC)	50323	81367	100,000	68310
absent(SAC)	49677	18633	0	31690

Table 6.38: Number of times parameters are present in the best fitted models for k=3, Model 6

	λ_{12}	λ_{13}	λ_{23}	λ_{123}
present($-2logL$)	0	100,000	100,000	100,000
absent($-2logL$)	100,000	0	0	0
present(SAC)	0	100,000	100,000	100,000
absent(SAC)	100,000	0	0	0

both the criterion functions are selecting the model (in most cases) where λ_{23} is present and the other interaction parameters are assumed to be zero. When k=2, SAC as well as $-2logL$ work very well in selecting the best fitted model. If we carefully look at the tables where the counts represent the absence or presence of each interaction parameter, we can say that the SAC works better than $-2logL$ in selecting the best fitted model where the smallest interaction parameter is absent. This shows that our proposed criterion function detects the model with maximum frequency when the parameter, with the smallest value in the original model, is assumed to be zero.

Chapter 7

Conclusion

In this dissertation, we propose a new method of loglinear model building by adding one or more new terms orthogonally to the terms of the existing model. When we add all the new non-existing terms to the existing model this way, we obtain a new saturated model (Saturated 2). We compare this new saturated model with the standard saturated model (Saturated 1). The proposed orthogonal extension provide the conditional independence structure among the categorical variables in the unsaturated model (section 2.4.1.1, p.10). This way of extending the loglinear models allows us to relate the parameters that are present in the model to the parameters that are assumed to be zero in an orthogonal way. We estimated the unknown model parameters by using the maximum likelihood estimation method. We compared the Saturated 1 model with the Saturated 2 model. We also established the necessary and sufficient conditions for the equivalence of the parameter estimates between these two models. We demonstrated the uniqueness and sum to zero properties of this orthogonal extension. Our proposed new criterion function is based on the extensions of the unsaturated models. Our proposed method

is computer intensive and comparable with the methods using the AIC and BIC. SAC is identifying the parameter with the smallest close-to-zero value more frequently than the deviance statistic in 100,000 realizations of the simulated data (Table 6.23). The deviance statistic is identifying the next close-to-zero value more frequently than the SAC (Table 6.10). When number of non-zero parameters is less than 3, SAC is performing overall better than the deviance statistic in correctly identifying the smallest value close-to-zero. The performance of SAC is more balanced in frequency for identifying the close-to-zero values in 100,000 realizations of the simulated data.

Bibliography

[1] Agresti, A. 2013. *Categorical Data Analysis*. Wiley, Third Edition.

[2] Aitchison, J., and Silvey, S.D. 1958. Maximum-likelihood estimation of parameters subject to restraints. *The Annals of Mathematical Statistics*.

[3] Aitchison, J., and Silvey, S.D. 1960. Maximum-likelihood estimation procedures and associated tests of significance. *J. Roy. Statist. Soc. Ser. B*, 22: 154-171.

[4] Anderson, A. H. 1974. Multidimensional contingency tables. *Scand. J. Statist.,* 1, 115-127.

[5] Barndorff-Nielsen, O. E. 1976. Factorization of likelihood functions for full exponential families. *J. Roy. Statist. Soc. Ser.* B 38: 37-44.

[6] Barndorff-Nielsen, O. E. 1978. *Information and Exponential Families*. Wiley.

[7] Bergsma, W., and Rudas, T. 2003. On conditional and marginal association. *Ann. Fac. Sci. Toulouse*, 11: 455-468.

[8] Birch, M. W. 1963. Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser.*, B(25): 220-233.

[9] Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. Springer.

[10] Bishop, Yvonne M., Fienberg, S. E., and Holland, P. W. 2007. *Discrete Multivariate Analysis*. Springer.

[11] Brown, L. D. 1988. *Fundamentals of statistical exponential families*. Hayward, Calif.: Institute of Mathematical Statistics.

[12] Casella, G., and Berger, R. L. 2002. *Statistical Inference*. Pacific Grove, Calif.: Duxbury.

[13] Christensen, R. 1997. *Loglinear Models and Logistic Regression*. Springer.

[14] Darroch, J. N., and Ratcliff, D. 1972. Generalized iterative scaling for loglinear models. *The Annals of Mathematical Statistics, 43,* 1470-1480.

[15] Fienberg, S. E. 1970. An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics, 41,* 907-917.

[16] Fienberg, Stephen E. 2007. *The Analysis of Cross-Classified Categorical Data*. Springer, Second edition.

[17] Fienberg, Stephen E., and Rinaldo, A. 2007. Three centuries of categorical data analysis: Loglinear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*.

[18] Ghosh, S., and Nyquist, H. 2016. Model fitting and optimal design for a class of binary response models. *Journal of Statistical Planning and Inference*.

[19] Gokhale, D. V., and Solomon K. 1978. *The Information in Contingency Tables*. Marcel Dekker, Inc.

[20] Goodman, L. A. 1968. The analysis of cross-classified data: independence, quasi-independence, and interaction in contingency tables with or without missing cells. *J. Amer. Statist. Assoc.*, 63:1091-1131.

[21] Goodman, L. A. 1972. Some multiplicative models for the analysis of cross-classified data. *Proceedings of the Sixth Berkley Symposium on Mathematical Statistics and Probability*, 649-696.

[22] Goodman, L. . and Hout, M. 1998. Statistical methods and graphical displays for analyzing how the association between two qualitative variables differs among countries, among groups, or over time: a modified regression-type approach. *Sociological Methodology, 28*, 175-230.

[23] Haberman, S. J. 1974. *The Analysis of Frequency Data* (Vol. IV). The University of Chicago Press.

[24] Klimova, A., and Rudas, T. 2015a. Iterative scaling in curved exponential families. *Scandinavian Journal of Statistics*.

[25] Klimova, A., and Rudas, T. 2015b. On the closure of relational models. *Journal of Multivariate Analysis*.

[26] Klimova, A., Rudas, T. and Dobra, A. 2011. Relational models for contingency tables. *Journal of Multivariate Analysis*.

[27] Koch, G. G., and Bhapkar, V. P. 1982. Chi-square tests. *Encyclopedia of Statistical Sciences*, 1.

[28] Koller, Daphne, and Friedman, Nir. 2009. *Probabilistic Graphical Models: Principles and Techniques*. Massachusetts Institute of Technology.

[29] Lauritzen, Stephen L. 1996. *Graphical Models*. Oxford Science Publications.

[30] Rao, A. Ramachandra, and Bhimasankaram, P. 2000. *Linear Algebra*. Hindustan Book Agency (India).

[31] Rao, C. R. 2006. *Linear Statistical Inference and Its Applications*. Wiley.

[32] Rohatgi, V. K., and Saleh, A. K. Md. E. 2001. *An Introduction to Probability and Statistics*. Wiley Series in Probability and Statistics, Second edition.

[33] Ross, Sheldon M. 2010. *Introduction to Probability Models*. Academic Press, 10th edition.

[34] Roy, S. N., and Kastenbaum, M. A. 1956. On the hypothesis of no "interaction" in a multiway contingency table. *Ann. Math. Stat.*, 27: 749-757.

[35] Roy, S. N., and Mitra, S. K. 1956. An Introduction to some nonparametric generalizations of analysis of variance and multivariate analysis. *Biometrica*, 43: 361-376.

[36] Rudas, T. 1998. *Odds Ratios in the Analysis of Contingency Tables*. Sage Publications, Inc.