MARKOVIAN TRANSITION PROBABILITIES: ESTIMATION AND TEST PROCEDURES

by

MAHBOOBEH ZANGENEH SIRDARI

Thesis submitted in fulfillment of the requirements for the degree of

Doctor of Philosophy

2013

ACKNOWLEDGEMENTS

There is not a proper way to say “thank you” to all the people who patiently have assisted me in the development of my thesis. But I should mention some of the people who have been extremely supportive of me in overcoming my weaknesses in many areas.

First and foremost I would like to express my genuine gratitude to my field supervisor, Prof. M. Ataharul Islam for his supervision, advice, guidance and constant support. I really was honored to have the opportunity to work under his supervision.

My grateful thanks also go to my supervisor, Dr. Norhashidah Awang for her supervision and help during my study. I would like to express my appreciation to Prof. Ahamdreza Soltani, my master supervisor, for encourage me to come to Malaysia and continue my study and all his guidance and support.

I would like to express my acknowledgement to the Dean, to the Deputy Dean, and all the office staffs of school of Mathematical Sciences for their support and co-operations towards my postgraduate affairs. I am grateful to Dr. Rafiqul Islam Chowdhury for permission to use “kernopt markov.gen” software, in this study. My acknowledgment also goes to the Health and Retirement Study (HRS) center for their kind permission to use RAND data.

Sincere thanks to all my friends for their kindness and helps, especially Dr. Babak Salamatinia for his moral support and helping me to edit my thesis and continue my work when I encountered difficulties; and Dr. Farshid Bateni for all his helpful guidance in writing my programs and computer problems. Thanks for their friendship and memories.

Last but not least, my deepest gratitude goes to my family who have always believed in me and supported me in whatever I have done, and express my deepest thanks to my sister Zahra for her love, encouragement and all support. I would like to dedicate this thesis to my dearest sister Nasim for her endless love support.

TABLE OF CONTENTS

LIST OF TABLES

Page

exponential form and testing for the association parameter

LIST OF ABREVIATIONS

BMI	Body Mass Index
EAE	Experimental Allergic Encephalomyelitis
EM	Expectation Maximization algorithm
HGLM	Hierarchical Generalized Linear Model
HRS	Health and Retirement Study
LR	Likelihood Ratio
LRT	Likelihood Ratio Test
MLE	Maximum Likelihood Estimate
MTD	Mixture Transition Distribution
MTDg	Generalization of MTD model
NIA	National Institute on Aging
OR	Odds Ratio
RS	Rao’s Score
SSA	Social Security Administration
W	Wald’s test

KEBARANGKALIAN PERALIHAN MARKOVAN: TATACARA PENGANGGARAN DAN UJIAN

ABSTRAK

Pembolehubah-pembolehubah kesudahan yang dijana daripada kajian-kajian longitud secara amnya adalah berkorelasi yang mana menimbulkan suatu cabaran sukar untuk memodelkan data sukatan-sukatan berulang. Adalah berfaedah untuk mengambil perhatian bahawa hubungan antara pembolehubah-pembolehubah bersandar dan juga hubungan antara pembolehubah bersandar dengan pembolehubah penerang boleh mendedahkan maklumat yang sangat berguna kepada penggubal dasar dan penyelidik dalam pelbagai bidang. Dalam memodelkan data dengan kesudahan-kesudahan yang bersandar serta kesudahan-kesudahan yang berkait dengan pembolehubah-pembolehubah penerang berpotensi, model-model sut dan bersyarat boleh digunakan. Kebanyakan hasil kerja yang ditemui dalam literatur adalah berdasarkan model-model sut dan hanya sebahagian kecil hasil kerja menggunakan model-model bersyarat dalam menangani kebersandaran. Dalam suatu kelas model yang berbeza, dikenali sebagai bentuk eksponen kuadratik, model-model dibangunkan supaya mengambil kira kaitan yang wujud antara pembolehubah-pembolehubah kesudahan. Model-model bersyarat, di sisi lain, adalah kebanyakannya berasaskan kepada andaian-andaian Markovan. Model-model ini semakin penting disebabkan sifat yang mendasari hubungan antara pembolehubah-pembolehubah kesudahan yang dijana daripada data longitud.

Dalam kajian ini, tiga objektif utama diutarakan: (i) penganggaran dan ujian bagi model bersyarat dengan kebersandaran kovariat untuk peringkat pertama dan peringkat lebih tinggi serta suatu ujian kebagusan penyuaian berdasarkan skor efisien

Rao telah dilanjutkan bagi kebarangkalian peralihan dengan kebersandaran kovariat untuk peringkat pertama dan peringkat lebih tinggi, (ii) suatu lanjutan model rantai Markov berdasarkan taburan sut dan tatacara pentakbirannya untuk peringkat pertama dan peringkat lebih tinggi dan lanjutan ujian skor efisien untuk menguji kebagusan penyuaian bagi model yang dicadangkan, dan (iii) suatu model lanjutan bagi data Bernoulli bivariat menggunakan bentuk eksponen kuadratik untuk rantai Markov dengan kebersandaran kovariat, tatacara pentakbirannya dan suatu cadangan lanjutan ujian skor efisien bagi menguji kebagusan penyuaian model tersebut. Model bentuk eksponen kuadratik yang dicadangkan mengambil kira respons-respons sut serta parameter berkaitan dengan hubungan antara dua pembolehubah hasil.

Kegunaan ujian yang dicadangkan telah diperiksa dengan suatu contoh data sebenar. Untuk menunjukkan penggunaan ujian yang dicadangkan, kami telah mengaplikasi tatacara penganggaran dan tatacara ujian ini pada data Kajian Kesihatan dan Persaraan (HRS), suatu set data tinjauan isirumah longitud bagi kajian persaraan dan kesihatan warga tua di Amerika Syarikat. Keputusan kajian menunjukkan kesesuaian teknik-teknik yang dicadangkan. Juga, keputusan kajian simulasi mempamerkan kadar penolakan hipotesis benar kekal dalam julat yang munasabah bagi saiz sampel agak kecil atau besar. Ujian kuasa untuk model yang dicadangkan bentuk eksponen kuadratik telah dilakukan berdasarkan kajian simulasi dan keputusan menunjukkan kesahan ujian yang dicadangkan dalam perbandingan dengan ujian alternatif.

MARKOVIAN TRANSITION PROBABILITIES: ESTIMATION AND TEST PROCEDURES

ABSTRACT

The outcome variables generated from longitudinal studies are generally correlated and do pose a formidable challenge to model the repeated measures data. It is worth noting that the relationship between dependent variables as well as between dependent and explanatory variables can reveal very useful information for the policy makers and researchers in various fields. In modeling for data with dependence in outcomes as well as when the outcomes are associated with the potential explanatory variables, marginal or conditional models can be used. Most of the works found in the literature are based on marginal models and only relatively few works employ the conditional models in addressing the dependence. In a different class of models, known as the quadratic exponential form, models are developed in order to take account of underlying associations among the outcome variables. The conditional models, on the other hand, are mostly based on the Markovian assumptions. These models are gaining importance increasingly due to the underlying nature of relationships among the outcome variables generated from longitudinal data.

In this study, three major objectives have been addressed: (i) the estimation and test for the first and higher order conditional models with covariate dependence and a goodness of fit test based on the Rao's efficient score has been extended for the first and higher order transition probabilities with covariate dependence, (ii) an extension of the Markov chain model based on the marginal probabilities and it's inferential

procedure for the first and higher orders and extension of the efficient score based test for testing goodness of fit for the proposed model, and (iii) an extended model for the bivariate Bernoulli data employing the quadratic exponential form for the Markov chain with covariate dependence, it's inferential procedure and a proposed extension of the efficient score test for testing goodness of fit of such model. The proposed quadratic exponential form model takes into account marginal responses as well as a parameter corresponding to association between two outcome variables.

The suitability of the proposed test has been examined with an example for real life data. To display an application of the proposed test, we have applied the estimation and test procedure for the Health and Retirement Study (HRS) data, a longitudinal household survey data set for the study of retirement and health among the elderly in the United States. The results indicate the suitability of the proposed techniques. Also, results of simulations demonstrate the rejection rate of the true hypothesis remains within reasonable range for relatively small or large sample sizes. The power test for the proposed model of the quadratic exponential form has been performed based on the simulation study and the results show the validity of the proposed test in comparison with the alternative tests.

CHAPTER 1
INTRODUCTION

1.1 Background

The importance of repeated measures data has been evidently increasing in various fields such as time series, survival analysis, environmental studies, economics, engineering, reliability, epidemiology, etc. The outcome variables generated from longitudinal studies are generally correlated and do pose a formidable challenge to model the repeated measures data. It is worth noting that the relationship between dependent variables as well as between dependent and explanatory variables can reveal very useful information for the policy makers and researchers in various fields.

In modeling for data with dependence in outcomes as well as when the outcomes are associated with the potential explanatory variables, the marginal or conditional models can be used. Most of the works found in the literature are based on marginal models and only relatively few works employ the conditional models in addressing the dependence. Lee and Nelder (2004), showed that the conditional models are fundamental and marginal predictions can be obtained from the conditional models. In a different class of models, known as the quadratic exponential form, models are developed in order to take account of underlying associations among the outcome variables. The conditional models, on the other hand, are mostly based on the Markovian assumptions. These models are gaining importance increasingly due to the underlying nature of relationships among the outcome variables generated from longitudinal data.

To test for the various associations in the models, the Likelihood Ratio and Wald tests are used. However, it has been observed that the efficient score tests can provide equally good tests and can provide an easier alternative.

1.2 Estimation and test procedure on Markov chain models

During the past, Markovian transition probabilities were estimated and tested in various fields of research. For instance, Bartlett (1951) was able to construct a likelihood ratio test for the goodness of fit by proving the asymptotic normality of certain frequency counts in Markov chains. The test procedure has been designed to test whether a sequence of observations is at most *r*-dependent. It considered that the transition probabilities are known, or at least depend upon a limited number of parameters which can be estimated. If the transition probabilities are completely unknown, a different test is needed. Hoel (1954) presented such a test. The derivation depended heavily upon Bartlett's results and methods, and was essentially a modification and amplification of some of his methods.

Anderson and Goodman (1957) obtained the maximum likelihood estimates and their asymptotic distribution for the transition probabilities in a Markov chain of arbitrary order when there are repeated observations of the chain. The likelihood ratio tests and usual chi-square tests used in contingency tables are obtained for testing the hypotheses. For testing stationary and order of higher-order Markov chain, Billingsley (1961b) used the Whittle’s formula, chi-square and maximum likelihood methods. McQueen and Thorley (1991) used the Markov chain methodology on annual stock returns. Albert (1994) proposed a class of Markov models for analyzing sequences of ordinal data from a relapsing-remitting disease which the state space was expanded to include information about the ordinal severity score as well as the relapsing-remitting status. He proposed a parameterization that reduced the number of parameters. It is noteworthy that most of the research works have been conducted for estimating the parameters based on the first order Markov chain. Recently there were few methods for higher order Markov chains reported, where the estimation and test procedures becomes complex due to increased order of model.

It is of great interest to develop methods of estimation and test procedure for higher order Markov chain models, but the problem is that the number of parameters increases exponentially with the order and it has prevented the use of models for second or higher orders, even when higher order dependence is present. There are some methods to reduce the number of parameters such as mixture transition distribution (MTD) model which was introduced by Raftery (1985) for lth-order Markov chains which combines realism for parsimony. Adke and Deshmukh (1988) obtained the limit distribution of a Markov chain of order $k > 1$ under condition weaker than conditions assumed by Raftery (1985).

More research on estimating and test procedures of parameters of Markov chain model are extended to the new methods where covariates and link functions are used and repeated measures are considered. For example, Muenz and Rubinstein (1985) proposed a model for Markov chain based on covariates and showed how the covariates relate to changes in state. An extensive covariate-dependent for higher order Markov models was improved by Islam and Chowdhury (2006). An influence of time-dependent covariates on the marginal probabilities of binary response has been studied by Azzalini (1994). It has been shown that the covariates relate only to the mean value of the process, independently of the association parameter. Liang and Zeger (1986) used the generalized estimating equations to overcome the difficulties related to 'lack of a rich class of models such as the multivariate Gaussian'. The key feature of their method is that one does not attempt to model the joint distribution of the subject profile; only the marginal distribution at each time point is modeled as a function for data autocorrelation. Azzalini (1994) explored the possibility of developing proper statistical methods for some of the situations for which generalized estimating equations provide a solution. An application of Markov models based on marginal probabilities is provided by Rahman and Islam (2007).

A goodness of fit test for the logistic regression model based on binary data was employed by Tsiatis (1980). He modified the model related to the probability of responses with a set of covariates. For testing the null hypothesis, efficient score test was used. The models for correlated binary data often focus on the dependence of marginal response probabilities on covariates and experimental conditions, although, there is a lack of study based on joint probability distributions that contain convenient estimation of marginal means and correlations for correlated binary data. In this case, logistic representations has been suggested by Cox (1972) which noted a probability distribution of this quadratic exponential form can be reparameterized in terms of marginal parameters of ready interpretation. This has been further discussed by Zhao and Prentice (1990) and they provided a comprehensive estimation procedure for the quadratic exponential form models and provided measures based on covariances (Yi et al., 2009). The quadratic exponential form models have been employed by Hudson et al. (2001) for analyzing familial aggregation of two disorders.

In their work, Van Duijn et al. (2009), compared the bias, standard errors, coverage rates and efficiency of maximum likelihood and maximum pseudo-likelihood estimators by using simulated social network data based on two versions of an empirically realistic network model. They assumed estimation of both the natural parameters and the mean-value parameters. Also, an improved pseudo-likelihood estimation method aimed at reducing bias was proposed by them.

Rao (2005) in a statistical literature on testing hypothesis, showed that the three asymptotic tests, Neyman-Pearson Likelihood Ratio (LR), Wald's statistic (W) and Rao's score (RS) are equivalent to the first-order of asymptotic, but for extension to second-order, they are different.

A large class of test statistics, including the LR, Rao’s score and Wald’s statistics, and their characteristics, is considered by Ghosh and Mukerjee (2001); with reference to the quasi-likelihood arising from an unbiased estimating function. They gave an explicit formula for the third-order power function under contiguous alternatives. Their work can be different based on the usual likelihood, because during working with quasi-likelihood the related Bartlett identities may not hold.

1.3 Problem statement

The problem of some previous methods is that, most of them have been obtained for the first order Markov chain and only a few methods worked on second and higher order such as Muenz and Rubinstein (1985). There are several methods defined and well researched for estimation of parameters such as mixture transition distribution (MTD) model (Adke and Deshmukh, 1988; Islam and Chowdhury, 2006; Raftery, 1985). However, most of the models do not associate explanatory variables in explaining the transition probabilities. Nevertheless, the test procedures of transition probabilities have not been well discovered. The extension of existing models is not easy for higher order. In order to evaluate the parameters using different tests such as stationarity, order of Markov chain, and goodness of fit test, commonly used methods such as the likelihood ratio test and the chi-square test are employed. The use of conventional methods i.e. the likelihood ratio test and the chi-square test require the evaluation of the model based on both null hypothesis and alternative. This could result in increase in the amount of work and time besides the fact that usually there are not sufficient data for alternative hypothesis. Therefore, in order to overcome these drawbacks the efficient scores test is employed in this research which only requires the estimation of parameters under the true null hypothesis.

Correlated outcomes are collected in many areas of research. The correlation between outcomes within individuals is of interest in statistical inferences. Most of the research has been on correlated binary data which are based on the dependence of marginal response probabilities on covariates and experimental conditions.

There are only few studies which focused on the joint probability distributions that contain convenient estimation of marginal means and correlations for binary outcomes (Cox, 1972; Cox and Wermuth, 1994, 2002; Yi et al., 2009; Zhao and Prentice, 1990).

1.4 Objectives of the Investigation

This study is planned and carried out to address the following objectives:

1. To propose higher-order Markov chain models with covariate dependence.
2. To develop alternative test procedures suitable for Markov chain models.
3. To introduce new link functions in the quadratic exponential models for analyzing transition probabilities in order to consider the order of the underlying Markov chain model.
4. To develop estimation and test procedures in the proposed models for higher order Markov chains.
5. To apply the proposed models to real life data.

1.5 Scope of study

In this thesis, the efficient score test which only requires the estimate of parameters under the null hypothesis is provided for the goodness of fit test which could be satisfied to test the order and stationarity of Markov chain model based on the conditional and marginal methods, by considering repeated measurements data. In a Markov chain framework, the current outcomes provide relationships of various orders with previous ones, over a period of time in a longitudinal analysis; which the relationships between outcome and risk factors can also be examined. This study extends the quadratic exponential model for displaying the estimation procedure for the nature and extent of dependence among the binary outcomes. In addition, a test procedure for testing the order of the underlying Markov chain is shown. The proposed models and the test procedures have been examined thoroughly with applications to real life data.

1.6 Organization of the thesis

In this thesis, we will study the development of methods on the estimation and test procedures of transition probabilities for higher order Markov chain models. This thesis contains six main chapters. In the chapter on Literature Review (Chapter 2), an overview of the reported results related to this study and the main basic knowledge about this study such as definition of models, methods of estimations and tests and others are discussed in detail. Chapter 3 introduces the goodness of fit test for the first order Markov chain based on conditional probabilities by considering covariates and their expansion for higher order Markov chains. Then the proposed test is applied to the real life data to examine the suitability of the techniques. Chapter 4 develops the goodness of fit test for the first and higher order Markov chains based on the marginal probabilities and a numerical example of the application of the proposed model is demonstrated. It also provides some simulation results for examining the suitability of the proposed models. In chapter 5 a quadratic exponential model is defined to test the order of Markov chain models, with its application for real life data. In Chapter 6 (Conclusions and Future Research), the overall conclusions based on the results and findings made in the present study are given in brief. Recommendations for future research based on the understanding and knowledge generated in the present study are also given in this chapter.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

The behavior of a Markov chain depends on the values used in the transition matrix, which specifies the probabilities that the system moves from one state to another in unit time. Standard texts assume that the values of such transition matrices are known. However, in most practical studies, this is not the case and the transition matrix needs to be estimated (Billingsley, 1961a).

There are various methods and models for estimating the transition probability matrix. These models were proposed for fitting first or higher order Markov chains. However, most of these studies focused on the first or second order Markov chains and there is a lack of realistic methods for analyzing the parameters of higher order Markov models. In this chapter a review of the past literatures about methods of estimating and test procedure for transition probability matrices are provided.

2.2 Repeated measurement data

In the past, all the models have been for data from studies which for each variable there is just one value, i.e. each person is observed only one time. In longitudinal studies (repeated measures data), each person is observed more than once. In experiments, values of dependent variable are compared before and after training program. For these studies data can be analyzed with the procedures used for cross-sectional data only by considering that the residuals are uniform. But in general such uniformity cannot be assumed because differences between measurements of effects of time or treatment show more variations on some observations. So repeated measurements model must be used.

In statistics, analysis of data becomes more complex when subjects are observed (measured) more than once and we collect repeated measures (longitudinal) data. For instance, this could arise in an experiment where values of an outcome variable before and after a treatment or interference are compared, or the study is about looking at changes over time in one or more outcome variables (Littell et al., 1996).

Let us consider a set of observations, which only take two values, 1 as success and 0 as failure. For the *i*th individual this response can be represented by a random variable, Y_i, called binary variable, without loss of generality code the two possible values of Y_i by 1 and 0 and

$$\mathrm{E}(Y_i) = \Pr(Y_i = 1) = \mu_i, \quad \Pr(Y_i = 0) = 1 - \mu_i.$$

Binary observations of Y_i of n individuals are usually assumed to be independent. The problem is to develop good methods of analysis for assessing any dependence, μ_i, on explanatory variables. (Cox and Snell, 1989)

2.3 Markov chain model

A Markov process $\{Y_t\}$ is a stochastic process with the property that, the probability of any particular future behavior of the process, when its current state is known exactly, is not altered by additional knowledge concerning its past behavior. A discrete-time Markov chain is a Markov process whose state space is a finite or countable set, and its time set is $\boldsymbol{T} = \{0, 1, 2, \dots\}$. Consider a discrete-time random variable X_t taking values in the finite set $\boldsymbol{E} = \{1, 2, \dots, m\}$. The first-order Markov hypothesis says that the present observation at time *t* is conditionally independent of those up to and including time $(t-2)$ given the immediate past [time $(t-1)$]. Thus it can be written,

$$\Pr(Y_t = j | Y_0 = i_0, \dots, Y_{t-2} = i_{t-1}, Y_{t-1} = i) = \Pr(Y_t = j | Y_{t-1} = i)$$

where $i_0, \dots, i_{t-1}, i, j \in \boldsymbol{E}$.

The above probability is called the first-order transition probability and is denoted by $p_{ij}(t)$. That is, $p_{ij}(t) = \Pr(Y_t = j | Y_{t-1} = i)$.

The transition probabilities are functions not only of the initial and final states, but also of the time of transition as well. When the transition probabilities are independent of time, Markov chain has stationary transition probability and we have a homogeneous Markov chain. Then $p_{ij}(t) = p_{ij}$ is independent of *t* and arrange in a matrix

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1\mathrm{m}} \\ p_{21} & p_{22} & \cdots & p_{2\mathrm{m}} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \tag{2.1}$$

and refer to $\boldsymbol{P} = [p_{ij}]$ as the Markov matrix or transition probability matrix of the process. A Markov process is completely defined once its transition probability matrix and initial state Y_0 (or the probability distribution of Y_0) are specified. That is

$$\Pr[Y_0 = i_0, Y_1 = i_1, \dots, Y_t = i_t] = \pi_{i_0} p_{i_0,i_1} \cdots p_{i_{t-2},i_{t-1}} p_{i_{t-1},i_t} ,$$

where $\pi_{i_0} = \Pr(Y_0 = i_0)$. Related computations show that (2.1) is equivalent to the Markov property in the form

$$\Pr(Y_{n+1} = j_1, \dots, Y_{n+m} = j_m | Y_0 = i_0, \dots, Y_n = i_n)$$
$$= \Pr(Y_{n+1} = j_1, \dots, Y_{n+m} = j_m | Y_n = i_n)$$

for all time points *n* and all states $i_0, \dots, i_n, j_0, \dots, j_m$.

The analysis of Markov chain concerns mainly the calculation of the probabilities of the possible realizations of the process. Central in these calculations are the *n*-step probability matrices $\mathbf{P}^{(n)} = \left[p_{ij}^{(n)}\right]$. Here $p_{ij}^{(n)}$ denotes the probability that process goes from state *i* to state *j* in *n* transitions. Formally,

$$p_{ij}^{(n)} = \Pr(Y_{t+n} = j | Y_t = i) .$$

The *n*-step transition probability of a Markov chain satisfy

$$p_{ij}^{(n)} = \sum_{k=0}^{\infty} p_{ik}\, p_{kj}^{(n-1)} ,$$

where $p_{ij}^{(0)} = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases}$.

From the theory of matrices, $\mathbf{P}^{(n)} = \mathbf{P} \times \mathbf{P}^{(n-1)}$. By iterating this formula, $\mathbf{P}^{(n)} = \mathbf{P}^{\mathbf{n}}$.

In some situations, the present depends not only on the last state, but on the last *l* observations. We have then an *l*th-order Markov chain whose transition probabilities are

$$\Pr(Y_t = j | Y_0 = i_0, \dots, Y_{t-2} = i_{t-2}, Y_{t-1} = i)$$
$$= \Pr(Y_t = j | Y_{t-l} = i_l, \dots, Y_{t-1} = i) = p_{i_l \dots ij} .$$

For instance, if we set $l = 2$ and $m = 3$, the corresponding transition matrix is

$$\boldsymbol{P} = \begin{array}{cc} Y_{t-2} & Y_{t-1} \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 1 & 2 \\ 2 & 2 \\ 3 & 2 \\ 1 & 3 \\ 2 & 3 \\ 3 & 3 \end{array} \begin{array}{c} \begin{array}{ccc} & Y_t & \\ 1 & 2 & 3 \end{array} \\ \begin{bmatrix} p_{111} & p_{112} & p_{113} \\ p_{211} & p_{212} & p_{213} \\ p_{311} & p_{312} & p_{313} \\ p_{121} & p_{122} & p_{123} \\ p_{221} & p_{222} & p_{223} \\ p_{321} & p_{322} & p_{323} \\ p_{131} & p_{132} & p_{133} \\ p_{231} & p_{232} & p_{233} \\ p_{331} & p_{332} & p_{333} \end{bmatrix} \end{array}.$$

Whatever the order is, there are $(m - 1)$ independent probabilities in each row of the matrix $\boldsymbol{P}$, the last one of which is completely determined by the others since each row is a probability distribution summing to 1. The total number of independent parameters to be estimated is thus equal to $m^l(m - 1)$.

2.4 Covariates

Consider random variables $Y_1, \dots, Y_n$, not in general binary, and suppose that

$$E(Y_i) = \sum_{s=1}^{p} \beta_s' X_{is} \ , \tag{2.2}$$

where $\{X_{is}\}\ (i = 1, \dots, n; s = 1, \dots, p)$ are known constants and $\beta_1, \dots, \beta_p$ are unknown parameters. In matrix forms

$$E(\boldsymbol{Y}) = \boldsymbol{\beta}'\boldsymbol{X} \ , \qquad E(Y_i) = \boldsymbol{\beta}'\boldsymbol{X}_i \quad (i = 1, \dots, n). \tag{2.3}$$

For binary data, Cox and Snell (1989) considered the model

$$\mu_i = \Pr(Y_i = 1) = E(Y_i) = \sum_{s=1}^{p} \beta_s' X_{is} \ ,$$

via to (2.2).

In representing the dependence of a probability on explanatory variables where the constraint $0 \leq \mu_i \leq 1$ is certainly satisfied, assumes a dependence for $i = 1, \dots, n$,

$$\mu_i = \frac{e^{\boldsymbol{\beta}' \boldsymbol{X}_i}}{1 + e^{\boldsymbol{\beta}' \boldsymbol{X}_i}}, \qquad 1 - \mu_i = \frac{1}{1 + e^{\boldsymbol{\beta}' \boldsymbol{X}_i}}. \tag{2.4}$$

Equations (2.4) are equivalent to

$$\theta_i = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \boldsymbol{\beta}' \boldsymbol{X}_i = \sum_{s=1}^{p} \beta_s' X_{is}, \tag{2.5}$$

or in other way

$$\boldsymbol{\theta} = \boldsymbol{\beta}' \boldsymbol{X} \tag{2.6}$$

The equation $\theta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right)$ is called the logistic transform of the probability μ_i and $\boldsymbol{\theta} = \boldsymbol{\beta}' \boldsymbol{X}$ is a linear logistic model.

In many ways (2.6) is the most useful analogue for binary response data of the linear model (2.3) for normally distributed data. For instance, let $Y_1, \dots, Y_n$ be independent binary random variable distributed in conformity with (2.5) and let $y_1, \dots, y_n$ be the observed values. Then the likelihood is

$$\frac{\prod_{i=1}^{n} e^{\boldsymbol{\beta}' \boldsymbol{X}_i y_i}}{\prod_{i=1}^{n}\left(1 + e^{\boldsymbol{\beta}' \boldsymbol{X}_i}\right)} = \frac{\exp\left(\sum_{s=1}^{p} \beta_s' t_s\right)}{\prod_{i=1}^{n}\left(1 + e^{\boldsymbol{\beta}' \boldsymbol{X}_i}\right)},$$

where $T_s = \sum_i X_{is} Y_i$ and t_s is its observed value. Since the Y_i's are binary, T_s is a random subtotal of the *s*th column of the matrix $\boldsymbol{X}$.

The goal of an analysis using this method is the same as that of any model-building technique used in statistics: To find the best fitting and most parsimonious, yet biologically reasonable model to describe the relationship between an outcome (dependence or response) variable and a set of independent (predictor or explanatory) variables. These independent variables are called covariates.

2.5 Methods of estimation and test procedures

If the transition probabilities are unknown, or else are specified functions of an unknown parameter, there arises the problem of making inferences about them from empirical data. For instance, for estimation and test procedures of parameters we need to know initial probabilities or distribution functions of parameters.

2.5.1 Maximum likelihood estimation

Billingsley (1961b) used the Whittle's formula, chi-square and maximum likelihood methods to estimate and test parameters. Let $\{y_1, y_2, \dots, y_{n+1}\}$ be a sample from a first-order Markov process with transition probabilities, p_{ij}, and initial probabilities, π_i. If $\{a_1, a_2, \dots, a_{n+1}\}$ is a sequence of $(n+1)$ states, then the probability that $y_1, y_2, \dots, y_{n+1}$ is in this sequence is just $\pi_{a_1} \, p_{a_1,a_2} \cdots p_{a_n,a_{n+1}}$. For $i, j = 1, \dots, s$, let n_{ij} be the number of transition from state *i* to state j, with $1 \leq m \leq n$, for which $a_m = i$ and $a_{m+1} = j$. The $s \times s$ matrix $\boldsymbol{F} = \{n_{ij}\}$ will be called the transition count of the sequence (Billingsley, 1961b). He has shown that

$$\sum_{i,j} \frac{\left(n_{ij} - n_i \, p_{ij}\right)^2}{n_i \, p_{ij}} \tag{2.7}$$

is asymptotically chi-square in distribution. If all the p_{ij} are positive, then the degrees of freedom is $s(s-1)$.This chi-square statistic is useful for testing whether the transition probabilities of the process have specified values p_{ij}. There arises naturally in the problem of testing whether these transition probabilities have a specified form $p_{ij}(\theta)$, where θ is an unknown parameter which should be estimated from the sample.

Results of this sort apply equally well, of course, if the number of samples is three or more. It must be assumed, however, that the number of samples is fixed, while the sample sizes go to infinity. A different theory is needed in the opposite case, that in which the samples are of fixed length (say *l*), while the number *n* of them goes to infinity (Billingsley, 1961b).

Berchtold and Raftery (2002) showed that given a set of observations, these parameters can be estimated by log-likelihood of the entire sequence of observations:

$$\ln L = \sum_{i_l \dots i_0 = 1}^{m} n_{i_l \dots i_0} \log\left(\hat{p}_{i_l \dots i_0}\right),$$

where $n_{i_l \dots i_0}$ denote the number of transitions of the type $Y_{t-l} = i_l, \dots, Y_{t-1} = i_1$, $Y_t = i_0$ in the data. The maximum likelihood estimate of the corresponding transition probability $p_{i_l \dots i_0}$ is then $\hat{p}_{i_l \dots i_0} = n_{i_l \dots i_0} / n_{i_l \dots i_1 +}$, where, $n_{i_l \dots i_1 +} = \sum_{i_0=1}^{m} n_{i_l \dots i_0}$.

2.5.2 Maximum pseudo likelihood estimation

In some statistical studies because of the complexity of the likelihood function or for more power, some modification for likelihood is needed. One of the special forms is pseudo-likelihood which is useful when independent variables are complex in the full likelihood (Cox and Reid, 2004).

Cox and Reid (2004), considered random vector $Y = (Y_1, \dots, Y_n)$ with density function $f(y, \theta)$, where θ is an unknown parameter. Maximum likelihood estimator of θ could be found from the independent observations which have approximately normal distribution with mean θ and variance of inverse Fisher information. If it is possible to indicate one or two dimensional distributions up to some order but not full q-dimensional, the univariate and bivariate densities for all $s, t = 1, \dots, q$ can be specified. Then from the vector $\boldsymbol{Y}$ the first and second order log-likelihood contributions can be written as

$$l_1(\theta; Y) = \sum_{s} \log f(y_s;\ \theta)\,,$$

$$l_2(\theta; Y) = \sum_{s>t} \log f(y_s, y_t;\ \theta) - \ aq l_1(\theta; Y)\,,$$

where a is an arbitrary constant. Considering $a = 0$ corresponds to take all possible bivariate distributions and $a = 1/2$ corresponds to take all possible conditional distributions; that is the pseudo-likelihood suggested by Besag (1974) for analyzing spatial data. It may represent the one-dimensional marginal distribution having no information about θ. For many applications it is difficult to construct the full likelihood function. Besag's Pseudo-likelihood is obtained from l_2 with $a = 1/2$ is a similar form in the analysis of spatial data.

For some applications where it is difficult to find the full joint distribution, the pseudo-likelihood function of pairs of observations can provide a useful model. In other applications, the pseudo-likelihood function for bivariate normal distribution is the likelihood for the quadratic exponential distribution, proposed by (Cox, 1972; Cox and Wermuth, 1994; Zhao and Prentice, 1990). The score function of the quadratic exponential is a special case of generalized estimating models where they show consistent estimators of parameters for the mean function. In a similar manner for a range of possible models for higher-order dependency, using the pseudo-likelihood l_2 may lead to consistent estimators of correlation parameters.

Billiot et al. (2008), used maximum pseudo-likelihood estimator for exponential family models of marked Gibbs point processes. Many researchers worked on this area to estimate the energy function based on marked Gibbs point processes. The likelihood function is the best method for estimation if the energy comes from a parametric family model. The main problem is that the likelihood function includes an unknown constant factor which its value depends on the parameters and is difficult to estimate. One way of solving the problem is using the pseudo-likelihood function. They showed that the pseudo-likelihood function contains all properties of estimators such as strong consistency and asymptotic normality.

In a recent paper, Van Duijn et al. (2009) compared the bias, standard errors, coverage rates and efficiency of the maximum likelihood and maximum pseudo-likelihood estimators based on two versions of an empirically realistic network model. Their results showed that it is better to use the maximum likelihood estimator than maximum pseudo-likelihood estimator and maximum bias-corrected pseudo-likelihood estimator; however, in situations when the maximum likelihood is not feasible, then the maximum pseudo-likelihood could be useful.

2.5.3 Score test

If the probability distribution of a set of variables contains k unknown parameters, the statistical hypothesis related to them can be simple or composite. When the hypothesis leads to a complete description of the values of the k parameters, it is a simple hypothesis; and if it leads to a collection of acceptable set, it is a composite hypothesis.

An important problem is the estimation of parameters under restrictions of alternative hypothesis test from which their validity may be tested before the final estimates. There is another problem in estimating parameters when there are restrictions arising from empirical considerations. The precision of the estimates can be enhanced by using the empirical relations. However, slightly imprecise relation may result in bias in the estimates. It is observed that it may still provide better results as compared to that of a less efficient estimate (Rao, 1948).

As shown by Rao (2005), the three asymptotic tests, Likelihood Ratio (LR), Wald's statistic (W) and Rao's score (RS) are equivalent to the first-order of asymptotic, but extension to second-order may be different. In general, if the null hypothesis is rejected, models for the alternative hypothesis should be considered. The score test depends on likelihood function only under true null hypothesis. If the score test is significant, there is no necessity to know about the alternatives.

Let $Y = (Y_1, \dots, Y_n)$ be an iid sample with density function $f(y, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is p-dimensional vector parameter, and denote the log likelihood by $l(\boldsymbol{\theta}; Y)$. The Fisher score vector of p components is defines as

$$\phi(\boldsymbol{\theta}) = \frac{\partial l}{\partial \boldsymbol{\theta}} = \left(\phi_1(\boldsymbol{\theta}), \dots, \phi_p(\boldsymbol{\theta})\right)$$

$$\phi_i(\boldsymbol{\theta}) = \frac{\partial l}{\partial \theta_i}, \quad i = 1, \dots, p.$$

The Fisher information matrix of $p \times p$ is given by

$$\boldsymbol{\mathcal{F}} = (\mathcal{F}_{rs}) = E[\phi(\boldsymbol{\theta})\phi'(\boldsymbol{\theta})].$$

where $\mathcal{F}_{rs} = E[\phi_r(\boldsymbol{\theta})\phi_s(\boldsymbol{\theta})]$. The maximum likelihood estimation of parameter, $\widehat{\boldsymbol{\theta}}$, is acquired from equations

$$\phi_i(\boldsymbol{\theta}) = 0, \quad i = 1, \dots, p.$$

Let $H_0\colon \boldsymbol{\theta} = \boldsymbol{\theta}_0$. To test the null hypothesis, three usual test statistics are as follows.

1. Likelihood ratio test (Neyman and Pearson, 1928)

   $$LRT = -2(\ln L_0 - \ln L_1)$$

   where L_0 is likelihood function under true null hypothesis and L_1 is likelihood function based on alternative hypothesis.

2. Wald test (Wald, 1943)

   $$W = \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \boldsymbol{\mathcal{F}}\left(\widehat{\boldsymbol{\theta}}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$$

   where $\boldsymbol{\mathcal{F}}\left(\widehat{\boldsymbol{\theta}}\right)$ is value of Fisher information matrix based on $\widehat{\boldsymbol{\theta}}$.

3. Rao score test (Rao, 1948)

   $$RS = [\phi(\boldsymbol{\theta}_0)]'[\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)]^{-1}[\phi(\boldsymbol{\theta}_0)].$$

All the three statistics have the same asymptotic chi-square distribution with *p* degrees of freedom.

A large class of test statistics, including the LR, Rao's score and Wald's statistics, and their characterization, is considered by Ghosh and Mukerjee (2001);

with reference to the quasi-likelihood arising from an unbiased estimating function. They mentioned that under the criteria of maximinity and average local power, Rao score test is more (or equally) efficient than LR test and Wald test.

As referred by Rao (1948), generally computation of RS statistic is simple because it depends only on estimates of parameters under null hypothesis. It is constant under transition of the parameters, but not for Wald test. The RS test has the same local efficiency as the Wald and LR tests. Parameters do not affect the distribution of RS test under null hypothesis parameter space; but in the same case for the LR test and sometimes Wald test, they are not applicable.

2.5.4 Goodness of fit test, stationarity, and order test for Markov chain

Bartlett (1951) was able to construct a likelihood ratio test for the goodness of fit by proving the asymptotic normality of certain frequency counts in Markov chains. His test is designed to test whether a sequence of observations is at most r-dependent. In developing the test it is assumed that the transition probabilities are known, or at least depend upon a limited number of parameters which can be estimated. If the transition probabilities are completely unknown, a different test is needed. Hoel (1954), presented a similar test. The derivation depended heavily upon Bartlett's results and methods, and was essentially a modification and amplification of some of his methods. Let us consider the problem of testing the hypothesis H_0 that a chain of order $(r-1)$ will suffice. More precisely, if $p_{ij\ldots kl}$ denotes the transition probability for an r-order Markov chain, the hypothesis to be tested is:

$$H_0 : p_{ij\ldots kl} = p'_{j\ldots kl} \quad (i = 1, 2, \ldots, s).$$

The test that will be constructed is an asymptotic version of the likelihood ratio test for composite hypotheses. Now, following Bartlett, the likelihood function is, except for neglected beginning term is given by:

$$L = \prod_{i\,..l} p_{ij\,...kl}^{n_{ij\,...kl}} \,,$$

where the indices range from 1 to s (corresponding to s possible states) and $n_{ij\,...kl}$ denotes the observed frequency of the *r*-chain state *ij... kl.* The maximum-likelihood estimate of $p_{ij\,...kl}$ is $\hat{p}_{ij\,...kl} = n_{ij\,...kl}/n_{ij\,...k.}$, where $n_{ij\,...k.} = \sum_{l=1}^{s} n_{ij\,...kl}$. Using primes to denote parameter under H_0, it follows from previous equation that

$$\hat{p}'_{ij\,...kl} = \hat{p}_{ij\,...kl} = n_{ij\,...kl}/n_{ij\,...k.} \,.$$

The likelihood ratio test for testing H_0 then assumed the form

$$\lambda = \frac{L_0\left(\hat{p}'_{ij\,...kl}\right)}{L\left(\hat{p}_{ij\,...kl}\right)}.$$

It was shown by Bartlett (1951), that $n_{ij\,...kl}$ are asymptotically normally distributed under mild regularity conditions. Anderson and Goodman (1957) obtained the maximum likelihood estimates and their asymptotic distribution for the transition probabilities in a Markov chain of arbitrary order when there are repeated observations of the chain. The likelihood ratio tests and chi-square tests of the form used in contingency tables are obtained for testing the following hypotheses: (a) that the transition probabilities of a first order chain are constant, (b) that in case the transition probabilities are constant, they are specified numbers, and (c) that the process is a *u*th order Markov chain against the alternative it is *r*th but not *u*th order. The stationary transition probabilities, p_{ij}, can be estimated by maximizing the probability $\prod_{i,j} p_{ij}^{n_{ij}}$ with respect to the p_{ij} subject of course to the restrictions $p_{ij} \geq 0$ and $\sum_{j=1}^{m} p_{ij} = 1$, $i = 1, \ldots, m$, when the n_{ij} are the actual observations.

The maximum likelihood estimates for p_{ij} is:

$$\hat{p}_{ij} = n_{ij}/n_i^* = \sum_{t=1}^{T} n_{ij}(t) \Big/ \sum_{k=1}^{m} \sum_{t=1}^{T} n_{ik}(t) = \sum_{t=1}^{T} n_{ij}(t) \Big/ \sum_{t=1}^{T-1} n_i(t)$$

where $n_{ij}(t)$ denote the number of individuals in state i at $t-1$ and j at t, and $n_i(t-1) = \sum_{j=1}^{m} n_{ij}(t)$. Hence this is also true for any other distribution in which the elementary probability is of the same form except for parameter-free factors, and the restrictions on the p_{ij} s are the same.

When the transition probabilities are not necessarily stationary, the general approach used in the preceding paragraph can still be applied, and the maximum likelihood estimates for the $p_{ij}(t)$s are found to be

$$\hat{p}_{ij}(t) = n_{ij}(t)/n_i(t) = n_{ij}(t) \Big/ \sum_{k=1}^{m} n_{ik}(t) \, .$$

For tests of hypotheses and confidence regions, Anderson and Goodman first considered testing the hypothesis that certain transition probabilities p_{ij} have specified values p_{ij}^0. The null hypothesis is $H_0: p_{ij} = p_{ij}^0$, $j = 1, \ldots, m$, for given i. if the null hypothesis is true,

$$\sum_{j=1}^{m} n_i^* \frac{\left(\hat{p}_{ij} - p_{ij}^0\right)^2}{p_{ij}^0} \tag{2.8}$$

has an asymptotic χ^2 distribution with $m-1$ degrees of freedom. A test for all p_{ij} $(i, j = 1, \ldots, m)$ can be obtained by summing (2.8) over all i, the result is a χ^2 variable with $m(m-1)$ degrees of freedom.

The null hypothesis $H_0: p_{ij}(t) = p_{ij} \ (t = 1, \ldots, T)$, is that the transition probabilities are constant. Under the alternative hypothesis, the estimates of the transition probabilities are

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t-1)},$$

and the likelihood function maximized is

$$\prod_{t=1}^{T} \prod_{i,j} \hat{p}_{ij}(t)^{n_{ij}(t)}.$$

The likelihood function maximized under the null hypothesis is

$$\prod_{t=1}^{T} \prod_{i,j} \hat{p}_{ij}^{n_{ij}(t)}.$$

The likelihood ratio criterion is defined

$$\lambda = \prod_{t} \prod_{i,j} \left[\frac{\hat{p}_{ij}}{\hat{p}_{ij}(t)}\right]^{n_{ij}(t)}. \tag{2.9}$$

When the null hypothesis is true, $-2 \log \lambda$ has a χ^2 distribution with $(T - 1)\, m\, (m-1)$ degrees of freedom. The likelihood ratio (2.9) is similar to the likelihood ratios obtained for standard tests of homogeneity in contingency tables. So,

$$\chi^2 = \sum_{i=1}^{m} \chi_i^2 = \sum_{i} \sum_{t,j} n_i(t-1) \frac{\left[\hat{p}_{ij}(t) - \hat{p}_{ij}\right]^2}{\hat{p}_{ij}},$$

is distributed as χ^2 with $(T-1)\, m\, (m-1)$ degrees of freedom.

For testing the order of Markov chain, they defined $p_{ij \ldots kl} \ (i, j, \ldots, k, l = 1, \ldots, m)$ as the transition probability of order *r*. The null hypothesis is that the