Mis-specification and Goodness-of-fit in Logistic Regression

Nuri H. Salem Badi

School of Mathematics and Statistics

University of Newcastle upon Tyne

A thesis submitted for the degree of

Doctor of Philosophy

September, 2013

Acknowledgements

I am extremely grateful to my supervisor Professor John N.S. Matthews, for his encouragement, guidance, advice and support throughout the work and completion of this research. I would also like to thank the postgraduate students and the staff in the Mathematics and Statistics Department at Newcastle University. Special praise must be reserved for my family, especially my mother and my wife for their personal support.

Abstract

The logistic regression model has become a standard model for binary outcomes in many areas of application and is widely used in medical statistics. Much work has been carried out to examine the asymptotic behaviour of the distribution of Maximum Likelihood Estimates (MLE) for the logistic regression model, although the most widely known properties apply only if the assumed model is correct. There has been much work on goodness-of-fit tests to address the last point. The first part of this thesis investigates the behaviour of the asymptotic distribution of the (MLE) under a form of model mis-specification, namely when covariates from the true model are omitted from the fitted model. When the incorrect model is fitted the maximum likelihood estimates converge to the least false values. In this work, key integrals cannot be evaluated explicitly but we use properties of the skew-Normal distribution and the approximation of the Logit by a suitable Probit function to obtain a good approximation for the least false values. The second part of the thesis investigates the assessment of a particular goodness-of-fit test namely the information matrix test (IM) test as applied to binary data models. Kuss (2002), claimed that the IM test has reasonable power compared with other statistics. In this part of the thesis we investigate this claim, consider the distribution of the moments of the IM statistic and the asymptotic distribution of the IM test (IMT) statistic. We had difficulty in reproducing the results claimed by Kuss (2002) and considered that this was probably due to the near singularity of the variance of IMT. We define a new form of the IMT statistic, $IMTR$, which addresses this issue.

Contents

List of Figures

Chapter 1

Review of Literature and Background

1.1 Introduction

The idea of regression analysis is to explain the dependence of a response variable on one or more covariates (sometimes known as predictor variables). In most statistical analyses, the goal of regression is to summarize observed data as simply, usefully and elegantly as possible. In some problems a theory may be available that specifies how the response varies as the values of the covariates change. So, the first step in regression analysis is to draw appropriate graphs to illustrate the data. In much work to analyse the data we wish to investigate how the changes in one or more variables affect other variables. We often assume that the mean response is a linear function of the covariates. This is the important instance of regression methodology called the linear regression: this method is most commonly used in regression when the outcome is continuous. So, in this case the changed in response variable which is effected by changes in covariates can be explained by fitting a linear regression model. For more details for applied linear regression, see Draper and Smith (1996) and Weisberg (2005).

In fact, in several fields, especially in medical statistics, we need to analysis dichotomous outcome variables. In this case the binary outcome takes only one of the two values, 0 or 1, to denote the absent and present variable respectively. Many examples of binary data are discussed by Cox and Snell (1989) and McCullagh and Nelder (1989). If we used the linear regression model for a binary outcome it would violate the fundamental assumption upon which both the linear model is based. These violations make the linear regression model inappropriate model for the dichotomous outcome. The logistic regression model has been become commonly used to study the association between a binary response variable Cox and Snell (1989). Its widespread application rests on its easy application and interpretation. A widely used reference for logistic regression is the book by Hosmer and Lemeshow now in its third edition, (with a third author, R.X. Sturdivant). It has been cited 31563 (Google Scholar 10 Sept. 2013).

The logistic regression model plays a major and important role in biostatistics analysis and that is why we are interested to examine this model. The general method of estimation the logistic regression parameter is maximum likelihood (ML). In a very general sense the ML method yields values for the unknown parameters that maximize the probability of the observed set of data.

After fitting a logistic regression model, one of the next essential steps is to investigate how well the proposed model fits the observed data; this is called as its goodness-of-fit test. There are many statistics used as goodness-of-fit test for logistic regression model. Hosmer et al. (1997), reported comparison between some of goodness-of-fit tests for logistic regression model. Kuss (2002), discusses the global goodness-of-fit tests for logistic regression model like the standard tests Pearson statistic χ^2, Residual Deviance (D), Residual Sum of Squares Test (RSS), Hosmer and Lemeshow Test and Information Matrix Test IMT.

The subject of the assessment the behaviour of Maximum likelihood estimates (MLE) and goodness-of-fit tests for logistic regression model is important, as the logistic model is widely used in medical statistics. Much work discusses the behaviour of the distribution of MLE for the logistic regression model under the correct model. In the first part of this thesis, our work considers this behaviour and investigates the MLE method under a mis-specified logistic regression model. Claeskens and Hjort (2008), discussed MLE method under the wrong model to find estimation of parameters in terms of the true parameters of the model, called the least false values.

In the second part of this thesis we have investigated the information matrix test (IMT), a goodness-of-fit test to the binary data model which is based on White (1982). Kuss (2002), claimed that the IMT has reasonable power compared with other statistics. Much work in the biostatistical literature has considered the goodness-of-fit tests for logistic regression model, but not work to examine the behaviour the distribution of this statistic. To study and investigate the IMT we need the least false values, which are considered in the first part of this thesis. In the end, under special circumstances sometimes we need other tests to confirm the results, we considered the bootstrap test which was discussed by Efron (1979).

In the rest of this introductory chapter, the expression of the binary data and the form of logistic regression model will be introduced, and then a brief literature review of goodness-of-fit statistic for logistic regression model will be provided.

1.2 Binary Data

Binary data are assumed to be distributed according to a Bernoulli distribution. Suppose that for each individual or experimental unit, the response variable Y_i takes only one of two possible values, 0 or 1. Observations of this nature arise, for instance, in medical trials where, at the end of trial period, the patient has either recovered, denote by $(Y = 1)$ or not $(Y = 0)$. Corresponding to this definition we may write

$$\Pr(Y_i = 1) = \pi_i,\ \Pr(Y_i = 0) = 1 - \pi_i$$

where $P_r(Y_i = 1)$ denotes the probability of success (present) and $P_r(Y_i = 0)$ denote the probability of failure (absent). We assume that such binary observations are available on n individuals, assumed to be independent.

1.2.1 Covariates and Link function

The principal objective of statistical analysis is to examine the relationship between the response variable and available covariates. So, it is important be able, to construct a formal model capable of describing the effect on π_i of changes in covariates. Let's suppose that we have response probability π and the covariates $x_1, x_2, \ldots, x_p$, then the dependence can be described by the linear combination for unknown parameters β_i as

$$\eta = \sum_{i=1}^{p} x_i \beta_i.$$

To investigate the relationship between the response probability π and the covariates, the main problem is that the probability π has to be between 0 and 1. However, the linear combination can take any real value in $(-\infty, \infty)$. A simple solution to solve this problem is to transform π to remove the range restrictions, and use a linear function of the covariates. There is a wide choice of link function $g(\pi)$ $g : (0, 1) \to R$ is available for this purpose. Four functions commonly used in practice are

- The logit link function

$$g_1(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

- The probit link function

$$g_2(\pi) = \Phi^{-1}(\pi)$$

- The complementary log-log link function

$$g_3(\pi) = \log[-\log(1-\pi)]$$

- The log-log link function

$$g_4(\pi) = \log[-\log(\pi)].$$

One of the most used link functions is the logit function

$$g_1(\pi) = \eta = \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$$

and so

$$\pi = \frac{\exp(\eta)}{1+\exp(\eta)},$$

which we also denote by

$$\pi = \text{expit}(\eta).$$

We can see, when the probability goes to 0, then the logit approaches $-\infty$, and at the other extreme, when the probability approaches 1 the logit approaches $+\infty$. Then, the logit link function maps the probabilities from the range $(0, 1)$ to the whole real line $(-\infty, +\infty)$. The behaviour of these link functions shows in the Figures as g_1, g_2, g_3 and g_4 respectively. Figure 1.1, shows compares the four functions, and Figure 1.2, shows comparison which $g_2(\pi)$, $g_3(\pi)$ and $g_4(\pi)$ ploted against $g_1(\pi)$ for values of π in the range $(0.01, 0.99)$. We can see the probit and the logit link function are almost linearly. The link function $g_3(\pi)$ is close to the $g_1(\pi)$, both being close to $\log(\pi)$ when π is small. $g_3(\pi)$ approaches ∞ much more slowly than $g_1(\pi)$ or $g_2(\pi)$ link function when π approaches 1, see Cartinhour (1990). In many cases especially in medical statistic we need to focus on regression model for dichotomous data, the logistic regression model is appropriate. For more information about binary data see McCullagh and Nelder (1989) and Cox and Snell (1989). For the logistic model, we model the effect of covariates by $g_1(\pi_i) = x_i^T\beta$. As will see in the following section, the use of g_1 has some theoretical advantages.

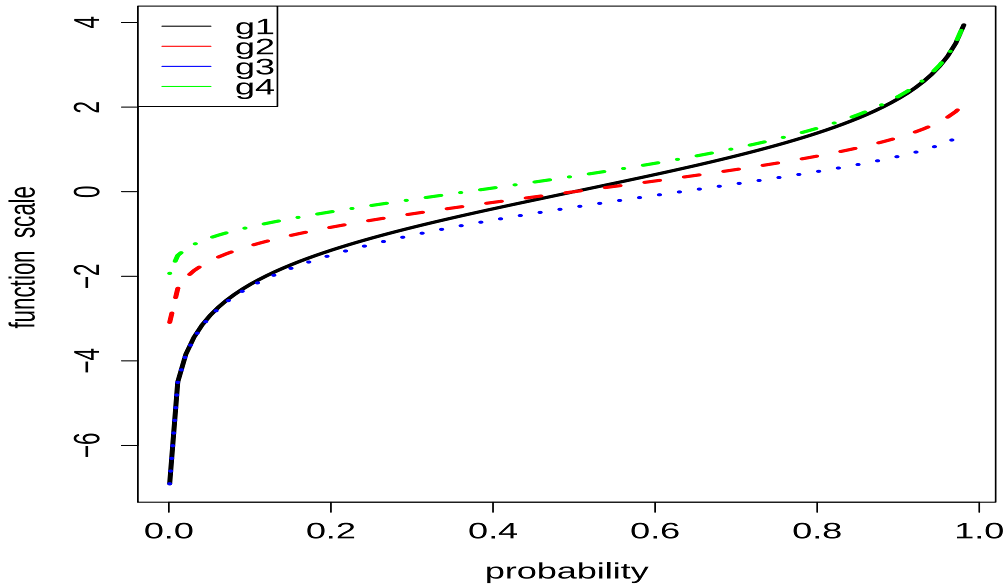

Figure 1.1: Plot of four link functions $g_1(\pi)$, $g_2(\pi)$, $g_3(\pi)$, and $g_4(\pi)$.

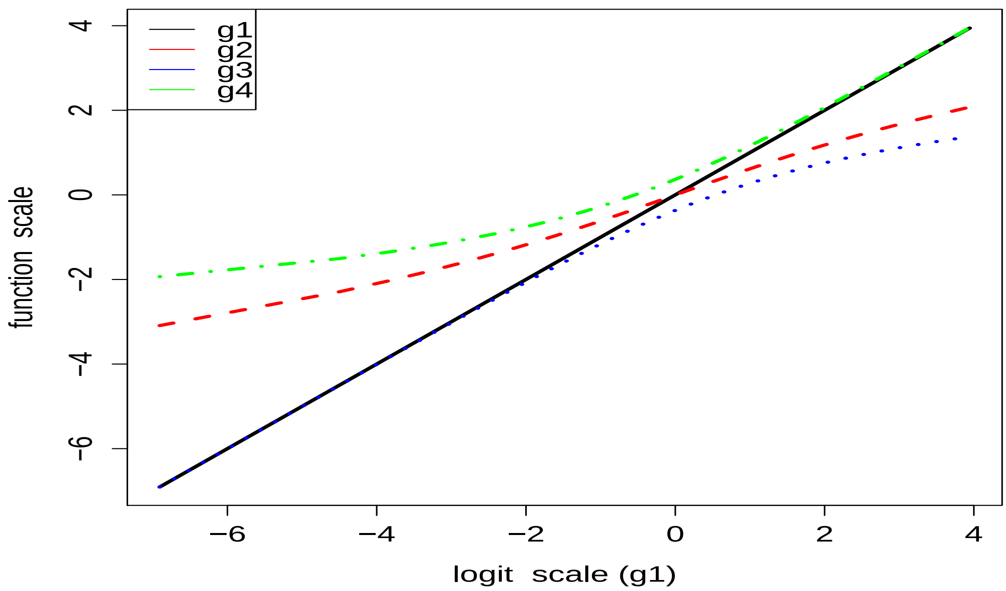

Figure 1.2: Plot of comparison of three link functions $g_2(\pi)$, $g_3(\pi)$, and $g_4(\pi)$ against the link function $g_1(\pi)$.

1.2.2 Binomial Distribution

The binomial distribution concerning to the binary data, let consider we have the response variable y_i is binary. The distribution of Y_i is Bernoulli distribution with parameter π_i. For $y_i \in \{0, 1\}$

$$\Pr(Y_i = y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i},$$

then the mean and variance are

$$E(Y_i) = \mu_i = \pi_i$$

and

$$\mathsf{var}(Y_i) = \sigma^2 = \pi_i(1 - \pi_i).$$

An extension is when Y_i is the number of successes in m_i independent trials. This might be useful when all m_i cases shows the same covariates vector. In this case the data are $Y_i \in \{0, 1, \ldots, m_i\}$ where, m_i denotes the number of observations in group i, and y_i is the number of successes in group i. Then Y_i is distributed as Binomial distribution with parameters π_i and m_i, and the probability distribution function of Y_i is

$$\Pr(Y_i = y_i) = \begin{pmatrix} m_i \\ y_i \end{pmatrix} \pi_i^{y_i}(1 - \pi_i)^{n_i - y_i},$$

Then, the mean and the variance of Y are

$$E(Y_i) = \mu_i = m_i\pi_i$$

and

$$\mathsf{var}(Y_i) = \sigma^2 = m_i\pi_i(1 - \pi_i).$$

Note that the Bernoulli distribution is the special case of the Binomial distribution when $m_i = 1$. In this thesis we only consider the case $m_i = 1$.

1.3 The Logistic Regression Model

The logistic regression model has become the standard analysis tool for binary responses. At present it is used in many fields, particularly in medical research, it is easy for calculation and analysis and interpretation of parameters. It is widely available in software. The goal of a logistic regression analyses is to find the best fitting model to describe the relationship between an outcome and covariates where the outcome is dichotomous. Nelder and Wedderburn (1972), considered the logistic regression model is a member of the class of the generalized linear models. For more details of logistic model see Dobson (1990) and Kleinbaum (1994), also Hosmer and Lemeshow (2000): see also Hosmer et al. (2013).

1.3.1 The Model

Although the general approach models $Y_i \sim binomial(m_i, \pi_i)$, in this thesis we consider the case $m_i = 1$. So, $Y_i \sim binomial(1, \pi_i)$ where π_i is the probability of success for each i, $i = 1, 2, \ldots, n$. Thus $E(Y_i) = \pi_i$ and $\mathsf{var}(Y_i) = \pi_i(1 - \pi_i)$. The logistic regression model can be expressed as

$$\pi_i = \text{expit}(\alpha + X_i^T \beta),$$

where X_i is a p-dimensional vector of covariates. In fitting the logistic regression model to a given set of data, the unknown parameters α and β are estimated by the maximum likelihood method (ML). In this case there are $(p+1)$ likelihood equations which are obtained by differentiating the log likelihood function by each of the $(p+1)$ parameters. The likelihood function is given by

$$L(\alpha, \beta) = \prod_{i=1}^{n} \pi_i^{y_i} [1 - \pi_i]^{1-y_i}.$$

So, the estimation of parameters require the maximization of the likelihood function or equivalently the maximization of the log likelihood function which denoted by

$$\ell(\alpha, \beta) = \log(L(\alpha, \beta)) = \sum_{i=1}^{n} [y_i \log \pi_i + (1 - y_i) \log(1 - \pi_i)]$$

By differentiation of log likelihood functions with respect to parameters we get the following:

$$\sum_{i=1}^{n} [y_i - \pi_i] = 0,$$

and

$$\sum_{i=1}^{n} x_{ij} [y_i - \pi_i] = 0$$

where, $j = 1, 2, \ldots, p$. These result in a solution for parameters α and β, denoted by $\hat{\alpha}$ and $\hat{\beta}$, and the fitted values for the logistic regression model are

$$\hat{\pi}_i = \text{expit}(\hat{\alpha} + x_i^T \hat{\beta}).$$

1.3.2 Example of Analysis of Binary Data By Logistic model

To illustrate analysis by a logistic regression model, let us consider an example shown by Hosmer and Lemeshow (2000). Table 1.1 lists age in year AGE, and presence or absence of evidence of significant coronary heart disease CHD for 100 subjects. Also ID denoted to an identifier variable ID and an age group variable AGRP. The outcome variable is CHD, which is consider a value of 1 to indicate CHD is present, or 0 to indicate that it is absent in the individual. So, it is interesting to explore the relationship between age and CHD. As we can see the outcome variable is binary, so the absence of CHD is (y=0) and the presence of CHD is (y=1). Table 1.2 shows the data by using the age group variable, AGRP, for each age group, the frequency of occurrence of each outcome and the proportion with CHD present is shown.

The logistic model

$$E(Y \mid x) = \pi(x) = \text{expit}(\alpha + \beta_1 AGE),$$

is used to fit the AGE variable, rather than the grouped version. Now we need to fit the logistic regression model to estimate the parameters α and β_1 by maximum likelihood method. The output of analysis the logistic regression model, shows in Table 1.3. Figure 1.3 shows the comparison between the fitted logistic model AGE with AGRP.

The maximum likelihood estimates of α and β_1 are $\hat{\alpha}$=-5.309 and $\hat{\beta} = 0.111$, and the fitted values are given by

$$\hat{\pi}(x) = \text{expit}(-5.309 + 0.111 \times AGE)$$

The Table 1.3, also contains estimates of the standard errors of the estimated coefficients (Std.Err), and the last column displays a p-value. For more applications of the logistic regression model, see Hilbe (2009) and Dobson and Barnett (2008).

ID	AGE	AGRP	CHD	ID	AGE	AGRP	CHD
1	20	1	0	51	44	4	1
2	23	1	0	52	44	4	1
3	24	1	0	53	45	5	1
4	25	1	0	54	45	5	1
5	25	1	1	55	46	5	0
6	26	1	0	56	46	5	1
7	26	1	0	57	47	5	0
8	28	1	0	58	47	5	0
9	28	1	0	59	47	5	1
10	29	1	0	60	48	5	0
11	30	2	0	61	48	5	1
12	30	2	0	62	48	5	1
13	30	2	0	63	49	5	0
14	30	2	0	64	49	5	0
15	30	2	0	65	49	5	1
16	30	2	1	66	50	6	0
17	32	2	0	67	50	6	1
18	32	2	0	68	51	6	0
19	33	2	0	69	52	6	0
20	33	2	0	70	52	6	1
21	34	2	0	71	53	6	1
22	34	2	0	72	53	6	1
23	34	2	1	73	54	6	1
24	34	2	0	74	55	7	0
25	34	2	0	75	55	7	1
26	35	3	0	76	55	7	1
27	35	3	0	77	56	7	1
28	36	3	0	78	56	7	1
29	36	3	1	79	56	7	1
30	36	3	0	80	57	7	0
31	37	3	0	81	57	7	0
32	37	3	1	82	57	7	1
33	37	3	0	83	57	7	1
34	38	3	0	84	57	7	1
35	38	3	0	85	57	7	1
36	39	3	0	86	58	7	0
37	39	3	1	87	58	7	1
38	40	4	0	88	58	7	1
39	40	4	1	89	59	7	1
40	41	4	0	90	59	7	1
41	41	4	0	91	60	8	0
42	42	4	0	92	60	8	1
43	42	4	0	93	61	8	1
44	42	4	0	94	62	8	1
45	42	4	1	95	62	8	1
46	43	4	0	96	63	8	1
47	43	4	0	97	64	8	0
48	43	4	1	98	64	8	1
49	44	4	0	99	65	8	1
50	44	4	0	100	69	8	1

Table 1.1: Age and Coronary Heart Disease(CHD) Status of 100 Subjects (from Hosmer and Lemeshow (2000))

	CHD			
Age Group	n	Absent	Present	Proportion
20-29	10	9	1	0.10
30-34	15	13	2	0.13
35-39	12	9	3	0.25
40-44	15	10	5	0.33
45-49	13	7	6	0.46
50-54	8	3	5	0.63
55-59	17	4	13	0.76
60-69	10	2	8	0.80
Total	100	57	43	0.43

Table 1.2: Frequency Table of Age Group by CHD

Variable	Coeff	Std.Err	z	$P > \|z\|$
AGE	0.111	0.0241	4.61	< 0.001
Constant	-5.309	1.1337	-4.68	< 0.001
	Log Likelihood=-53.67656			

Table 1.3: Results of Fitting the Logistic Regression Model to the Data in 1.1

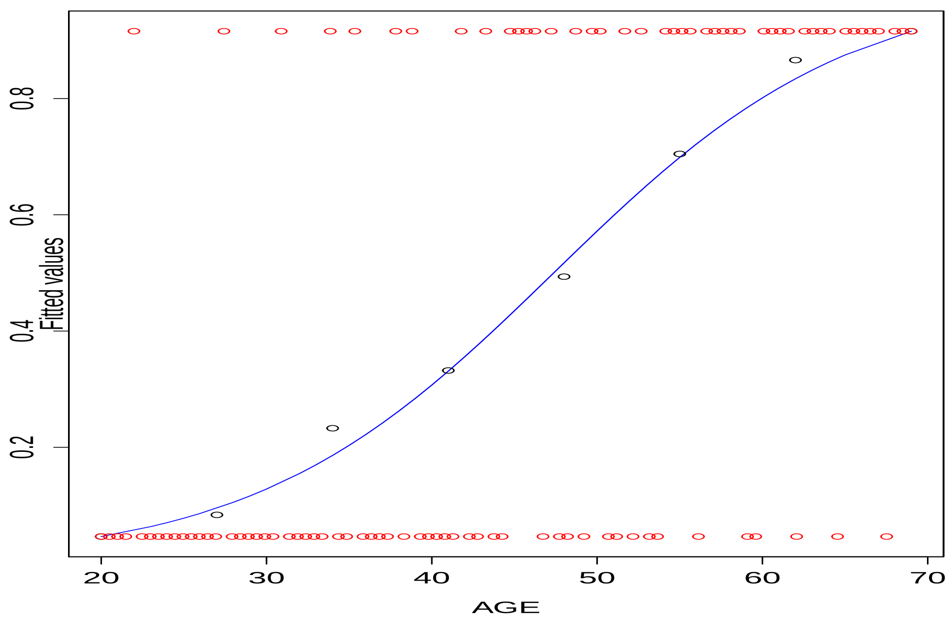

Figure 1.3: Plot of fitted model by AGE (blue) and plot the data of AGRP (black points) and CHD (red points) for (Hosmer and Lemeshow (2000) data).

1.4 Goodness-of-fit Tests

If we look to the Figure 1.3, there are two models fitted to explain the relationship between π and x. The fitted model with AGE covariates denoted by blue line which estimated by two coefficients α and β. However, we have another model with AGRP denoted by black points which makes no assumptions of the form of the relationship between π and age but requires 8 parameters. The goodness-of-fit is very important to decide if the more succinct model is adequate.

The subject of assessment of goodness-of-fit in logistic regression model has attracted the attention of many researchers. It plays an important role in judging the fitted model. After fitting the logistic regression model, the next step is to examine how well the proposed model fits the observed data: this is called as its goodness-of-fit. Goodness-of-fit tests are methods to determine the suitability of the fitted model, and many approaches have been proposed as goodness-of-fit tests for the logistic regression model.

Goodness-of-fit tests for the logistic regression can be split into three types: i) Those based an examination of residuals; ii) Those based an test which group the observation; iii) Those which do not group observation. Methods in i) are more general and subjective assessments of a model and are not considered in this thesis. This is not to undervalue then they are often the most valuable approach to model assessment. The observed values for Bernoulli regression are just 0s and 1s and this makes graphical approaches less easy to handle. The focus of this work is the test statistics. In 1.4.1, tests using grouping are considered, with those that do not need to group the data being discussed in 1.4.2.

1.4.1 Goodness-of-fit Tests with Grouping

Hosmer et al. (1980), proposed and developed approaches involving grouping based on the values of the estimated probabilities obtained from the fitted logistic model. Two grouping methods were proposed. The first approach is based on grouping the data according to percentiles of the estimated probabilities, and the second approach is based on grouping the data according to fixed cutoff values of the estimated probabilities. Tests with grouping based on estimated probabilities were proposed and developed by Hosmer et al. (1980), Lemeshow and Hosmer (1982), Hosmer and Lemeshow (1989) and Hosmer et al. (1997). Brown (1980), developed a score test statistic which essentially compares two fitted model.

Hosmer and Lemeshow Test $\hat{C}$

The calculation of this test dependent upon grouping of estimated probabilities $\hat{\pi}(x_i)$ which use g groups. The first group contains the $n_1 = n/g$ observations which have the smallest estimated probabilities, the second group contains $n_2 = n/g$ values have the next smallest estimated probabilities and the last group contains the $n_g = n/g$ observation with the largest $\hat{\pi}(x_i)$: here n is the size of the sample and g the total number of groups. Before defining a formulae to calculate $\hat{C}$ we will consider some notions. The statistic test $\hat{C}$ is obtained by calculating Pearson chi-square statistic from the $2\times g$ table with two rows and g columns of observed and expected frequencies. In the row with $y = 1$ summing of the all estimated probabilities in a group give the estimated expected value. In the row with $y = 0$ estimated expected value is obtained by summing one minus the estimated probabilities over all subjects in the group. We can denotes the observed number of subjects have had the event present ($y = 1$) and absent ($y = 0$) respectively in each group columns g $(s = 1, 2, 3, \ldots, g)$:

$$O_{1s} = \sum_{i=1}^{n_s} y_i \, , O_{0s} = \sum_{i=1}^{n_s} (1 - y_i)$$

where n_s is the number of the observation in group g. The expected number of subjects of present and absent respectively is denoted by:

$$E_{1s} = \sum_{i=1}^{n_s} \hat{\pi}_i \, , E_{0s} = \sum_{i=1}^{n_s} (1 - \hat{\pi}_i)$$

Then $\hat{C}$ is simply obtained by calculation the Pearson χ^2 statistic for the observed and expected frequencies from the $2 \times g$ table as:

$$\hat{C} = \sum_{s=1}^{g} \sum_{j=0}^{1} \frac{(O_{js} - E_{js})^2}{E_{js}}.$$

from which it following

$$\hat{C} = \sum_{s=1}^{g} \frac{(O_s - n_s\bar{\pi}_s)^2}{n_s\bar{\pi}_s(1 - \bar{\pi}_s)},$$

where, n_s is the total number of values in s^{th} group, O_s is the number of responses for the number of covariates in the s^{th} group, defining as

$$O_s = \sum_{i=1}^{n_s} y_i$$

where, $O_s = O_{1s} + O_{0s}$, and $\bar{\pi}_s$ is the average of the estimated probabilities which are defined as:

$$\bar{\pi}_s = \sum_{i=1}^{n_s} \frac{m_i \hat{\pi}_i}{n_s}$$

Use of an extensive set of simulations proved that when $m_i = 1$, where m_i is the individual binomial denominator and the fitted logistic model is the correct model, then the distribution of $\hat{C}$ is approximated by the χ^2 distribution with $(g-2)$ degrees of freedom Hosmer et al. (1980).

Hosmer and Lemeshow Test $\hat{H}$

The second grouping strategy was proposed from Hosmer and Lemeshow denoted by $\hat{H}$, this method depends upon grouping the estimated probabilities in groups based on fixed cutpoint, so each group contains all subjects with fitted probability located in specific intervals. For example, the cutpoint of the first group is $0.0 \leq \hat{\pi}(x_i) < 0.1$, then this group contains all subjects with estimated probabilities located in this interval; the second group contains all subjects with estimated probabilities located between cutpoint $0.1 \leq \hat{\pi}(x_i) < 0.2$ and the last group has interval $0.9 \leq \hat{\pi}(x_i) < 1.0$. The calculation of $\hat{H}$ uses exactly the same formulae used to calculate $\hat{C}$: the only difference between the two approaches is in the construction of the groups. The distribution of $\hat{H}$ is approximated by the χ^2 distribution with $(g-2)$ degrees of freedom.

Although Hosmer and Lemeshow tests are good, it requires grouping, and choice of g is

- g is arbitrary but almost everywhere in the literature and in software a value of 10, or very similar is chosen.

- Smaller values of g might be chosen for smaller n.

- Sparse data causes a problem for H and lead to uneven group widths for C.

1.4.2 Goodness-of-fit Tests Without Grouping

Deviance and Pearson Chi-Square Tests

Two of the most commonly used goodness-of-fit measures, are the Pearson's chi-squared χ^2 and the deviance D goodness-of-fit test statistics but the behaviour of these tests are unstable with bernoulli data; see McCullagh (1986). The general idea of the deviance is make comparison between two models the first model is full model with p parameters and the second model is a model with q parameters, where $(q < p)$. The deviance can write as

$$D = -2\log\left(\frac{\hat{L}_s}{\hat{L}_r}\right) = -2(\ell_s - \ell_r),$$

Where $\hat{L}_r$, $\hat{L}_s$ are the likelihoods for the full and small model and ℓ_r, ℓ_s denoted to the log-likelihood: Asymptotically this is χ^2 in $p-q$ df. The residual deviance is the case when the large model is saturated and has n parameters. In case of the logistic regression model McCullagh (1986), introduced specific form when $m_i = 1$; the residual deviance can then be found as

$$D = -2\sum_{i=1}^{n}\{\hat{\pi}_i \log\hat{\pi}_i + (1-\hat{\pi}_i)\log(1-\hat{\pi}_i)\},$$

In this case the deviance is invalid as a goodness-of-fit test, because it is a function of $\hat{\pi}_i$, which does not compare the observed values with fitted values.

Also, Pearson chi-square goodness of fit statistic can be written:

$$X^2 = \sum_{i=1}^{n}\frac{(y_i-\hat{\pi}_i)^2}{\hat{\pi}_i(1-\hat{\pi}_i)} = n$$

which is equal to the sample size: this is not a useful goodness-of-fit test.

Residual Sum of Squares Test

Copas (1989), proposed a method, which used the unweighted residual sum of squares a goodness-of-fit test to assess the model adequacy. The idea of this approach is to keep all the individual values of m_i but to give less weight in cases of m_i are small. The unweighted residual sum of squares statistic considers only the numerator of the Pearson chi-squares statistic, which is the summation again over the individual observations,the statistic can be written:

$$RSS = \sum_{i=1}^{n}(y_i-\hat{\pi}_i)^2.$$

Of course, the relative weighting for varying m_i is not relevant for our case where $m_i = 1$. Hosmer et al. (1997), discussed how to compute the moments and asymptotic distribution of the RSS statistic. They give useful expressions for the mean and variance which are easier to compute than the expressions given by Copas (1989). The proposed asymptotic mean and variance of RSS are respectively, $E[RSS - S(W)] \cong 0$

and var$[RSS - S(W)] \cong d^T(I - M)Wd$, where $M = WX(X^TWX)^{-1}X^T$, $W =$ diag $[\pi_i(1-\pi_i)]$, $S(W) = \sum_{i=1}^{n}[diag(\pi_i(1-\pi_i))]$ and d is vector with elements $d_i = (1-2\pi_i)$. Used the standardized statistic to assess significance by referring the following to the standard normal

$$\frac{[RSS - S(W)]}{\sqrt{\mathsf{var}[RSS - S(W)]}}.$$

R^2 Test

Several R^2 type statistics have been used for goodness-of-fit in logistic regression, such as that proposed by Cox and Snell (1989)

$$R_g^2 = 1 - \left(\frac{\hat{L}_c}{\hat{L}_0}\right)^{n/2}$$

where, $\hat{L}_c$ represents the log-likelihood evaluated at the ML estimation parameters and L_0 represents the log-likelihood of the model containing only an intercept. Another version due to Nagelkerke (1991) is

$$\bar{R}_g^2 = \frac{R_g^2}{\max(R_g^2)}$$

where, $\max(R_g^2) = 1 - (\hat{L}_0)^{2/n}$.

1.5 Information Matrix tests: IMT and IMT_{DIAG}

The Information Matrix test (IMT) is a test for general misspecification, proposed by White (1982). The two well-known expressions for the information matrix coincide only if the correct model has been specified and the IMT takes advantage of this fact. The IMT avoids the grouping necessary for tests like the Hosmer-Lemeshow test. Many researchers, (Lancaster (1984), Newey (1984), Davidson and Mackinnon (1984) and Orme (1988)) pointed out the behaviour of the asymptotic distribution of IMT statistic and dispersion matrix. Chesher (1984) discussed the information matrix test and showed that it is useful with binary data models. Kuss (2002), made comparisons between some goodness-of-fit tests in logistic regression models with sparse data. The results of his simulation showed that the IMT has reasonable power compared with other tests. However, Kuss did not give information about the asymptotic distribution of the IMT statistic. Also he did not focus exclusive in the case $m_i = 1$. Although the IMT is extensively discussed in the econometrics literature, it is less well known in the biostatistics literature.

There are several forms of the IMT, some of which give rather unstable behaviour. The reason for this will be explored, and potential corrections suggested in the later part of this thesis. A complication in this analysis is that the test statistic is parameter dependent and must be evaluated at the MLE of the parameters of the fitted model. As such we need to the limiting values of these parameters under what may well be a wrong model.

1.6 Model Mis-specification

It is well known in linear regression that if an outcome is dependent on several covariates but if only a subset of these is fitted, then the parameters estimates obtained will, in general, be inconsistent. The exception is if the fitted and omitted covariates are orthogonal. The situation for logistic regression is less well-known and analytical progress is limited. The topic is of interest in its own right, given the widespread use of logistic regression in biostatistical applications, as well as being necessary for the proper investigation of the use of the IMT for logistic regression. The first part of this thesis will examine this issue in detail for a variety of models.

1.7 Thesis Outline

Part 1: Chapter 2-4

In order to investigate the behaviour of the logistic regression under a mis-specified model, we propose to find expressions for the least false values for some specific forms of covariates. Different distributions of covariates have been considered and compared by simulation. The second chapter poses the main idea of the first part of this thesis: we use the skew normal distribution and the probit function as an approximation to the expit to find the least false values under a logistic model with missing covariates. We present the idea and behaviour of MLE under wrong model as discussed in Claeskens and Hjort (2008). We describe the skew normal distribution and the relationship between the probit function and expit function. We use these to obtain explicit forms for the least false values when the covariates have a multivariate Normal distribution. Chapter 3 introduces the least false value when the covariates assumptions are violated: we consider when the covariates are drawn from three different distributions: multivariate t, multivariate Uniform and bivariate Log-normal distributions. In all cases the least false values are evaluated by simulation. and compared to the values that would be found from the formulae derived for normal covariates. Chapter 4 introduces the behaviour of the MLE when one of the covariates is binary. The form of the least false values for logistic model with one binary covariate and some multivariate normal covariates, some of which are omitted, are also evaluated by simulation. This result is then applied to randomized trials and illustrated by a real example.

Part 2: Chapters 5-8

In chapter 5 consider the basic idea of the IMT statistic and its theory is introduced. The IMT and IMT_{DIAG} are defined for a logistic regression. Chapter 6 considers the moments of the IMT statistic. We calculate the covariance matrix of the IMT statistic under missing covariates and using the least false values. Formulae for the variance of IMT and IMT_{DIAG} are derived and evaluated by simulation. Chapter 7 investigates the asymptotic distribution of the IMT statistic and proposes a new form for the IMT, namely the reduced IMT, $IMTR$. Chapter 8 considers bootstrapping the IMT. In the chapter 9 concluding remarks are made.

Chapter 2

Least false values under missing covariates logistic model

2.1 Introduction

The subject of the behaviour of maximum likelihood estimation (MLE) method in logistic regression model has attracted the attention of many scientists and researchers. Cox (1970) developed the analysis of the binary data and application of the maximum likelihood: see also Cox and Snell (1989). Nelder and Wedderburn (1972) introduced the generalized linear model and used special techniques to obtain the maximum likelihood estimates of the parameters, with observations distributed according to some exponential family. McCullagh and Nelder (1989) discussed the generalized linear model and behaviour of the maximum liklihood (ML) method for binary outcome. The ML method under the wrong logistic model has been discussed extensively by Claeskens and Hjort (2008, p.23). In this chapter we will examine the behaviour of MLE method when the wrong logistic model has been fitted. The idea is to try to find in terms of the true parameters of the model the least false values which are obtained by maximising the incorrect likelihood function. We will use the relationship between $\text{expit}(u) = e^u/(1 + e^u)$ function and probit function $\Phi(\cdot)$, and use the properties of the multivariate skew-Normal distribution to compute a good approximation to the least false values under wrong logistic model.

Firstly, we will give an example for a linear regression model. Secondly, we will define and discuss the properties of the skew-normal distribution and the approximation of $\text{expit}(\cdot)$ in terms of $\Phi(\cdot)$, before we discuss the behaviour of the least false values under the wrong logistic model.

2.2 Least False Value for Linear Regression Model

Before we discuss the behaviour of MLE in case of logistic model, it is instructive to consider the example of the linear regression model. In this section, we explain the behaviour of MLE and compute least false values, in the case of missing covariates from a linear regression model. Let consider we have true linear model

$$E(Y) = \alpha + X\beta,$$

where X can be partitioned as $(X_f|X_a)$, so we can write the model as

$$\alpha + X_f\beta_f + X_a\beta_a,$$

where, β_f and β_a are vectors with $p \times 1$ and $q \times 1$ dimensions respectively. But we fit the model when $E(Y)$ is taken to be

$$E(Y) = \alpha + X_f\beta_f.$$

Then, the important question in this case is what is $E_{Y|X}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}_f\end{bmatrix}$? We know that the expectation of the estimators of the parameters for linear regression model is

$$E_{Y|X}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}_f\end{bmatrix} = (X_F^TX_F)^{-1}X_F^TE(Y)$$

where, $X_F = \begin{bmatrix} 1_n & X_f \end{bmatrix}$ has dimension $n \times (p+1)$ and 1_n is an n-dimensional vector of ones. So, we can write

$$E_{Y|X}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}_f\end{bmatrix} = (X_F^TX_F)^{-1}X_F^T\begin{bmatrix} X_F & X_a \end{bmatrix}\begin{bmatrix}\alpha\\ \beta_f\\ \beta_a\end{bmatrix}$$

where, X_a has dimension $n \times q$, then

$$E_{Y|X}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}_f\end{bmatrix} = \begin{bmatrix} I_{p+1} & (X_F^TX_F)^{-1}(X_F^TX_a) \end{bmatrix}\begin{bmatrix}\alpha\\ \beta_f\\ \beta_a\end{bmatrix}$$

so, we get

$$E_{Y|X}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}_f\end{bmatrix} = \begin{bmatrix}\alpha\\ \beta_f\end{bmatrix} + (X_F^TX_F)^{-1}(X_F^TX_a)\beta_a$$

The above is a standard result when we take the X to be fixed. This could be because the X are fixed or because we have conditioned on them. However, for comparison with later results we wish the unconditioned $E((\hat{\alpha}, \hat{\beta}_f)^T)$. We consider X having a normal distribution $X \sim N(0, \Omega)$. So, corresponding to the partition of X_f and X_a we can write the partition of Ω as

$$\Omega = \begin{bmatrix} \Omega_{ff} & \Omega_{fa} \\ \Omega_{af} & \Omega_{aa} \end{bmatrix}$$

Then, the least false values α^* and β^* compute by $E(\hat{\alpha}, \hat{\beta}_f)^T \to (\alpha^*, \beta_f^*)^T$, when $E(\hat{\alpha}, \hat{\beta}_f)^T$ is over joint distribution of (Y, X) i.e, $E_X E_{Y|X}(\hat{\alpha}, \hat{\beta}_f)^T$. Then,

$$E_X E_{Y|X} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_f \end{bmatrix} = \begin{bmatrix} \alpha^* \\ \beta_f^* \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta_f \end{bmatrix} + E\left[(X_F^T X_F)^{-1}(X_F^T X_a)\right] \beta_a \tag{2.1}$$

We can see that, the second part of (2.1) contained the expectation with X_F, X_a which is not independent and we cannot compute it directly, we will use the properties of Wishart distribution to solve the equation (2.1). See Mardia et al. (1979, p.66), the Wishart distribution is discussed where this distribution defined as: If $M(p \times p)$ can be written as $M = X^T X$ where $X(m \times p)$ is a data matrix from $N_p(0, \Omega)$, then M is said to have a Wishart distribution with scale matrix Ω and degrees of freedom parameter m and write $M \sim W_p(\Omega, m)$, $M^{-1} \sim invW_p(\Omega^{-1}, m)$ and

$$E(M^{-1}) = \frac{\Omega^{-1}}{m - p - 1}.$$

Now,we may consider $X_F = \begin{bmatrix} 1_n & HX_f \end{bmatrix}$, where, the centring matrix $H = I_n - \frac{1}{n} 1_n 1_n^T$, then

$$X_F^T X_F = \begin{bmatrix} n & 0 \\ 0 & X_f^T H X_f \end{bmatrix}$$

and, replacing X_a with HX_a

$$X_F^T H X_a = \begin{bmatrix} 0 \\ X_f^T H X_a \end{bmatrix}$$

Then,we can write

$$\beta_f^* = \beta_f + E\left[(X_F^T H X_F)^{-1}(X_F^T H X_a)\right] \beta_a \tag{2.2}$$

We need to use the properties of the Wishart distribution,to evaluate this expectation. Consider M and its sub-matrices, as

$$M = \begin{bmatrix} X_f^T H X_f & X_f^T H X_a \\ X_a^T H X_f & X_a^T H X_a \end{bmatrix} = \begin{bmatrix} M_{ff} & M_{fa} \\ M_{af} & M_{aa} \end{bmatrix}.$$

In this case $M_{ff} \sim W_p(\Omega_{ff}, n-1)$, a p-dimensional wishart distribution, and $M_{aa} - MafM_{ff}^{-1}M_{af} \sim W_q(\Omega_{aa} - \Omega_{af}\Omega_{ff}^{-1}\Omega_{fa}, n-1-p)$ and this is independent of M_{ff},M_{fa}. Now, we will back to work out the calculation of the expectation on equation (2.1), so will starting to compute

$$E_X((X_F^T X_F)^{-1} X_F^T X_a).$$

Now, $M \sim W_{p+q}(\Omega, n-1)$ and $M^{-1} \sim invW_{p+q}(\Omega, n-1)$ where IW denotes the inverse wishart distribution. From this it follows Mardia et al. (1979, p.67).

$$E(M^{-1}) = \frac{\Omega^{-1}}{n-p-q-2},$$

Also

$$M^{-1} = \left[\begin{array}{cc} (M_{ff} - M_{fa} M_{aa}^{-1} M_{af})^{-1} & -M_{ff}^{-1} M_{fa} (M_{aa} - M_{af} M_{ff}^{-1} M_{fa})^{-1} \\ -M_{ff}^{-1} M_{fa} (M_{aa} - M_{af} M_{ff}^{-1} M_{fa})^{-1} & (M_{aa} - M_{af} M_{ff}^{-1} M_{fa})^{-1} \end{array} \right].$$

Then, considering the off-diagonal block

$$-E(M_{ff}^{-1} M_{fa} (M_{aa} - M_{af} M_{ff}^{-1} M_{fa})^{-1}) = -(n-p-q-2)^{-1} \Omega_{ff}^{-1} \Omega_{fa} (\Omega_{aa} - \Omega_{af} \Omega_{aa}^{-1} \Omega_{fa})^{-1}.$$

But, $M_{aa} - M_{af} M_{ff}^{-1} M_{fa} \sim W_q(\Omega_{aa} - \Omega_{af} \Omega_{ff}^{-1} \Omega_{fa}, n-1-p)$, and is independent of M_{ff} and M_{fa}. So,

$$E((M_{aa} - M_{af} M_{ff}^{-1} M_{fa})^{-1}) = \frac{(\Omega_{aa} - \Omega_{af} \Omega_{ff}^{-1} \Omega_{fa})^{-1}}{n-p-q-2}.$$

Now, by independence

$$E(M_{ff}^{-1} M_{fa} (M_{aa} - M_{af} M_{ff}^{-1} M_{fa})^{-1}) = E(M_{ff}^{-1} M_{fa}) E((M_{aa} - M_{af} M_{ff}^{-1} M_{fa})^{-1})$$

and so,

$$E(M_{ff}^{-1} M_{fa}) = \Omega_{ff}^{-1} \Omega_{fa}.$$

Finally, we can write the least false values in equation (2.1), in terms of covariance matrix and the parameters of the true model as

$$E_X E_{Y|X} \left[\begin{array}{c} \hat{\alpha} \\ \hat{\beta}_f \end{array} \right] \to \left[\begin{array}{c} \alpha^* \\ \beta_f^* \end{array} \right] = \left[\begin{array}{c} \alpha \\ \beta_f + \Omega_{ff}^{-1} \Omega_{fa} \beta_a \end{array} \right]. \tag{2.3}$$

So, in particular, $\beta_f^* = \beta_f + \Omega_{ff}^{-1} \Omega_{fa} \beta_a$. Note that $\beta_f^* = \beta_f$ if $\beta_a = 0$ or if $\Omega_{fa} = 0$.

2.3 MLE Under the True Logistic Model

The behaviour of MLE for binary outcome has been discussed more extensively by McCullagh and Nelder (1989). The logistic model when $Y_i \sim$ binomial(m_i, π_i) with $m_i = 1$ can be fitted using the method of maximum likelihood to estimate the parameters. The first step is to construct the likelihood function which is a function of the unknown parameters and data, then choose those values of the parameters that maximize this function. The log-likelihood function is:

$$l(\pi; Y) = \sum_{i=1}^{n} [y_i \text{log}(\pi_i) + (1 - y_i) \log(1 - \pi_i)]$$

Where, in this case we have

$$g(\pi_i) = \eta_i = \log(\pi_i/1 - \pi_i) = (\alpha + \sum_{j=1}^{p} x_{ij}\beta_j).$$

where β is a p-dimensional vector, x_i are a vector of covariates for i^{th} individual and $i = 1, \cdots, n$. The y_i are the realisations of n independent random variables $Y_i \in \{0, 1\}$ and $\Pr(Y_i = 1|x_i) = \pi_i$. Thus we have

$$l(\beta) = \sum_{i=1}^{n} y_i(\alpha + x_i^T\beta) - \sum_{i=1}^{n} \log\left[1 + \exp(\alpha + x_i^T\beta)\right]$$

To estimate the parameters α and β_j we differentiate the log-likelihood function with respect to α and β_j, giving :

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n}\{y_i - \text{expit}(\alpha + x_i^T\beta)\} = \sum_{i=1}^{n}(y_i - \pi_i).$$

And

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \pi_i}{\pi_i(1 - \pi_i)} \frac{\partial \pi_i}{\partial \beta_j}.$$

$$= \sum_{i=1}^{n}\{x_{ij}y_i - x_{ij}\text{expit}(\alpha + x_i^T\beta)\}$$

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n}(y_i - \pi_i)x_{ij}. \tag{2.4}$$

The MLE for α and β_j can be found by setting $\frac{\partial l}{\partial \alpha} = 0$ and $\frac{\partial l}{\partial \beta_j} = 0$ in equation (2.4) in each of the p equations.

If the fitted model is the true model then, the asymptotic distribution of $\hat{\beta}_j$ is $\hat{\beta} \sim$

$N(\beta, I(\beta)^{-1})$ where $I(\beta)$ is the $(p \times p)$ Fisher's information matrix, its $(r,s)^{th}$ element is defined as

$$I_{rs} = \left[-E \left\{ \frac{\partial^2 l}{\partial \beta_r \partial \beta_s} \right\} \right].$$

If is evaluated at MLE $\hat{\beta}$.

2.4 MLE Under the Wrong Model

Claeskens and Hjort (2008) discussed how the maximum likelihood method used to estimate the parameters of a given regression model is affected when the assumed model is incorrect. If the data are independent and identically distributed, the log likelihood function in case of the density $f(y_i, \theta)$ for an individual observation, we can write as:

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f(y_i, \theta).$$

The important question here is, if we fit a model for Y as $f(y \mid \theta)$ when true model is $g(y)$, what value do we estimate for θ ? We have for each value of θ, by the weak large numbers, in probability, as $n \to \infty$

$$n^{-1}\ell_n(\theta) \to E\left(\log f(Y \mid \theta)\right),$$

As Y is actually distributed according to the density $g(y)$ the right hand side is

$$A = \int g(y) \log f(y \mid \theta) dy.$$

We know the ML estimator $\hat{\theta}$ maximises $\ell_n(\theta)$ and so maximises the above. The Kullback-Leibler (KL) divergence is

$$KL(g(y), f(y,\theta)) = \int g(y) \log \frac{g(y)}{f(y,\theta)} dy = \int g(y) \log g(y) - A. \tag{2.5}$$

If the value θ^* minimises the $KL(g(y), f(y,\theta))$ then $\hat{\theta} \to \theta^*$. The value θ^* is called the least false (LF) value.

When there are covariates in the model the Y_i will no longer be identically distribution (although we still assume independence). As such the above argument needs modificated to accommodate covariates. Introduce covariates X, which have distribution function $F(x)$, then the above is adapted as follows.
For model $f(y \mid x, \theta)$, consider KL conditional on X:

$$KL_X(g(Y \mid X), f(Y \mid X, \theta)) = \int g(y \mid X) \log \frac{g(y \mid X)}{f(y \mid X, \theta)} dy.$$

Overall we get

$$KL(g, f_\theta) = E_X(KL_X) = \int\int g(y \mid X) \log \frac{g(y \mid X)}{f(y \mid X, \theta)} dydF(x),$$

and then

$$KL(g, f_\theta) = \int\int g(y \mid X) \log g(y \mid X) dydF(x) - \int\int g(y \mid X) \log f(y(X, \theta)) dydF(x)$$

Now to solve likelihood function and find the least false parameter θ^* which minimises $KL(g, f_\theta)$, then

$$E_X \left(E_g \left(\frac{\partial \log f(Y \mid X, \theta)}{\partial \theta} \right) \right) |_{\theta^*} = 0. \tag{2.6}$$

2.4.1 Application to Logistic Regression

Now, we will apply the MLE method under wrong model on logistic regression model. The idea is to use this method to obtain equations which determine the least false value θ^* for a logistic regression. To explain the behaviour of the MLE in this case we will partition of the vector covariates X, as previous (X_f, X_a). The model is

$$f(Y \mid \theta) = p(Y = y \mid \theta)$$

where, $\theta = (\alpha, \beta)$, so, we can write the logistic model as

$$f(Y \mid \theta) = (1 - \pi)^{1-Y} \pi^Y,$$

where,

$$\pi = \text{expit}(\alpha + \beta_f X_f)$$

is the fitted model. However, this model is mis-specified because the true model is.

$$\pi = \text{expit}(\alpha + \beta_f X_f + \beta_a X_a)$$

Then the ML equations follow,

$$\log f(Y \mid \theta) = (1 - Y) \log(1 - \pi) + Y \log \pi$$

$$= \log(1 - \pi) + Y \log \frac{\pi}{1 - \pi}$$

$$= Y(\alpha + \beta_f X_f) - \log(1 + e^{\alpha + \beta_f X_f}),$$

therefore

$$\frac{\partial \log f(Y \mid \theta)}{\partial \alpha} = Y - \text{expit}(\alpha + \beta_f X_f),$$

and

$$\frac{\partial \log f(Y \mid \theta)}{\partial \beta} = YX_f - X_f \text{expit}(\alpha + \beta_f X_f).$$

So, expectation of these equations are zero when $\theta = \theta^* = (\alpha^*, \beta^*)$. From the above equations where Y is binary, the expectation in this case becomes

$$E_X(E_{Y|X}(Y)) = E_X(\text{expit}(\alpha^* + \beta_f^* X_f)),$$

and

$$E_X(X_f E(Y \mid X)) = E_X(X_f \text{expit}(\alpha^* + \beta_f^* X_f)).$$

The $E(Y \mid X)$ is $\Pr(Y = 1 \mid X)$ and this is given by the true model

$$\Pr(Y = 1 \mid X) = \text{expit}(\alpha + \beta_f^T X_f + \beta_a^T X_a).$$

But we fit the model without X_a. The least false values, α^* and β_f^*, can be found in terms of α, β_f and β_a and the parameters of the distribution of the covariates as from

$$E\left[\text{expit}(\alpha^* + \beta_f^{*T} X_f)\right] = E\left[\text{expit}(\alpha + \beta^T X)\right] \tag{2.7}$$

$$E\left[X_{fj}\text{expit}(\alpha^* + \beta_f^{*T} X_f)\right] = E\left[X_{fj}\text{expit}(\alpha + \beta^T X)\right]. \tag{2.8}$$

where, X_{fj} is the j^{th} element of X_f $(j = 1, \ldots, p)$.

These equations can be solved approximately if X follows a multivariate normal distribution, by approximating expit$(\cdot)$ and using the skew-normal distribution. So, before solving the above equations to find the least false values, we will briefly review the required properties of the skew-normal distribution and the approximation of expit$(\cdot)$.

2.4.2 Previous Work on mis-specification in Logistic Regression

The behaviour of the Mis-specified logistic regression model has been discussed by several researchers, including Lee (1982), Gail et al. (1984), Robinson and Jewell (1991), Neuhaus et al. (1991), Neuhaus and Jewell (1993) and Drake and McQuarrie (1995). Gail et al. (1984), derive conditions on the components of generalized linear model such that omitting covariates related to outcome will not result in asymptotic biases. These conditions are not met for some models, in particular the logistic regression. Gail et al. (1984), derive formulae for estimating bias both for the method of moments and maximum likelihood estimators.

Gail et al. (1984) worked on randomized trials and omitted all covariates except treatment effect. Subsequent contributions, Neuhaus et al. (1991), Neuhaus and Jewell (1993) and Drake and McQuarrie (1995), all of focused on behaviour of bias more by epidemiological applications than trials. Lee (1982), worked on general mis-specified for the multinomial logistic probability model. He tried to find conditions on the random variables response variable Y, X and Z such that, if the true model is

$$\Pr(Y \mid X, Z) = \text{expit}(\alpha_0 + \alpha_1 X + \beta Z)$$

and the fitted model is

$$\Pr(Y \mid X) = \text{expit}(\alpha_0^* + \alpha_1^* X)$$

then the results of analysis given $\alpha_1^* \to \alpha_1$, i.e. estimate of α_1 is unaffected by omission of Z. In fact if, conditional on the Y, the variable Z is independent of X, then the coefficient of X in the mis-specified model is unaffected by omission of Z.

Gail et al. (1984), Robinson and Jewell (1991), Neuhaus et al. (1991), Neuhaus and Jewell (1993) and Drake and McQuarrie (1995) attempt to find solution to the mis-specified likelihood functions equations, by using Taylor series approximations to produce expressions for bias. In fact Taylor series approximation is useful, but limited to small parameter values and can be difficult technically. For example, Drake and McQuarrie (1995) could obtain only a first-order of Taylor series for the model with some covariates omitted, which was also restricted to two scalar covariates X_1 and X_2. Their result is useful for epidemiological studies; but it provides less insight when applied to randomized trials. Consider the true model is

$$E(Y \mid X_1, X_2, T) = \text{expit}(\alpha + \beta_1 X_1 + \beta_2 X_2 + \gamma T),$$

while the fitted model is

$$E(Y \mid X_1, T) = \text{expit}(\alpha^* + \beta_1^* X_1 + \gamma^* T),$$

then the first order solution for the bias $\gamma^* - \gamma$ is

$$\frac{1}{2}\beta_2 \left[E(X_2 \mid T = 1) - E(X_2 \mid T = 0) - [E(X_1 \mid T = 1) - E(X_1 \mid T = 0)] \times B\right]$$

where B depends on the variances and covariances of the covariates. We can see clearly for randomized control trials the bias vanishes, because in this case

$$E(X_j \mid T = 1) - E(X_j \mid T = 0) = 0 \; , j = 1, 2$$

Our proposed work in the first part of this thesis, using properties of the extended skew-Normal distribution, we derive closed-form approximations for the least-false values from a logistic regression with missing covariates, which are not restricted to small parameter values, to avoid the weaknesses of Taylor series. We consider all parameter values and assume the covariates have normal distributions, with some of them omitted.

2.5 Skew-Normal Distribution

The skew-Normal distribution has been discussed and extended by many researchers, the earlier work developed a systematic treatment of this distribution has been given by Azzalini (1985) and Azzalini (1986). More discussion and numerical evidence of the presence of skewness in real data by Hill and Dixon (1982). Cartinhour (1990), have introduced a multivariate extension skew-Normal distribution, also, discussed by Azzalini and Dalla Valle (1996). Some have developed theorems for skew-normal distribution and related multivariate families: see Henze (1986), Chiogna (1998), Azzalini and Capitanio (1999) and Azzalini (2005). Other discussion for quadratic forms and flexible class of skew-symmetric distribution discussed by Loperfido (2001) and Ma and Genton (2004).

2.5.1 Definition

A random variable U is called skew normal with parameter λ, so $U \sim SN(\lambda)$, if its density function is :

$$f(u; \lambda) = 2\phi(u)\Phi(\lambda u) \tag{2.9}$$

where $u \in R$, $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution function of standard normal distribution respectively, that defined by Azzalini (1985). To more demonstrate the impact of λ on shape of density function in equation (2.9), we consider simple example for $(\lambda = 0, 2, 4, 8)$ and set of suitable variables u, is exhibited graphically in Figure 2.1. Location and scale parameters can be introduced if the random variable has density

$$f(u; \lambda) = 2\phi\left(\frac{u-\zeta}{\eta}\right)\Phi\left(\frac{\lambda(u-\zeta)}{\eta}\right).$$

More general case proposed by Arnold et al. (1993) and Arnold and Beaver (2002).

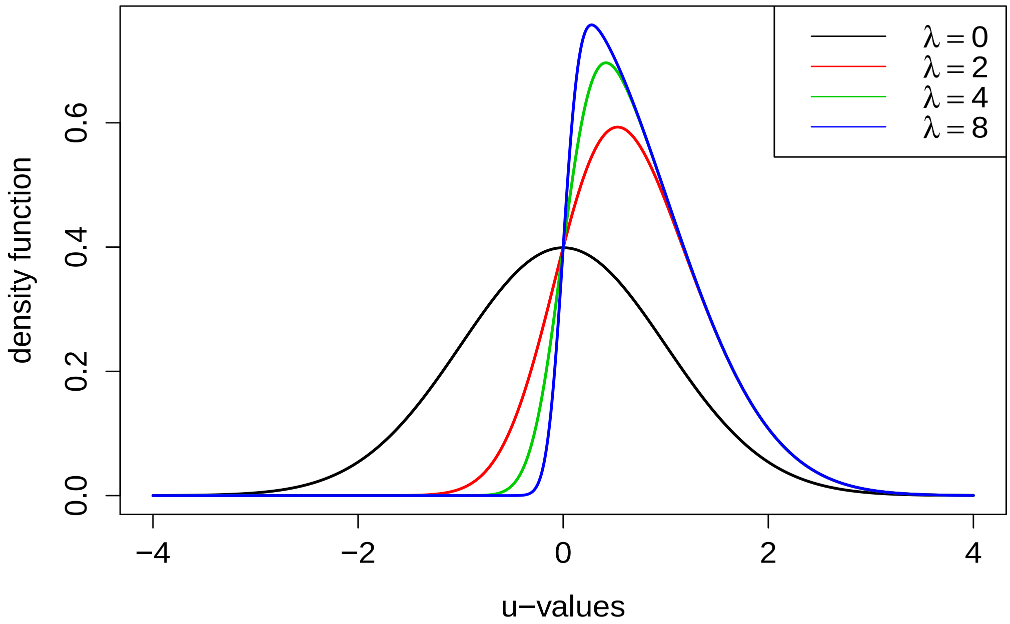

Figure 2.1: plot of pdf of expression (2.9) for different λ.

2.5.2 Extended Multivariate Skew-Normal Distribution

In general case, Arnold and Beaver (2000), discussed extends the skew normal distribution and properties of this family. We can defined the extend multivariate skew-normal distribution as; a p-dimensional random variable U has extended skew-normal distribution, $ESN(\vartheta, \Omega, \lambda, \nu)$, if it has density:

$$\frac{\phi_p(u; \vartheta, \Omega)\Phi(\lambda^T(u-\vartheta)+\nu)}{\Phi(\nu/\sqrt{1+\lambda^T\Omega\lambda})}$$

where ν is a scalar, Ω is dispersion matrix has $p \times p$ dimensional and parameters ϑ and λ are p-dimensional. The $\phi_p(.; \vartheta, \Omega)$ is the density of a p-dimensional normal variable with mean ϑ and dispersion matrix Ω where $\Phi(.)$ is the cumulative distribution function of a univariate standard normal variable. It follows that,

$$E(U) = \vartheta + \frac{\Omega\lambda}{\sqrt{1+\lambda^T\Omega\lambda}}\frac{\phi(\bar{\nu})}{\Phi(\bar{\nu})}$$

where $\bar{\nu} = \frac{\nu}{\sqrt{1+\lambda^T \Omega \lambda}}$ and $\phi(.)$ here, is the density of a univariate standard normal distribution. The moment-generating function of the distribution $ESN(\vartheta, \Omega, \lambda, \nu)$ is

$$M(t) = E(\exp(t^T U)) = \frac{\exp\left(\frac{1}{2} t^T \Omega t\right) \Phi\left(\frac{\nu + \lambda^T \Omega t}{\sqrt{1+\lambda^T \Omega \lambda}}\right)}{\Phi\left(\frac{\nu}{\sqrt{(1+\lambda^T \Omega \lambda)}}\right)}.$$

2.6 Relationship Between Probit Function and expit Function

In this section, we are going to consider the approximation of expit$(\cdot)$ by $\Phi(\cdot)$, the distribution function of a standard normal variable. Gumbel (1961) reported that, the logistic distribution closely resembles the normal distribution which discussed the shape of distribution both are symmetrical and noted some properties. Johnson and Kotz (1970, p.5), point out the comparison of logistic and normal cumulative distribution function. The approximation form defined as:

$$\text{expit}(u) \approx \Phi(ku)$$

where $k = (16\sqrt{3})/(15\pi)$. The approximation between expit(u) and $\Phi(ku)$, is shown in Figure (2.2). Moreover, Figure (2.3) appeared the ratio between expit(u) and $\Phi(ku)$. We can see that, the ratio is poor when the value of u is negative, but only for values smaller than we are likely to estimate. That is because $\Phi(\cdot)$ tends to zero much quicker than expit$(\cdot)$ as $u \longrightarrow -\infty$.

The logistic distribution has a shape similar to the normal distribution, which makes it useful to replace the normal distribution by the logistic distribution to simplify the analysis. The idea is, suppose that if the cumulative function of the standard normal is

$$Q_1(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp -(\frac{1}{2} v^2) dv,$$

and the cumulative distribution function of the standard logistic is

$$Q_2(x) = \left\{1 + \exp\left(-\frac{\pi x}{\sqrt{3}}\right)\right\}^{-1}.$$

I.e both have mean 0 and variance 1. So we expect

$$\text{expit}(x) = Q_2\left(\frac{x}{\pi\sqrt{3}}\right) \approx \Phi\left(\frac{x}{\pi\sqrt{3}}\right).$$

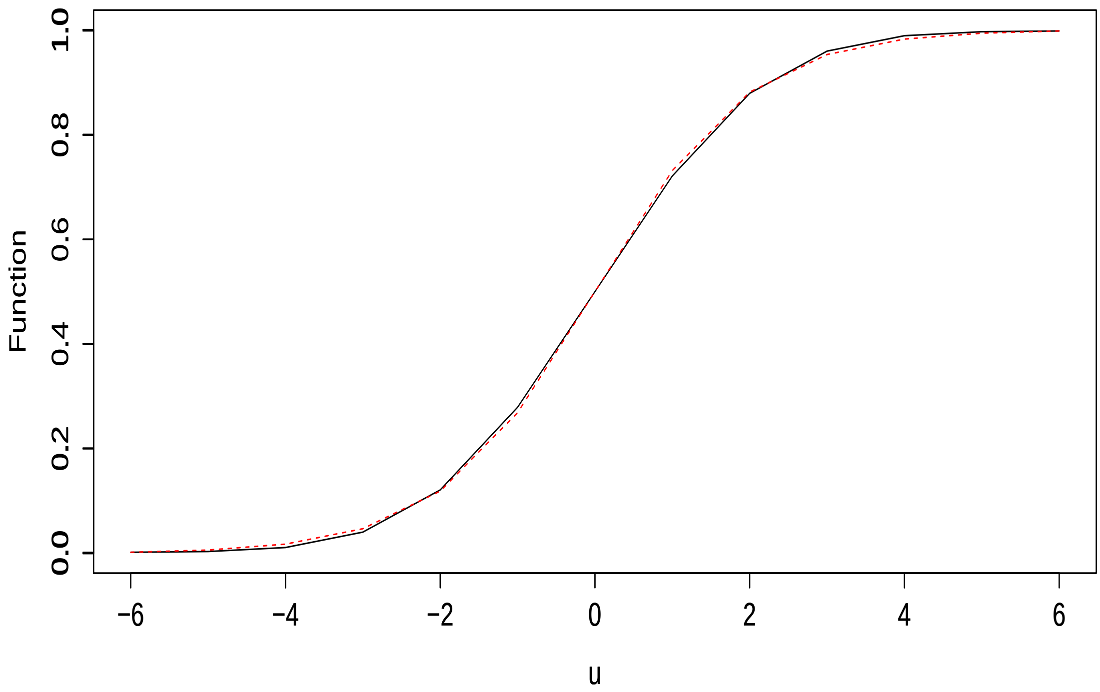

Figure 2.2: Comparison between shape of $\text{expit}(u)$ and $\Phi(ku)$ function.

Now working to make comparison between $Q_1(x)$ and $Q_2(x)$ to explain the similarity of the distribution shape. The differences $[Q_1(x) - Q_2(x)]$ has shown in Figure (2.4), which appeared the maximum value of the differences approximately is about 0.023, when $x = 0.7$. Consideration of the difference only required positive x because $Q_i(-x) = 1 - Q_i(x)$, $i = 1, 2$. By changing the scale of x in cumulative function of standard normal distribution, we plot $Q_2(x) - Q_1(kx)$ against x for a range of values of $k = \frac{n}{n-1}$. The change has been shown in Figure (2.4) as well. It shows that, the best quotient is $k = \frac{16}{15}$ satisfy reduced the maximum value of differences $[Q_2(x) - Q_1(16x/15)]$ to less than 1%. Further discussion that, although the shape of the logistic and normal distribution is a close to similar, there is some differences in some cases of parameters value, which may be because, the logistic distribution has long tails. Moreover, may also note that the curve of the standard normal has points of inflection at $x = \pm 1$, whereas the logistic curve are $x = \pm(\sqrt{3}/\pi)\log(2+\sqrt{3}) = \pm 0.53$. The above results have been discussed by Johnson and Kotz (1970).

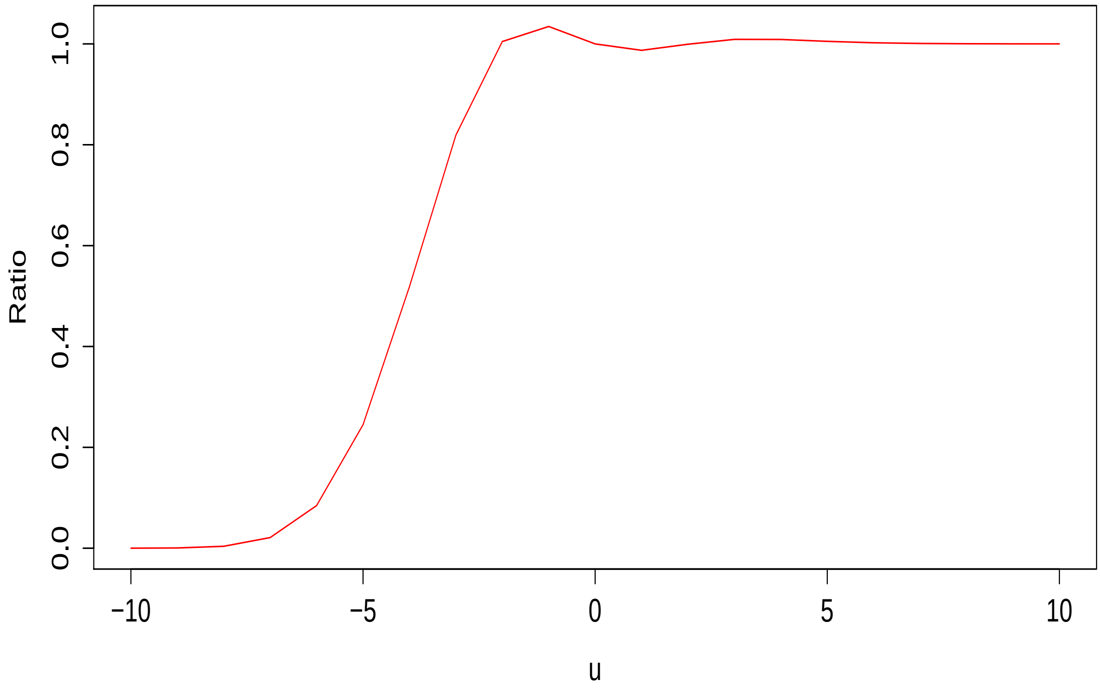

Figure 2.3: Plot of the Ratio between the expit(u) and $\Phi(ku)$ function.

2.7 Least False Values Under Missing Covariates Logistic Model

This section presents the main result of this chapter, namely the least false values for the logistic regression model with missing covariates. The main point is, suppose that we model a binary outcome, Y, using a logistic regression, i.e.

$$\Pr(Y = 1|X_f) = \text{expit}(\alpha + \beta_f^T X_f),$$

but that the true model includes more covariates, i.e.

$$\Pr(Y = 1|X) = \text{expit}(\alpha + \beta_f^T X_f + \beta_a^T X_a).$$

Now,to find the least false values in terms of parameters of the true logistic model, we have three things to do. First, use the approximation form expit$(\cdot) \approx \Phi(k\cdot)$ which discussed in section 2.6. Second, use the properties of the skew-normal distribution which have been shown in section 2.5. Finally, we use the two equations which

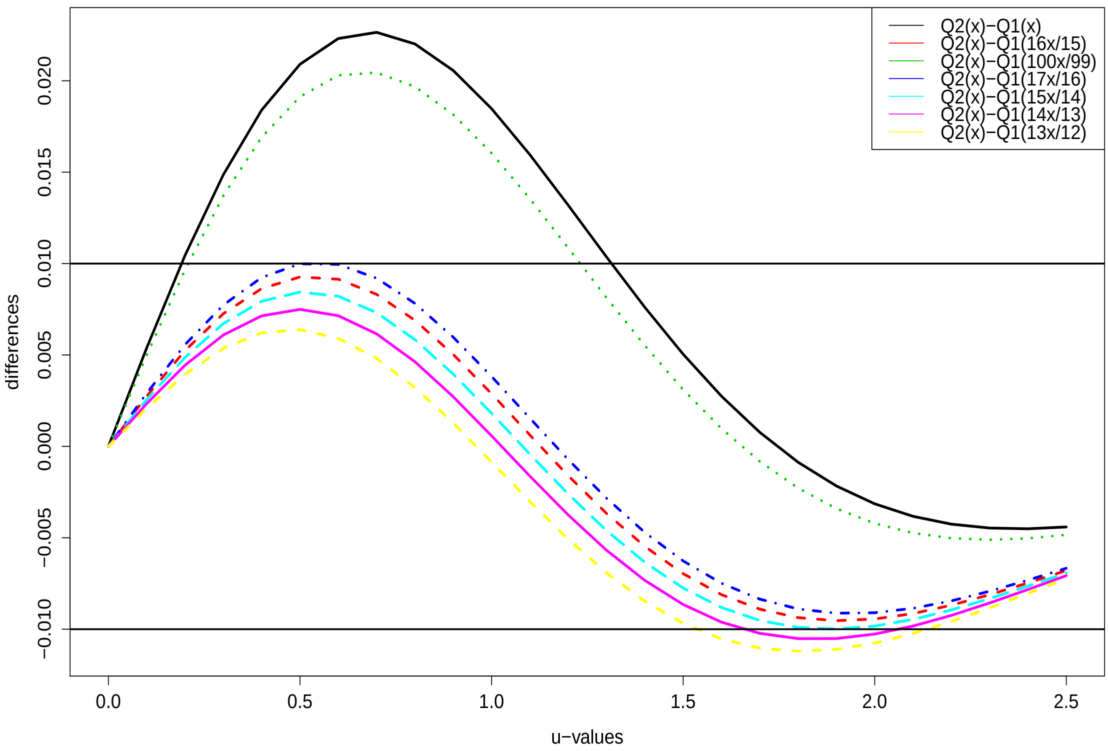

Figure 2.4: Comparision of Logistic and Normal Cumulative Distribution.

determine the $MLEs$, as we have discussed in section 2.4 about MLE under the wrong model to find the least false values. Let us assume that X has $(p+q)$-dimensional multivariate Normal distribution, where p and q denote the dimensions of X_f and X_a respectively. The presence of an intercept in the above models means that we may assume, wlog, that $E(X) = 0$. If $\mathsf{var}(X) = \varOmega$, then also suppose that the partition of this matrix corresponding to X_f and X_a is:

$$\varOmega = \begin{pmatrix} \varOmega_{ff} & \varOmega_{fa} \\ \varOmega_{af} & \varOmega_{aa} \end{pmatrix},$$

then we can apply the approximation to (2.7) and (2.8) using $\text{expit}(u) \approx \varPhi(ku)$, which this leads to

$$E_X\left(\varPhi\left(k[\alpha^* + \beta_f^{*T} X_f]\right)\right) = E_X\left(\varPhi\left(k[\alpha + \beta^T X]\right)\right) \tag{2.10}$$

Now we use the properties of skew-normal distribution, which discussed more expansively in section 2.5, in this case the density function of skew-normal distribution where $E(X) = 0$ is

$$f(X,\alpha,\beta)=\frac{\Phi(k(\alpha+\beta^TX))\phi(X)}{\Phi\left(\frac{k\alpha}{\sqrt{1+k^2\beta^T\Omega\beta}}\right)}.$$

Then we can write the right hand of (2.10) as

$$E_X\left(\Phi\left(k[\alpha+\beta^TX]\right)\right)=\Phi\left(\frac{k\alpha}{\sqrt{1+k^2\beta^T\Omega\beta}}\right)\int\frac{\Phi(k(\alpha+\beta^TX))\phi(X)}{\Phi\left(\frac{k\alpha}{\sqrt{1+k^2\beta^T\Omega\beta}}\right)}dX.$$

Note that, the integration of the second part of the above equation, equal one corresponding to the density function of the skew-normal distribution. So,

$$E_X\left(\Phi\left(k[\alpha+\beta^TX]\right)\right)=\Phi\left(\frac{k\alpha}{\sqrt{1+k^2\beta^T\Omega\beta}}\right)$$

and then, applying an analogous result to the left hand side of (2.10), we obtain

$$\Phi\left(\frac{k\alpha^*}{\sqrt{1+k^2\beta_f^{*T}\Omega_{ff}\beta_f^*}}\right)=\Phi\left(\frac{k\alpha}{\sqrt{1+k^2\beta^T\Omega\beta}}\right),$$

which is

$$\frac{\alpha^*}{\sqrt{1+k^2\beta_f^{*T}\Omega_{ff}\beta_f^*}}=\frac{\alpha}{\sqrt{1+k^2\beta^T\Omega\beta}}. \tag{2.11}$$

Turning our attention to (2.8) and using the results for the expectation of a SN distribution, we obtain

$$\frac{\Omega_{ff}\beta_f^*}{\sqrt{1+k^2\beta_f^{*T}\Omega_{ff}\beta_f^*}}\phi\left(\frac{\alpha^*}{\sqrt{1+k^2\beta_f^{*T}\Omega_{ff}\beta_f^*}}\right)=\frac{(\Omega\beta)_1}{\sqrt{1+k^2\beta^T\Omega\beta}}\phi\left(\frac{\alpha}{\sqrt{1+k^2\beta^T\Omega\beta}}\right), \tag{2.12}$$

where $\phi(\cdot)$ is the standard Normal density, and $(\Omega\beta)_1$ denotes the first p elements of $\Omega\beta$, which is $\Omega_{ff}\beta_f+\Omega_{fa}\beta_a$. Using the result in (2.11), we can simplify (2.12) to

$$\beta_f^*=R(\beta_f+\Omega_{ff}^{-1}\Omega_{fa}\beta_a) \tag{2.13}$$

where

$$R^2=(1+k^2\beta_f^{*T}\Omega_{ff}\beta_f^*)/(1+k^2\beta^T\Omega\beta)$$

Now,using (2.13) we obtain

$$\beta_f^{*T}\Omega_{ff}\beta_f^*=R^2[(\beta_f^T\Omega_{ff}+\beta_a^T\Omega_{af}\Omega_{ff}^{-1}\Omega_{ff})(\beta_f+\Omega_{ff}^{-1}\Omega_{fa}\beta_a)]$$

$$= R^2(\beta_f^T \Omega_{ff}\beta_f + 2\beta_a^T \Omega_{af}\beta_f + \beta_a^T \Omega_{af}\Omega_{ff}^{-1}\Omega_{fa}\beta_a)$$

Now, let $A = \beta_f^{*T}\Omega_{ff}\beta_f^*$, therefore the above amounts to

$$A = \frac{1+k^2A}{1+k^2\beta^T\Omega\beta}(\beta^T\Omega\beta - \beta_a^T\widetilde{\Omega}\beta_a),$$

where $\widetilde{\Omega} = \Omega_{aa} - \Omega_{af}\Omega_{ff}^{-1}\Omega_{fa}$. From this we get

$$A[1 + k^2\beta^T\Omega\beta - k^2(\beta^T\Omega\beta - \beta_a^T\widetilde{\Omega}\beta_a)] = (\beta^T\Omega\beta - \beta_a^T\widetilde{\Omega}\beta_a).$$

Therefore

$$A = \frac{\beta^T\Omega\beta - \beta_a^T\widetilde{\Omega}\beta_a}{1+k^2\beta_a^T\widetilde{\Omega}\beta_a}$$

and hence R^2 can be written

$$\frac{1+k^2A}{1+k^2\beta^T\Omega\beta} = \frac{1+k^2\frac{\beta^T\Omega\beta - \beta_a^T\widetilde{\Omega}\beta_a}{1+k^2\beta_a^T\widetilde{\Omega}\beta_a}}{1+k^2\beta^T\Omega\beta}$$

$$= \frac{1}{1+k^2\beta_a^T\widetilde{\Omega}\beta_a}$$

It follows that

$$\beta_f^* = \frac{1}{\sqrt{1+k^2\beta_a^T\widetilde{\Omega}\beta_a}}(\beta_f + \Omega_{ff}^{-1}\Omega_{fa}\beta_a) \tag{2.14}$$

and

$$\alpha^* = \frac{\alpha}{\sqrt{1+k^2\beta_a^T\widetilde{\Omega}\beta_a}}. \tag{2.15}$$

If we make comparison between equations (2.14) with (2.3) the least false value in case of linear regression model. Note that (2.14) includes a denominator, such that $\beta_f^* \neq \beta_f$ even when $\Omega_{fa} = 0$, although, of course, $\beta_f^* = \beta_f$ if $\beta_a = 0$. However, (2.3) had $\beta_f^* = \beta_f$ if $\Omega_{fa} = 0$ without restrictions on β_a.

2.8 Example When the True Logistic Model has Two Covariates

Let us consider simple example of the computing of least false estimates. Suppose that we have true logistic regression model with two scalar covariates, i.e.

$$\pi_i = \text{expit}(\alpha + \beta X),$$

where X here is two scalar covariates X_1, X_2, then we can write the true model as

$$\pi_i = \text{expit}(\alpha + \beta_1 X_1 + \beta_2 X_2),$$

and we fit standard logistic regression model with one covariate, i.e.

$$\pi_i = \text{expit}(\alpha + \beta_1 X_1).$$

We know that, as the same idea which has been discussed in general case, to find the last false values it will be using in this example to find the two least false parameters α^* and β_1^*. So the least false values in this case satisfy

$$E\left[\text{expit}(\alpha^* + \beta_1^* X_1)\right] = E\left[\text{expit}(\alpha + \beta_1 X_1 + \beta_2 X_2)\right]$$

$$E\left[X_1 \text{expit}(\alpha^* + \beta_1^* X_1)\right] = E\left[X_1 \text{expit}(\alpha + \beta_1 X_1 + \beta_2 X_2)\right]$$

We discussed in the previous section when the covariates has zero mean. Here it is convenient, for later use, to record the result when the covariates have means not equal to zero. Now, consider in this case $X \sim N(\mu, \Omega)$ and we can define $Z = X - \mu$, where $\mu^T = (\mu_1, \mu_2)$, $X^T = (X_1, X_2)$, $Z^T = (Z_1, Z_2)$ and

$$\Omega = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix},$$

assuming $\mathsf{var}(X_1) = \mathsf{var}(X_2) = \sigma^2$. Then, we can write the true model as

$$\Pr(Y|X) = \text{expit}(\alpha + \beta_1 X_1 + \beta_2 X_2)$$
$$\Pr(Y|X) = \text{expit}(\alpha + \beta_1\mu_1 + \beta_2\mu_2 + \beta_1 Z_1 + \beta_2 Z_2)$$

and the fitted model as

$$\Pr(Y|X_1) = \text{expit}(\alpha_1^* + \beta_1^* X_1)$$

$$\Pr(Y|X_1) = \text{expit}(\alpha_1^* + \beta_1^*\mu_1 + \beta_1^* Z_1).$$

Then, the least false equations are written as

$$E\left[\text{expit}(\alpha^* + \beta_1^*\mu_1 + \beta_1^* Z_1)\right] = E\left[\text{expit}(\alpha + \beta_1\mu_1 + \beta_2\mu_2 + \beta_1 Z_1 + \beta_2 Z_2)\right]$$

$$E\left[Z_1 \text{expit}(\alpha^* + \beta_1^*\mu_1 + \beta_1^* Z_1)\right] = E\left[Z_1 \text{expit}(\alpha + \beta_1\mu_1 + \beta_2\mu_2 + \beta_1 Z_1 + \beta_2 Z_2)\right]$$

As we found in general case we can write

$$\alpha^* + \beta_1^*\mu_1 = \frac{\alpha + \beta_1\mu_1 + \beta_2\mu_2}{\sqrt{1 + k^2\beta_2^2\sigma^2(1-\rho^2)}}$$

where in this case, $\widetilde{\Omega} = \sigma^2(1-\rho^2)$. Also, the least false value β_1^* is

$$\beta_1^* = \frac{\beta_1 + \rho\beta_2}{\sqrt{1 + k^2\beta_2^2\sigma^2(1-\rho^2)}}.$$

Similar to the general case the least false value α^* in this case is

$$\alpha^* = \frac{\alpha + \beta_2(\mu_2 - \rho\mu_1)}{\sqrt{1 + k^2\beta_2^2\sigma^2(1-\rho^2)}}.$$

2.8.1 The Least False Value when $\sigma_1^2 \neq \sigma_2^2$

The previous discussion considered the covariance matrix Ω with equal variance. Now if we change this assumption on X to make it drawn from a normal distribution with different variances $\sigma_1^2 \neq \sigma_2^2$. Then, in this case the matrix of Ω is

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

and $\widetilde{\Omega} = \sigma_2^2(1-\rho^2)$. Then, the final expression of the least false value β_1^* in this case is

$$\beta_1^* = \frac{\beta_1 + \rho\frac{\sigma_2}{\sigma_1}\beta_2}{\sqrt{1 + k^2\beta_2^2\sigma_2^2(1-\rho^2)}},$$

and α^* is

$$\alpha^* = \frac{\alpha + \beta_2(\mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1)}{\sqrt{1 + k^2\beta_2^2\sigma_2^2(1-\rho^2)}}.$$

We can see clearly, when $\rho = 0$ or $\mu_1 = \mu_2 = 0$ the expression of least false values are still affected by value of denominator. Only when $\beta_2 = 0$, i.e the fitted model is, in fact, correct, do we obtain $\beta_1^* = \beta_1$ and $\alpha^* = \alpha$.

2.9 Use Logit Link Function Instead of Probit Link Function

In this section we will discuss the point of why use the logit function as link function for logistic regression model instead use probit function directly without the need to use the approximation form $\text{expit}(u) \approx \Phi(ku)$. We know that we use a transformation as link function for logistic model which maps the unit interval onto the whole real line. The commonly used in practice are, the logit function

$$g(\pi) = \log(\pi/(1-\pi))$$

and the probit function

$$g(\pi) = \Phi^{-1}(\pi)$$

So, the question here is why does not use the probit function directly as link function without use the approximation of logit to find the least false value? McCullagh and Nelder (1989), discussed mostly the logit function because of its simple interpretation as the logarithm of the odds ratio. In general, logit is the canonical link function for binomial distribution, which makes it mathematically convenient and provides sufficient statistics. Aldrich and Nelson (1984), discussed the comparison between logit and probit function, the logistic regression model with probit link function is given by:

$$\Pr(Y = 1 \mid X) = \Phi(\beta^T X)$$

and the logit model as we discussed befor is

$$\Pr(Y = 1 \mid X) = \text{expit}(\beta^T X).$$

As we have discussed, the likelihood equation for logistic regression model is

$$\sum_{i=1}^{n} [Y_i - \text{expit}(\beta^T x_i)]x_{ij} = 0,\ j = 1, \cdots, p$$

and that the least false equation come from

$$E_X E_{Y|X} \left[(Y - \text{expit}(\beta^T X))X_j \right] = 0.$$

However, when we set $\pi_i = \Phi(\beta^T x_i)$, the likelihood is

$$L = \prod_{i=1}^{n} (\Phi(\beta^T x_i))^{y_i} (1 - \Phi(\beta^T x_i))^{1-y_i}$$

and the log-liklihood function is

$$l = \sum_{i=1}^{n} y_i \log \Phi(\beta^T x_i) + (1 - y_i) \log(1 - \Phi(\beta^T x_i))$$

then,

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \left[y_i \frac{\phi(\beta^T x_i)}{\Phi(\beta^T x_i)} - (1 - y_i) \frac{\phi(\beta^T x_i)}{\Phi(-\beta^T x_i)} \right] x_{ij}$$

and,

$$\sum_{i=1}^{n} [y_i - \Phi(\beta^T x_i)] W_i x_{ij} = 0,\ j = 1, \cdots, p$$

where

$$W_i = \frac{\phi(\beta^T x_i)}{\Phi(\beta^T x_i)\Phi(-\beta^T x_i)}.$$

Therefore, the least false equations become

$$E_X E_{Y|X}((Y - \Phi(\beta^T X))X_j W) = 0,$$

where

$$W = \frac{\phi(\beta^T X)}{\Phi(\beta^T X)\Phi(-\beta^T X)}.$$

The weight value W in the ML function in case of probit model complicates the function and we cannot use the properties of the skew-normal distribution to find the least false values when the wrong logistic model has been fitted. While in the case of logit model the weight $W = 1$ which makes it easier to use the properties of the skew-normal distribution, as discussed in previous section and find the least false values under missing covariates logistic model.

2.10 Simulation Study of Multivariate Normal Distribution

The goal of this simulation, is to assess the approximation computed for the least false values for logistic regression model. We are interested to application on case of the covariates generated by multivariate Normal distribution. Applied on different cases with different variance and different correlation to check on the behaviour of the formulae of the least false values under missing covariate.

2.10.1 Design of Simulation Study

We looking in this simulation for check the approximation of the last false values for a true logistic regression model has five covariates $p = 5$ is

$$\pi_i = \text{expit}(\alpha + \beta^T X)$$

where, $\beta^T = (\beta_1, \beta_2, \ldots, \beta_5)$, $X = (x_{i1}, \ldots, x_{i5})$ and in the fitted model there are two covariates. We designed the simulation as follows:

- We choose X as a draw from the multivariate normal distribution $X \sim N_5(0, \Omega)$.

- We consider the 5×5 covariance matrix Ω is

$$\Omega = \sigma^2 \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix},$$

where,

$$\Omega_{11} = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{21} & 1 \end{bmatrix}, \ \Omega_{21} = \begin{bmatrix} \rho_{31} & \rho_{32} \\ \rho_{41} & \rho_{42} \\ \rho_{51} & \rho_{52} \end{bmatrix}, \ \Omega_{22} = \begin{bmatrix} 1 & \rho_{34} & \rho_{35} \\ \rho_{43} & 1 & \rho_{45} \\ \rho_{53} & \rho_{54} & 1 \end{bmatrix}, \Omega_{21}^T = \Omega_{12}.$$

- Use three different variance $\sigma^2 = 0.1, 0.5, 1.5$.

- We consider 6 different cases of correlation which is each case of Ω_{ij} has same ρ_{ij} designed as:(0.1,0.1,0.2), (0.2,0.2,0.4), (0.7,0.8,0.7), (0.8,0.7,0.9), (0.1,-0.2,0.4), (0.2,-0.2,-0.2). Values are chosen to assume Ω is positive definite.

- We choose the parameters $\beta_1, \ldots, \beta_5$ and α to give us two cases $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$. As we can calculate the unconditional $\Pr(Y = 1)$,

$$\Pr(Y = 1) = \int \Pr(Y = 1) f(X) dX,$$

$$\Pr(Y = 1) = \int \operatorname{expit}(\alpha + \beta^T X) \phi(X) dX,$$

$$\Pr(Y = 1) \approx \int \Phi(k(\alpha + \beta^T X)) \phi(X) dX,$$

and as we computed by properties skew-normal distribution we get

$$\Pr(Y = 1) \approx \Phi\left(\frac{k\alpha}{\sqrt{1 + k^2 \beta^T \Omega \beta}}\right).$$

Choose $\beta_1 = 0.25, \beta_2 = 0.35, \beta_3 = 0.40, \beta_4 = 0.3, \beta_5 = 0.2$ and adjust α, so that over the covariates $\Pr(Y = 1) \simeq 10\%$ ($\alpha = -2.2$) and $\Pr(Y = 1) \simeq 60\%$ ($\alpha = 0.4$).

- Large sample size has been used $n = 500$, $n = 10000$ and $N = 1000$ number of simulation.

$\sigma^2 = 0.1$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	0.3919	0.3974	0.99	0.3520	0.3296	1.06	0.47194	0.4290	1.10
0.2	0.2	0.4	0.4048	0.3969	1.01	0.4212	0.3969	1.06	0.4999	0.4962	1.01
0.7	0.8	0.7	0.4030	0.3996	1.01	0.6791	0.6730	1.01	0.7473	0.7729	0.97
0.8	0.7	0.9	0.4025	0.3978	1.01	0.6908	0.5967	1.15	0.6381	0.6961	0.92
0.1	-0.2	0.4	0.3932	0.3969	0.99	0.1027	0.0857	1.19	0.2092	0.1849	1.13
0.2	-0.2	-0.2	0.4068	0.3955	1.01	0.0756	0.0998	0.75	0.2481	0.1995	1.24
$\sigma^2 = 0.5$											
0.1	0.1	0.2	0.3989	0.3874	1.02	0.2975	0.3214	0.93	0.4308	0.4183	1.02
0.2	0.2	0.4	0.3950	0.3854	1.02	0.4182	0.3854	1.08	0.4794	0.4818	0.99
0.7	0.8	0.7	0.4205	0.3992	1.05	0.6594	0.6712	0.98	0.7762	0.7705	1.01
0.8	0.7	0.9	0.3968	0.3894	1.01	0.5566	0.5842	0.95	0.6408	0.6815	0.94
0.1	-0.2	0.4	0.3838	0.3856	0.99	0.0927	0.0832	1.11	0.2109	0.1796	1.17
0.2	-0.2	-0.2	0.4024	0.3955	1.01	0.1165	0.0998	1.17	0.1722	0.1995	0.87
$\sigma^2 = 1.5$											
0.1	0.1	0.2	0.3337	0.3656	0.91	0.3268	0.3033	1.07	0.4046	0.3947	1.02
0.2	0.2	0.4	0.3993	0.3606	1.10	0.3693	0.3606	1.02	0.4298	0.4507	0.95
0.7	0.8	0.7	0.3925	0.3954	0.99	0.6775	0.6659	1.01	0.7781	0.7648	1.01
0.8	0.7	0.9	0.3689	0.3706	0.99	0.5579	0.5560	1.00	0.6239	0.6486	0.96
0.1	-0.2	0.4	0.3675	0.3609	1.01	0.0863	0.0779	1.10	0.1661	0.1681	0.99
0.2	-0.2	-0.2	0.3495	0.3869	0.90	0.0996	0.0967	1.03	0.1995	0.1934	1.03

Table 2.1: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Normal distribution in case $\Pr(Y = 1) \simeq 60\%$, $n = 500$ and R_i denote to the Ratio

2.10.2 Results and Discussion

In this part we will show the results and discuss the simulation studies. We report the accuracy of the estimation parameters of the logistic regression model has two covariates when the true model has five covariates. Tables shows comparison between the least false values which is computed by approximation of $\text{expit}(u) \approx \Phi(ku)$ and skew-Normal distribution properties and values of estimated parameters by fitted logistic regression model. R_1, R_2, R_3 denote the ratios of the mean of the simulated fits to the computed least false value.

Table 2.1 and Table 2.2, shows the results of simulation of data generated by multivariate Normal distribution in cases of $\Pr(Y = 1) \simeq 60\%$ and $\Pr(Y = 1) \simeq 10\%$ respectively with sample size $n = 500$. Table 2.3 and Table 2.4, shows the results of simulation with sample size $n = 10000$. We can see clearly the results show ratios close to one. The same behaviour results found in both cases of $\Pr(Y = 60\%)$ and $\Pr(Y = 10\%)$, where is the ratio found close to one. That is meaning the approximation form of the least false values works well, although the probability of outcome Y is very low about 10%, but a good results and reasonable behaviour have been found. Some issues of low ratio a raised in case of sample size $n = 500$, that there are some estimated values were very small close to zero which affect on ratio.

$\sigma^2 = 0.1$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	-2.218	-2.185	1.01	0.3108	0.3296	0.94	0.4344	0.4290	1.01
0.2	0.2	0.4	-2.208	-2.183	1.01	0.4305	0.3969	1.08	0.4841	0.4962	0.98
0.7	0.8	0.7	-2.221	-2.198	1.01	0.6800	0.6730	1.01	0.7959	0.7729	1.02
0.8	0.7	0.9	-2.226	-2.188	1.01	0.6183	0.5967	1.03	0.6868	0.6961	0.98
0.1	-0.2	0.4	-2.2132	-2.183	1.01	0.106	0.0857	1.23	0.1990	0.184	0.97
0.2	-0.2	-0.2	-2.193	-2.194	0.99	0.1043	0.0997	1.04	0.1989	0.1995	1.07
$\sigma^2 = 0.5$											
0.1	0.1	0.2	-2.148	-2.131	1.01	0.3364	0.3214	1.04	0.4414	0.4183	1.05
0.2	0.2	0.4	-2.148	-2.120	1.01	0.4200	0.3854	1.08	0.4830	0.4818	1.00
0.7	0.8	0.7	-2.2031	-2.191	1.01	0.7068	0.6709	1.05	0.7607	0.7705	0.99
0.8	0.7	0.9	-2.161	-2.142	1.01	0.6046	0.5842	1.03	0.6729	0.6815	0.99
0.1	-0.2	0.4	-2.126	-2.120	1.00	0.0939	0.0832	1.12	0.2012	0.1796	1.12
0.2	-0.2	-0.2	-2.217	-2.175	1.01	0.1202	0.0988	1.21	0.1965	0.1977	0.99
$\sigma^2 = 1.5$											
0.1	0.1	0.2	-2.062	-2.011	1.02	0.3380	0.3033	1.11	0.4325	0.3947	1.09
0.2	0.2	0.4	-2.014	-1.983	1.01	0.4165	0.3610	1.15	0.4410	0.4507	0.98
0.7	0.8	0.7	-2.233	-2.175	1.02	0.6910	0.6659	1.03	0.8057	0.7648	1.05
0.8	0.7	0.9	-2.043	-2.039	1.00	0.5435	0.5560	0.98	0.6473	0.6486	0.99
0.1	-0.2	0.4	-1.973	-1.9854	0.99	0.0454	0.0779	0.58	0.1877	0.1681	1.11
0.2	-0.2	-0.2	-2.123	-2.128	0.99	0.0664	0.0967	0.68	0.1953	0.1934	1.01

Table 2.2: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Normal distribution in case $\Pr(Y = 1) \simeq 10\%$, $n = 500$ and R_i denote to the Ratio

$\sigma^2 = 0.1$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	0.3973	0.3974	0.99	0.3261	0.3296	0.99	0.4246	0.4290	0.99
0.2	0.2	0.4	0.3966	0.3969	0.99	0.4062	0.3969	1.02	0.4963	0.4962	1.00
0.7	0.8	0.7	0.4009	0.4000	1.00	0.6705	0.6730	0.99	0.7774	0.7729	1.01
0.8	0.7	0.9	0.3985	0.3981	1.00	0.5951	0.5967	0.99	0.6950	0.6961	0.99
0.1	-0.2	0.4	0.3944	0.3969	0.99	0.0946	0.0857	1.10	0.1827	0.1849	0.99
0.2	-0.2	-0.2	0.3981	0.3990	0.99	0.1025	0.0997	1.02	0.2014	0.1995	1.01
$\sigma^2 = 0.5$											
0.1	0.1	0.2	0.3809	0.3874	0.98	0.3341	0.3214	1.04	0.4154	0.4183	0.99
0.2	0.2	0.4	0.3787	0.3854	0.98	0.3777	0.3854	0.98	0.4764	0.4818	0.99
0.7	0.8	0.7	0.3997	0.3992	1.00	0.6728	0.6712	1.00	0.7652	0.7705	0.99
0.8	0.7	0.9	0.3872	0.3894	0.99	0.5794	0.5842	0.99	0.6696	0.6815	0.98
0.1	-0.2	0.4	0.3824	0.3856	0.99	0.0820	0.0832	0.99	0.1857	0.1796	1.03
0.2	-0.2	-0.2	0.3950	0.3955	0.99	0.0990	0.0998	1.00	0.1956	0.1977	0.99
$\sigma^2 = 1.5$											
0.1	0.1	0.2	0.3589	0.3656	0.98	0.2952	0.3033	0.97	0.3915	0.3947	0.99
0.2	0.2	0.4	0.3341	0.3606	0.93	0.3517	0.3606	0.98	0.4481	0.4507	0.99
0.7	0.8	0.7	0.3977	0.3954	1.01	0.6678	0.6661	1.00	0.7617	0.7648	0.99
0.8	0.7	0.9	0.3631	0.3706	0.98	0.5592	0.5560	1.01	0.6517	0.6486	1.01
0.1	-0.2	0.4	0.3589	0.3609	0.99	0.0854	0.0779	1.09	0.1629	0.1681	0.97
0.2	-0.2	-0.2	0.3669	0.3869	0.95	0.0903	0.0967	0.93	0.1963	0.1934	1.01

Table 2.3: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Normal distribution in case $\Pr(Y = 1) \simeq 60\%$, $n = 10000$ and R_i denote to the Ratio

$\sigma^2 = 0.1$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	-2.188	-2.185	1.00	0.3251	0.3296	0.99	0.4340	0.4290	1.01
0.2	0.2	0.4	-2.183	-2.183	1.00	0.3981	0.3971	1.00	0.5058	0.4962	1.02
0.7	0.8	0.7	-2.198	-2.198	1.00	0.6730	0.6730	1.00	0.7725	0.7729	0.99
0.8	0.7	0.9	-2.191	-2.190	1.00	0.6094	0.5967	1.02	0.6885	0.6961	0.99
0.1	-0.2	0.4	-2.182	-2.183	0.99	0.0820	0.0857	0.96	0.1802	0.1849	0.97
0.2	-0.2	-0.2	-2.193	-2.194	0.99	0.1043	0.0997	1.04	0.1989	0.1995	0.99
$\sigma^2 = 0.5$											
0.1	0.1	0.2	-2.131	-2.131	1.00	0.3292	0.3214	1.02	0.4261	0.4183	1.02
0.2	0.2	0.4	-2.112	-2.120	0.99	0.3953	0.3854	1.03	0.4842	0.4820	1.00
0.7	0.8	0.7	-2.191	-2.191	1.00	0.6812	0.6709	1.01	0.7664	0.7705	0.99
0.8	0.7	0.9	-2.141	-2.142	0.99	0.5848	0.5842	1.00	0.6851	0.6815	1.01
0.1	-0.2	0.4	-2.116	-2.120	0.99	0.0875	0.0832	1.05	0.1834	0.1796	1.02
0.2	-0.2	-0.2	-2.175	-2.175	1.00	0.1003	0.0988	1.01	0.1968	0.1977	0.99
$\sigma^2 = 1.5$											
0.1	0.1	0.2	-2.002	-2.011	0.99	0.3031	0.3033	0.99	0.4117	0.3947	1.04
0.2	0.2	0.4	-1.967	-1.983	0.99	0.3610	0.3610	1.00	0.4600	0.4507	1.02
0.7	0.8	0.7	-2.174	-2.175	0.99	0.6603	0.6659	0.99	0.7700	0.7648	1.01
0.8	0.7	0.9	-2.031	-2.038	0.99	0.5518	0.5560	0.99	0.6607	0.6486	1.02
0.1	-0.2	0.4	1.963	-1.985	0.99	0.0798	0.0779	1.02	0.1781	0.1681	1.06
0.2	-0.2	-0.2	-2.129	-2.128	1.00	0.0982	0.0967	1.02	0.2011	0.1934	1.04

Table 2.4: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Normal distribution in case $\Pr(Y = 1) \simeq 10\%$, $n = 10000$ and R_i denote to the Ratio

The parameter selection and correlation selection may have a slight effect in a few cases.

2.11 Simulation Study of Bivariate Normal Distribution

The previous simulation discussed the case of multivariate normal distribution, which found reasonable results in different cases with different correlation and variance. In this section we are going to apply simulation in case of the least false values with covariates from a bivariate normal distribution with a single fitted covariate, and assess the formulae which was computed and discussed in section 2.6. So, this simulation designed to examine the approximation form of the least false values and check on the behaviour of the MLE in this case.

2.11.1 Design of Simulation

To achieve the target of this simulation we will use the same assumption which used in previous simulation, but we consider some adjusted. As consider we have true logistic regression model has two covariates draw from bivariate normal distribution with mean zero and variance Σ

$$\pi_i = \text{expit}(\alpha + \beta_1 x_1 + \beta_2 x_2)$$

and we fitted the standard logistic regression model

$$\pi_i = \text{expit}(\alpha + \beta_1 x_1).$$

- Use different cases of variance $(\sigma_1^2, \sigma_2^2) : (0.1, 0.1), (0.1, 0.3), (0.6, 0.4)$ and different cases of correlation $\rho : (-0.6, -0.2, 0.1, 0.3, 0.8)$.

- We choose the parameters α, β_1 and β_2 to give two cases: choose $\beta_1 = 0.25$, $\beta_2 = 0.35$ and adjust $\alpha = -2.8, -0.3$ which make over the covariates $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$.

- Sample size $n = 10000$ and $N = 1000$ number of simulation.

2.11.2 Results and Discussion

This simulation designed to examine the behaviour of MLE and compute the least false values with bivariate normal covariates. The results showed the comparison between the parameters estimated values $\hat{\alpha}, \hat{\beta}_1$ and the least false values α^*, β_1^* also showed the ratio which denoted by R_1 and R_2. All cases of results with different variances and correlation reported in tables. Table 2.5, reported the simulation results of last false values using different values of ρ_{ij} and variance where x_1, x_2 have normal distribution and fit model with x_1 in case $\Pr(Y = 1) \simeq 10\%$. Table 2.6, reported the results of case $\Pr(Y = 1) \simeq 60\%$. We can see clearly that, the results appeared the same behaviour which found in the case of multivariate normal covariates. Some slightly differences which appeared when the correlation is negative $\rho = -0.2$, $\sigma_1^2 = 0.1$ and $\sigma_2^2 = 0.3$ in case of $\Pr(Y = 1) \simeq 10\%$. So, that is meaning the expression of the least false value which computed in case of the bivariate normal covariates works well at most cases of variance and correlation. However, the least false values appeared slightly sensitive in some cases by negative correlation when $\Pr(Y = 1) \simeq 10\%$.

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estemated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	-2.796	-2.796	1.00	0.039	0.039	0.99
-0.2	-2.795	-2.794	1.00	0.185	0.179	1.03
0.1	-2.796	-2.794	1.00	0.281	0.284	0.99
0.3	-2.797	-2.795	1.00	0.355	0.354	1.00
0.8	-2.799	-2.797	1.00	0.535	0.529	1.01
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	-2.790	-2.789	1.00	-0.112	-0.113	0.99
-0.2	-2.787	-2.783	1.00	0.135	0.127	1.05
0.1	-2.784	-2.782	1.00	0.307	0.308	0.99
0.3	-2.785	-2.784	1.00	0.428	0.429	0.99
0.8	-2.795	-2.794	1.00	0.733	0.733	1.00
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	-2.787	-2.785	1.00	0.077	0.078	0.99
-0.2	-2.780	-2.780	1.00	0.191	0.191	1.00
0.1	-2.780	-2.780	1.00	0.278	0.276	1.01
0.3	-2.780	-2.779	1.00	0.334	0.333	1.00
0.8	-2.794	-2.791	1.00	0.478	0.477	1.00

Table 2.5: Simulation results of last false values using different values of ρ_{ij} and variance where x_1, x_2 have normal distribution and fit model with x_1 in case $\Pr(Y = 1) \simeq 10\%$

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estemated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	-0.299	-0.299	1.00	0.041	0.039	1.03
-0.2	-0.298	-0.299	0.99	0.176	0.179	0.98
0.1	-0.299	-0.299	1.00	0.285	0.284	1.00
0.3	-0.299	-0.299	0.99	0.353	0.354	0.99
0.8	-0.300	-0.300	1.00	0.532	0.530	1.00
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	-0.298	-0.298	1.00	-0.114	-0.113	1.01
-0.2	-0.297	-0.298	0.99	0.129	0.127	1.01
0.1	-0.296	-0.298	0.99	0.308	0.309	0.99
0.3	-0.298	-0.298	1.00	0.429	0.429	1.00
0.8	-0.298	-0.299	0.99	0.732	0.733	0.99
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	-0.298	-0.298	1.00	0.077	0.078	0.99
-0.2	-0.296	-0.297	0.99	0.191	0.191	1.00
0.1	-0.296	-0.297	0.99	0.275	0.276	0.99
0.3	-0.297	-0.298	0.99	0.330	0.333	0.99
0.8	-0.298	-0.299	0.99	0.476	0.477	0.99

Table 2.6: Simulation results of last false values using different values of ρ_{ij} and variance where x_1, x_2 have normal distribution and fit model with x_1 in case $\Pr(Y = 1) \simeq 60\%$

2.12 Conclusion

Corresponding to the simulation analysis, we found a good result in all cases when the covariates are draw from the multivariate and bivariate normal distributions. The results appeared the MLE has reasonable behaviour with the least false values in case of missing covariates, which computed in terms of the true parameters. As we know, the normal distribution is symmetric distribution; also we made asymptotic normality distribution on covariates when compute the least false values. To examine the behaviour of the MLE and the formulae for least false values given in (2.14) and (2.15), we should apply and consider another symmetric distribution, such as t-multivariate distribution and uniform multivariate distribution. Moreover, we are interested to examine the behaviour of the least false values with covariates draw from distribution more skewed, say, log normal distribution. We will consider all this assumption and discuss it in the next chapter.

Chapter 3

Least false values for logistic regression model when covariate assumptions are violated

3.1 Introduction

The previous chapter discussed the least false values with multivariate covariates and bivariate covariates, the simulation providing us reasonable results. That was applied to covariates draw from multivariate and bivariate normal distribution. In this chapter we are interested to consider the model with symmetric distribution different from multivariate normal distribution. As we know, the behaviour of the MLE maybe affected by the assumption of normality on the covariates. So we will discuss in this chapter two of symmetric multivariate distribution, say, t-distribution and multivariate uniform distribution. Moreover, we are interested to examine the behaviour of the least false values when the covariates are skewed and we use lognormal distribution for this study.

3.2 Simulation of Multivariate t and Multivariate Uniform Distribution

The goal of this simulation is to use the same computed formulae of the last false value which used in the previous chapter, to assess the approximation computed for the least false values for logistic regression model and with multivariate t and uniform distribution.

3.2.1 Design of simulation

We use the same assumption which used in simulation in previous chapter. Let consider we have a true logistic regression model which has five covariates $p = 5$ is

$$\pi_i = \text{expit}(\alpha + \beta^T X)$$

where, $\beta^T = (\beta_1, \beta_2, \ldots, \beta_5)$, $X = (x_{i1}, \ldots, x_{i5})$ and the logistic regression model has two covariates has been fitted. We designed the simulation as follows:

- We choose X as a draw from one of two multivariate distribution; either

- Multivariate Uniform distribution, or

- Multivariate t-distribution.

- We are generating multivariate Uniform covariates by related with standard Normal distribution as:

- $Z \sim MVN(0, R)$ where R is the correlation matrix.

- $U = \Phi(Z) \to [0, 1]$, (element wise).

- $X_U \sim 5\sigma(U - \frac{1}{2}) \to [-2\frac{1}{2}\sigma, 2\frac{1}{2}\sigma]$.

- We consider the 5×5 covariance matrix Ω is

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.$$

As we know, the mean of the uniform distribution is $U = 1/2$ and the variance is $\mathsf{var}(U) = 1/12$. So, in this case we have $\mathsf{var}(X_U) = 25/12$ and $\mathsf{cov}(X_{Ui}, X_{Uj})$ =25 $\mathsf{cov}(U_i, U_j)$, where the covariance is

$$\mathsf{cov}(U_i, U_j) = \frac{\arcsin(\frac{\rho_{ij}}{2})}{2\pi}.$$

Then, the components of covariance matrix Ω are

$$\Omega_{11} = 25 \begin{bmatrix} \frac{1}{12} & \mathsf{cov}(U_1, U_2) \\ \mathsf{cov}(U_2, U_1) & \frac{1}{12} \end{bmatrix},$$

$$\Omega_{22} = 25 \begin{bmatrix} \frac{1}{12} & \mathsf{cov}(U_3, U_4) & \mathsf{cov}(U_3, U_5) \\ \mathsf{cov}(U_4, U_3) & \frac{1}{12} & \mathsf{cov}(U_5, U_4) \\ \mathsf{cov}(U_5, U_3) & \mathsf{cov}(U_5, U_4) & \frac{1}{12} \end{bmatrix},$$

$$\Omega_{21} = 25 \begin{bmatrix} \mathsf{cov}(U_3, U_1) & \mathsf{cov}(U_3, U_2) \\ \mathsf{cov}(U_4, U_1) & \mathsf{cov}(U_4, U_2) \\ \mathsf{cov}(U_5, U_1) & \mathsf{cov}(U_5, U_2) \end{bmatrix}, \Omega_{21}^T = \Omega_{12}.$$

- We consider 6 different cases of correlation which is each case of Ω_{ij} has same ρ_{ij} designed as:(0.1,0.1,0.2), (0.2,0.2,0.4), (0.7,0.8,0.7), (0.8,0.7,0.9), (0.1,-0.2,0.4), (0.2,-0.2,-0.2). Values are chosen to assume Ω is positive definite.

- We generating multivariate t-distribution with various value of degrees of freedom df, which changes the shape of the distribution, we choose three cases $df = (5, 10, 200)$ and use variance $\sigma^2 = 0.5$ in each case.

- Use the same assumption on different cases of correlation and variance, also use the same assumption on chose the true parameters as $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$ which used in the simulation of multivariate normal distribution in the previous chapter.

- Large sample size has been used $n = 500$, $n = 10000$ and $N = 1000$ number of simulation.

3.2.2 Results and Discussion

The results concerning two simulation data generated by multivariate Uniform distribution and multivariate t-distribution. The results of this simulation with Uniform distribution, showed in Table 3.1 and Table 3.2, in cases of $\Pr(Y = 1) \simeq 60\%$ and $\Pr(Y = 1) \simeq 10\%$ respectively with two sample size $n = 500$, $n = 10000$. The same results appeared, the ratio found nearly close to one in almost cases. A few cases appeared low ratio in case of sample size $n = 500$, which there are some estimated value were very small (i.e , when $\Omega_{11} = 0.1, \Omega_{12} = -0.2, \Omega_{22} = 0.4$ the parameter estimated was $\beta_1 = 0.0809, \beta_1^* = 0.0639$ and the ratio was $R_2 = 0.79$). In general we found the least false values in this case have the same behaviour of the multivariate normal covariates.

The results of the second part of this simulation, concerning for results of data generated by multivariate t-distribution which showed in Table 3.3 and Table 3.4 in cases of $\Pr(Y = 1) \simeq 60\%$ and $\Pr(Y = 1) \simeq 10\%$ respectively with sample size $n = 500$. Table 3.5 and Table 3.6 shows the results in case of sample size $n = 10000$. The results of four cases with different degree of freedom $df = 200, 10, 5$ and one case of variance has been used $\sigma^2 = 0.5$. Comparing these results with case of Normal distribution, more clearly when the degree of freedom larger enough we can reported that the results have the same behaviour. Moreover, we can say that the ratio appeared nearly close to one in all cases of correlation and degree of freedom, some slightly differences with low ratio appeared in few cases when degree of freedom

$n = 500$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	0.3831	0.3548	1.07	0.2831	0.2913	0.97	0.3932	0.3800	1.03
0.2	0.2	0.4	0.3275	0.3485	0.94	0.3433	0.3437	1.00	0.4454	0.4308	1.03
0.7	0.8	0.7	0.4161	0.3929	1.05	0.6727	0.6584	1.02	0.7491	0.7566	0.99
0.8	0.7	0.9	0.3294	0.3594	0.92	0.5420	0.5339	1.01	0.5640	0.6238	0.90
0.1	-0.2	0.4	0.3557	0.3489	0.98	0.0639	0.0809	0.79	0.1556	0.1682	0.93
0.2	-0.2	-0.2	0.3971	0.3811	0.93	0.1142	0.1005	1.13	0.1825	0.3811	0.93
$n = 10000$											
0.1	0.1	0.2	0.3344	0.3548	0.94	0.2779	0.2913	0.95	0.3648	0.3800	0.96
0.2	0.2	0.4	0.3309	0.3485	0.95	0.3299	0.3437	0.96	0.4129	0.4308	0.96
0.7	0.8	0.7	0.3921	0.3929	1.00	0.6562	0.6584	0.99	0.7563	0.7566	0.99
0.8	0.7	0.9	0.3350	0.3594	0.93	0.5262	0.5339	0.99	0.6194	0.6238	0.99
0.1	-0.2	0.4	0.3186	0.3489	0.91	0.0766	0.0809	0.95	0.1463	0.1682	0.87
0.2	-0.2	-0.2	0.3830	0.3811	1.00	0.0971	0.1005	0.97	0.1935	0.1957	0.99

Table 3.1: Simulation results of last false values using different values of ρ_{ij} by generated variables from multivariate Uniform distribution in case $\Pr(Y = 1) \simeq 60\%$, $n = 500$, $n = 10000$ and R_i denote to the Ratio

$n = 500$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	-2.015	-1.951	1.03	0.2864	0.2913	0.98	0.3977	0.3800	1.04
0.2	0.2	0.4	-1.901	-1.916	0.99	0.6911	0.3437	1.04	0.7496	0.4308	1.03
0.7	0.8	0.7	-2.182	-2.161	1.01	0.6911	0.6584	1.04	0.7496	0.7566	0.99
0.8	0.7	0.9	-2.009	-1.977	1.01	0.5728	0.5339	1.07	0.5876	0.6238	0.94
0.1	-0.2	0.4	-2.109	-2.096	1.01	0.1333	0.1005	1.32	0.1925	0.1957	0.98
0.2	-0.2	-0.2	-1.927	-1.919	1.00	0.0892	0.0809	1.10	0.1871	0.1682	1.11
$n = 10000$											
0.1	0.1	0.2	-1.942	-1.951	0.99	0.2986	0.2913	1.02	0.3861	0.3800	1.01
0.2	0.2	0.4	-1.900	-1.916	0.99	0.3551	0.3437	1.03	0.4483	0.4308	1.04
0.7	0.8	0.7	-2.168	-2.161	1.00	0.6586	0.6584	1.00	0.7649	0.7566	1.01
0.8	0.7	0.9	-1.963	-1.977	0.99	0.5217	0.5339	0.98	0.6557	0.6238	1.05
0.1	-0.2	0.4	-1.883	-1.919	0.98	0.0794	0.0809	0.98	0.1885	0.1682	1.12
0.2	-0.2	-0.2	-2.089	-2.096	0.99	0.1022	0.1005	1.01	0.1938	0.1957	0.99

Table 3.2: Simulation results of last false values using different values of ρ_{ij} by generated variables from multivariate Uniform distribution in case $\Pr(Y = 1) \simeq 10\%$, $n = 500$,$n = 10000$ and R_i denote to the Ratio

is $df = 5$ and $n = 500$, which have the same behaviour found in case of the normal multivariate covariates when the estimated value was very small.

Overall, if we assume normality on covariates, but the covariates are drawn from a multivariate t-distribution with variety of degree of freedom and multivariate Uniform distribution, which use large sample size $n = 10000$. We found that, for different combination of correlations and variances, are appeared the results from (2.14) and (2.15) still appear to hold.

$df = 200$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	0.380	0.387	0.98	0.382	0.321	1.19	0.437	0.418	1.04
0.2	0.2	0.4	0.389	0.385	1.01	0.409	0.385	1.06	0.422	0.481	0.88
0.7	0.8	0.7	0.390	0.398	0.98	0.676	0.670	1.01	0.801	0.770	1.03
0.8	0.7	0.9	0.393	0.389	1.01	0.549	0.584	0.94	0.682	0.681	1.00
0.1	-0.2	0.4	0.382	0.385	0.99	0.107	0.083	1.29	0.165	0.179	0.92
0.2	-0.2	-0.2	0.395	0.395	1.00	0.098	0.099	0.99	0.174	0.197	0.88
$df = 10$											
0.1	0.1	0.2	0.376	0.387	0.97	0.325	0.321	1.01	0.426	0.418	1.01
0.2	0.2	0.4	0.393	0.385	1.02	0.352	0.385	0.91	0.514	0.481	1.06
0.7	0.8	0.7	0.392	0.398	0.99	0.676	0.670	1.01	0.783	0.770	1.01
0.8	0.7	0.9	0.375	0.389	0.96	0.604	0.584	1.03	0.606	0.681	0.89
0.1	-0.2	0.4	0.399	0.385	1.03	0.072	0.083	0.87	0.194	0.179	1.08
0.2	-0.2	-0.2	0.376	0.395	0.95	0.141	0.098	1.43	0.175	0.197	0.89
$df = 5$											
0.1	0.1	0.2	0.364	0.387	0.94	0.390	0.321	1.21	0.359	0.418	0.86
0.2	0.2	0.4	0.402	0.385	1.04	0.332	0.385	0.86	0.466	0.481	0.97
0.7	0.8	0.7	0.400	0.398	1.00	0.635	0.670	0.95	0.806	0.770	1.04
0.8	0.7	0.9	0.375	0.389	0.96	0.564	0.584	0.97	0.722	0.681	1.06
0.1	-0.2	0.4	0.352	0.385	0.91	0.086	0.083	1.03	0.147	0.179	0.82
0.2	-0.2	-0.2	0.379	0.395	0.96	0.086	0.098	0.88	0.186	0.198	0.94

Table 3.3: Simulation results of last false values using different values of ρ_{ij} and $\sigma^2 = 0.5$ by generated variables from multivariate t-distribution in case $\Pr(Y = 1) \simeq 60\%$, $n = 500$ and R_i denote to the Ratio

$df = 200$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	-2.160	-2.131	1.01	0.313	0.321	0.98	0.425	0.418	1.01
0.2	0.2	0.4	-2.100	-2.120	0.99	0.369	0.385	0.96	0.529	0.482	1.09
0.7	0.8	0.7	-2.224	-2.191	1.01	0.692	0.670	1.03	0.789	0.770	1.02
0.8	0.7	0.9	-2.165	-2.142	1.01	0.575	0.584	0.99	0.738	0.681	1.08
0.1	-0.2	0.4	-2.140	-2.120	1.01	0.061	0.083	0.74	0.239	0.179	1.33
0.2	-0.2	-0.2	-2.212	-2.175	1.01	0.110	0.098	1.11	0.200	0.197	1.01
$df = 10$											
0.1	0.1	0.2	-2.129	-2.131	0.99	0.338	0.321	1.05	0.472	0.418	1.13
0.2	0.2	0.4	-2.118	-2.120	0.99	0.376	0.385	0.98	0.485	0.482	1.01
0.7	0.8	0.7	-2.227	-2.191	1.01	0.660	0.670	0.98	0.799	0.770	1.03
0.8	0.7	0.9	-2.176	-2.142	1.01	0.618	0.584	1.05	0.589	0.681	0.86
0.1	-0.2	0.4	-2.105	-2.120	0.99	0.089	0.083	1.06	0.171	0.179	0.96
0.2	-0.2	-0.2	-2.160	-2.175	0.99	0.073	0.098	0.74	0.258	0.197	1.30
$df = 5$											
0.1	0.1	0.2	-2.097	-2.131	0.98	0.299	0.321	0.93	0.429	0.418	1.02
0.2	0.2	0.4	-2.118	-2.120	0.99	0.430	0.385	1.11	0.498	0.482	1.03
0.7	0.8	0.7	-2.225	-2.191	1.01	0.669	0.671	0.99	0.779	0.771	1.01
0.8	0.7	0.9	-2.114	-2.142	0.99	0.709	0.584	1.21	0.550	0.682	0.81
0.1	-0.2	0.4	-2.021	-2.12	0.95	0.068	0.083	0.83	0.200	0.179	1.11
0.2	-0.2	-0.2	-2.180	-2.175	1.00	0.096	0.098	0.97	0.234	0.197	1.18

Table 3.4: Simulation results of last false values using different values of ρ_{ij} and $\sigma^2 = 0.5$ by generated variables from multivariate t-distribution in case $\Pr(Y = 1) \simeq 10\%$, $n = 500$ and R_i denote to the Ratio

$df = 200$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	0.381	0.387	0.99	0.325	0.321	1.01	0.425	0.418	1.02
0.2	0.2	0.4	0.381	0.385	0.99	0.394	0.385	1.02	0.478	0.481	0.99
0.7	0.8	0.7	0.399	0.398	1.00	0.667	0.670	0.99	0.775	0.770	1.01
0.8	0.7	0.9	0.390	0.389	1.00	0.570	0.584	0.98	0.688	0.681	1.01
0.1	-0.2	0.4	0.375	0.385	0.97	0.075	0.083	0.91	0.160	0.179	0.90
0.2	-0.2	-0.2	0.394	0.395	0.99	0.103	0.098	1.05	0.187	0.197	0.95
$df = 10$											
0.1	0.1	0.2	0.373	0.387	0.96	0.317	0.321	0.99	0.413	0.418	0.99
0.2	0.2	0.4	0.376	0.385	0.98	0.368	0.385	0.96	0.463	0.481	0.96
0.7	0.8	0.7	0.394	0.398	0.99	0.664	0.670	0.99	0.773	0.770	1.00
0.8	0.7	0.9	0.385	0.389	0.99	0.553	0.584	0.95	0.676	0.681	0.99
0.1	-0.2	0.4	0.371	0.385	0.96	0.077	0.083	0.92	0.175	0.179	0.98
0.2	-0.2	-0.2	0.395	0.395	1.00	0.097	0.098	0.99	0.195	0.197	0.99
$df = 5$											
0.1	0.1	0.2	0.379	0.387	0.98	0.299	0.321	0.93	0.386	0.418	0.92
0.2	0.2	0.4	0.380	0.385	0.99	0.357	0.385	0.93	0.458	0.481	0.95
0.7	0.8	0.7	0.398	0.398	1.00	0.666	0.671	0.99	0.771	0.771	1.00
0.8	0.7	0.9	0.380	0.389	0.98	0.557	0.584	0.96	0.670	0.681	0.98
0.1	-0.2	0.4	0.371	0.385	0.96	0.068	0.083	0.82	0.167	0.179	0.93
0.2	-0.2	-0.2	0.391	0.395	0.99	0.093	0.098	0.94	0.185	0.198	0.94

Table 3.5: Simulation results of last false values using different values of ρ_{ij} and $\sigma^2 = 0.5$ by generated variables from multivariate t-distribution in case $\Pr(Y = 1) \simeq 60\%$, $n = 10000$ and R_i denote to the Ratio

$df = 200$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.1	0.2	-2.124	-2.131	0.99	0.330	0.321	1.03	0.419	0.418	1.00
0.2	0.2	0.4	-2.119	-2.121	0.99	0.395	0.385	1.03	0.504	0.482	1.05
0.7	0.8	0.7	-2.196	-2.191	1.00	0.672	0.670	1.00	0.770	0.770	1.00
0.8	0.7	0.9	-2.141	-2.142	0.99	0.589	0.584	1.01	0.676	0.681	0.99
0.1	-0.2	0.4	-2.121	-2.120	1.00	0.082	0.083	0.99	0.172	0.179	0.96
0.2	-0.2	-0.2	-2.181	-2.180	1.00	0.099	0.0988	1.01	0.205	0.197	1.03
$df = 10$											
0.1	0.1	0.2	-2.116	-2.131	0.99	0.328	0.321	1.02	0.414	0.418	0.99
0.2	0.2	0.4	-2.095	-2.120	0.99	0.393	0.385	1.02	0.484	0.482	1.01
0.7	0.8	0.7	-2.191	-2.191	1.00	0.672	0.670	1.00	0.771	0.770	1.00
0.8	0.7	0.9	-2.132	-2.142	0.99	0.571	0.584	0.98	0.686	0.682	1.01
0.1	-0.2	0.4	-2.096	-2.120	0.99	0.093	0.083	1.10	0.196	0.179	1.10
0.2	-0.2	-0.2	-2.160	-2.175	0.99	0.098	0.098	1.00	0.198	0.197	1.01
$df = 5$											
0.1	0.1	0.2	-2.104	-2.131	0.99	0.322	0.321	1.00	0.418	0.418	1.00
0.2	0.2	0.4	-2.079	-2.120	0.98	0.379	0.385	0.98	0.466	0.482	0.97
0.7	0.8	0.7	-2.188	-2.191	0.99	0.674	0.671	1.00	0.766	0.771	0.99
0.8	0.7	0.9	-2.119	-2.142	0.99	0.552	0.584	0.95	0.688	0.682	1.01
0.1	-0.2	0.4	-2.07	-2.12	0.98	0.082	0.083	0.99	0.176	0.179	0.98
0.2	-0.2	-0.2	-2.150	-2.175	0.99	0.109	0.098	1.10	0.186	0.198	0.94

Table 3.6: Simulation results of last false values using different values of ρ_{ij} and $\sigma^2 = 0.5$ by generated variables from multivariate t-distribution in case $\Pr(Y = 1) \simeq 10\%$, $n = 10000$ and R_i denote to the Ratio

3.3 Least False Values With Log Normal Covariates

3.3.1 Introduction

The previous section discussed the least false values for logistic regression model with multivariate covariates draw from t-distribution and Uniform distribution. As we know that, the Normal, t and Uniform distributions have symmetric shape. In this section, we are interested to investigate the behaviour of the least false value under missing covariates when the covariates have more skewed distribution, say, log normal. To examine this target, let us consider we have two covariates X_1, X_2 draw from bivariate normal distribution $X \sim (\mu, \Sigma)$, where $\mu^T = (\mu_1, \mu_2)$ and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

We will consider three cases, as follows: $(X_1, \exp(X_2))$, $(\exp(X_1), X_2)$ and $(\exp(X_1), \exp(X_2))$. These cases contain one of covariates or both of them are log normal. We know that, the expression of the least false value in this case dependent upon correlation between of the two covariates need to compute, and we need to use the variance and mean corresponding to log normal covariates. So, in each case compute the least false values using the appropriate dispersion matrix for the covariates. The calculations, are shown in the following section.

3.3.2 Least False Value in case of covariates $(X_1, \exp(X_2))$

In this case the first covariate is normal distribution with mean μ_1 and variance $\mathsf{var}(X_1) = \sigma_1^2$. The second covariate distributed lognormal distribution with mean $\exp(\mu_2 + \frac{1}{2}\sigma_2^2)$ and variance $(\exp(\sigma_2^2) - 1)\exp(2\mu_2 + \sigma_2^2)$. Now work to find the correlation between the two covariates $(X_1, \exp(X_2))$, we consider the correlation is

$$\rho_{X_1,\exp(X_2)} = \mathsf{cor}(X_1, \exp(X_2) = \frac{\mathsf{cov}(X_1, \exp(X_2))}{\sigma_1\sqrt{\mathsf{var}(\exp(X_2))}}$$

$$= \frac{E(X_1\exp(X_2)) - E(X_1)E(\exp(X_2))}{\sigma_1\sqrt{\mathsf{var}(\exp(X_2))}}.$$

It seems clear to compute $\rho_{X_1,\exp(X_2)}$, just we need work out to find $E(X_1\exp(X_2))$. We have the moment generating function of the bivariate normal distribution, Which written as

$$M(t_1,t_2) = E(\exp(t_1X_1 + t_2X_2)) = \exp(t_1\mu_1 + t_2\mu_2 + \tfrac{1}{2}t^T\Sigma t)$$
$$= \exp(t_1\mu_1 + t_2\mu_2 + \tfrac{1}{2}\sigma_1^2t_1^2 + \tfrac{1}{2}\sigma_2^2t_2^2 + \rho\sigma_1\sigma_2t_1t_2).$$

Now, differentiate $M(t_1,t_2)$ with respect for t_1, $D(t_1,t_2)$, say, we get

$$D(t_1,t_2) = \frac{\partial M(t_1,t_2)}{\partial t_1} = E(X_1\exp(t_1X_1 + t_2X_2))$$

and if put $t_1 = 0$ and $t_2 = 1$, we get $D(0,1) = E(X_1\exp(X_2)$, which we need compute, then

$$D(t_1,t_2) = (\mu_1 + \sigma_1^2t_1 + \rho\sigma_1\sigma_2t_2)M(t_1,t_2)$$

and we get

$$E(X_1 \mid \exp(x_2)) = D(0,1) = (\mu_1 + \rho\sigma_1\sigma_2)\exp(\mu_2 + \tfrac{1}{2}\sigma_2^2)$$

and then,

$$\mathsf{cov}(X_1,\exp(X_2) = (\mu_1+\rho\sigma_1\sigma_2)\exp(\mu_2+\tfrac{1}{2}\sigma_2^2) - \mu_1\exp(\mu_2+\tfrac{1}{2}\sigma_2^2) = \rho\sigma_1\sigma_2\exp(\mu_2+\tfrac{1}{2}\sigma_2^2),$$

Finally, the correlation is

$$\mathsf{cor}(X_1,\exp(X_2)) = \frac{\rho\sigma_2\exp(\mu_2 + \frac{1}{2}\sigma_2^2)}{\sqrt{\exp(2\mu_2 + \sigma_2^2)(\exp(\sigma_2^2) - 1)}}$$

$$\rho_{X_1,\exp(X_2)} = \mathsf{cor}(X_1,\exp(X_2)) = \frac{\rho\sigma_2}{\sqrt{\exp(\sigma_2^2) - 1}}.$$

So, the least false values in this case are

$$\alpha^* = \frac{\alpha + \beta_2(\mu_2 - \rho_{X_1,\exp(X_2)}\frac{\mathsf{var}(\exp(X_2))}{\sigma_1}\mu_1)}{\sqrt{1 + k^2\beta_2^2\mathsf{var}(\exp(X_2))(1 - \rho^2_{X_1,\exp(X_2)})}},$$

and

$$\beta_1^* = \frac{\beta_1 + \rho_{X_1,\exp(X_2)}\frac{\mathsf{var}(\exp(X_2))}{\sigma_1}\beta_2}{\sqrt{1 + k^2\beta_2^2\mathsf{var}(\exp(X_2))(1 - \rho^2_{X_1,\exp(X_2)})}}.$$

3.3.3 Least False Value in case of covariates $(\exp(X_1), X_2)$

In this case the first covariate is distributed lognormal distribution and the second covariate is normal, we will using the same steps which have used in previous section, to find the correlation $\mathsf{cor}(\exp(X_1), X_2)$, and compute the least false values. so, we have differentiate $M(t_1, t_2)$ in terms of t_2 and we get, by analogy with the previous case,

$$\rho_{\exp(X_1),X_2} = \mathsf{cor}(\exp(X_1), X_2) = \frac{\rho\sigma_1}{\sqrt{\exp(\sigma_1^2) - 1}}.$$

Then, the least false values are

$$\alpha^* = \frac{\alpha + \beta_2(\mu_2 - \rho_{\exp(X_1),X_2}\frac{\sigma_2}{\mathsf{var}(\exp(X_1))}\mu_1)}{\sqrt{1 + k^2\beta_2^2\sigma_2^2(1 - \rho^2_{\exp(X_1),X_2})}},$$

and

$$\beta_1^* = \frac{\beta_1 + \rho_{\exp(X_1),X_2}\frac{\sigma_2}{\mathsf{var}(\exp(X_1))}\beta_2}{\sqrt{1 + k^2\beta_2^2\sigma_2^2(1 - \rho^2_{\exp(X_1),X_2})}}.$$

3.3.4 Least False Value in case of covariates $(\exp(X_1), \exp(X_1))$

The final case, we have two covariates that are log normal, as the same steps which used before, in this case we need to compute $E(\exp(X_1)\exp(X_2))$ which can find it by put $t_1 = 1, t_2 = 1$ in the moment generating function $M(t_1, t_2)$. So,

$$E(\exp(X_1)\exp(X_2)) = M(1,1) = \exp(\mu_1 + \mu_2 + \tfrac{1}{2}\sigma_1^2 + \tfrac{1}{2}\sigma_2^2 + \rho\sigma_1\sigma_2)$$

and the covariance is

$$\mathsf{cov}(\exp(X_1)\exp(X_2)) = \exp(\mu_1 + \mu_2 + \tfrac{1}{2}\sigma_1^2 + \tfrac{1}{2}\sigma_2^2)(\exp(\rho\sigma_1\sigma_2) - 1),$$

finally the correlation is

$$\rho_{\exp(X_1),\exp(X_2)} = \mathsf{cor}(\exp(X_1), \exp(X_2)) = \frac{\exp(\rho\sigma_1\sigma_2) - 1}{\sqrt{(\exp(\sigma_1^2) - 1)(\exp(\sigma_2^2) - 1)}}.$$

The least false values in this case are

$$\alpha^* = \frac{\alpha + \beta_2(\mu_2 - \rho_{\exp(X_1),\exp(X_1)}\frac{\mathsf{var}(\exp(X_2))}{\mathsf{var}(\exp(X_1))}\mu_1)}{\sqrt{1 + k^2\beta_2^2\mathsf{var}(\exp(X_2))(1 - \rho^2_{\exp(X_1),\exp(X_2)})}},$$

and

$$\beta_1^* = \frac{\beta_1 + \rho_{(\exp(X_1),\exp(X_2))}\frac{\mathsf{var}(\exp(X_2))}{\mathsf{var}(\exp(X_1))}\beta_2}{\sqrt{1 + k^2\beta_2^2\mathsf{var}(\exp(X_2))(1 - \rho^2_{\exp(X_1),\exp(X_2)})}}.$$

Finally, we have different correlation which need to use in the expression of the least false values α^* and β^* related in each case, we need to examine it by simulation.

3.4 Simulation Study of Log-Normal Distribution

This part of simulation is to assess the least false values form when the covariates have two covariates draw from bivariate normal distribution and log normal distribution. This example to application on three cases of covariates $(X_1, \exp(X_2))$, $(\exp(X_1), X_2)$ and $(\exp(X_1), \exp(X_2))$, check the behaviour of the approximation least false values in different cases of covariates, correlation and variance.

3.4.1 Design of Simulation

To get to the goal, let us consider we have true logistic regression model has two covariates $p = 2$ is

$$\pi_i = \text{expit}(\alpha + \beta_1 X_1 + \beta_2 X_2)$$

and the standard logistic regression model has been fitted

$$\pi_i = \text{expit}(\alpha + \beta_1 X_1)$$

- Choose X as a draw from bivariate Normal distribution $X \sim (0, \Sigma)$.
- Apply on three cases of covariates:

- Mixed covariates $(X_1, \exp(X_2))$.
- Mixed covariates $(\exp(X_1), X_2)$.
- Log normal covariates $(\exp(X_1), \exp(X_2))$.

- Use four cases of combinations of variance (σ_1^2, σ_2^2) :(0.1,0.1), (0.1,0.3), (0.5, 0.2) and (0.6,0.4).
- Use five cases of correlation ρ in each case of variance (-0.6, -0.2, 0.1, 0.3, 0.8).
- We choose the parameters α, β_1 and β_2 to give two cases: choose $\beta_1 = 0.25$, $\beta_2 = 0.35$ and adjust $\alpha = -2.8, -0.3$ which make over the covariates $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$.
- Sample size $n = 10000$ and $N = 1000$ number of simulation.

3.4.2 Results and Discussion

The goal of the simulation is to compute the least false values and compare the results with computed formulae. The results of the three cases reported on tables, which shows the comparison between the estimated parameters by fitted standard logistic model and the computed least false values. The same notation which used in simulation of bivariate normal distribution in previous chapter, α, β_1 denoted to parameters estimated, α^*, β_1^* denoted to computed least false values and R_1, R_2 are the ratios.

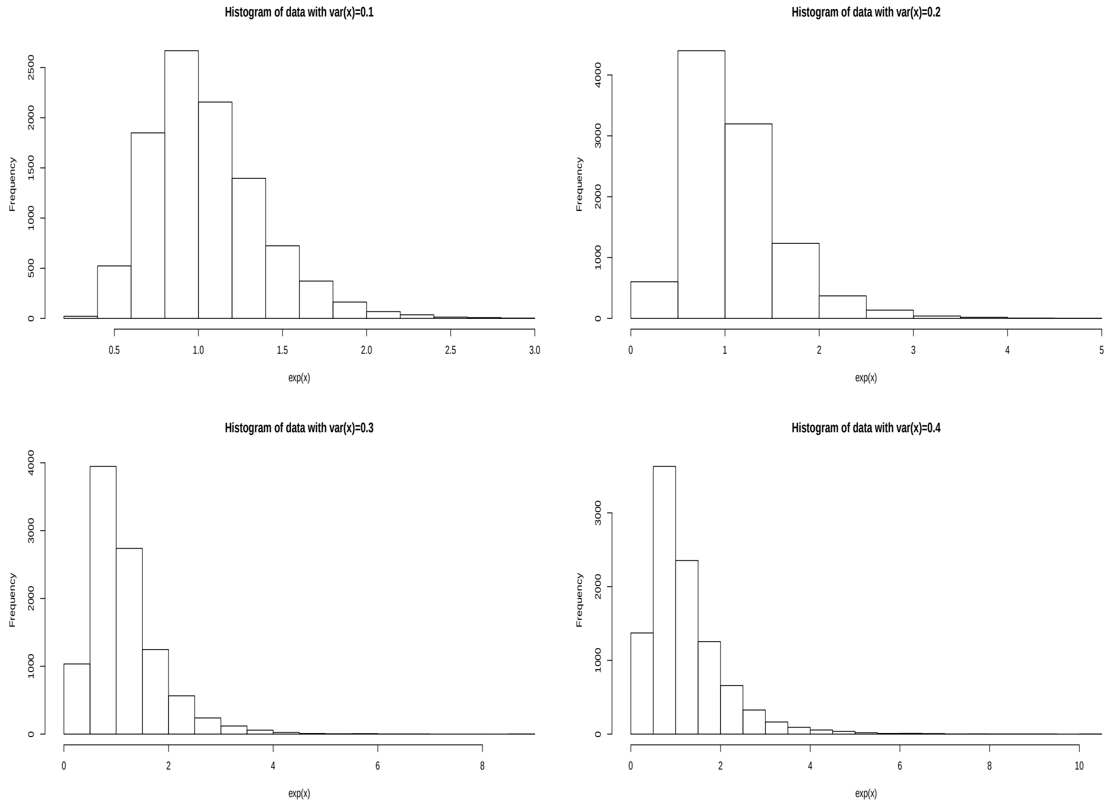

Figure 3.1: Plot of the histogram of a log normal covariate with different variance.

Figure (3.1), shows the histogram of the log normal data with different value of variance, which explains the skew shape affected by the value of variance. So, in case of σ_2^2=(0.1, 0.2, 0.3 and 0.4) the skewness equal (1.01, 1.52, 2.45 and 2.98) respectively.

3.4.3 Results of Case $(X_1, \exp(X_2))$

In this part we will show the result of case when the true model has two covariates, in this case the true model is

$$\pi_i = \text{expit}(\alpha + \beta_1 X_1 + \beta_2 \exp(X_2))$$

and the standard model has been fitted

$$\pi_i = \text{expit}(\alpha + \beta_1 X_1).$$

Table 3.7 and Table 3.8, reported the results in case of $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$ respectively. We can see clearly that, the results broadly show the same behaviour which was found in the case of multivariate normal covariates. Some slightly different which appeared in some cases by low ratio $R_2 = (1.25, 0.86, 0.40, 0.93)$ when the correlation is negative in case of $\Pr(Y = 1) \simeq 10\%$. These results corresponding to the values of coefficients which is appeared very small and close to zero which affect on the value of ratios. However, the ratio R_1 corresponding to the estimate of parameter α show close to one in all cases of different combination of variances and correlations. The results in case of $\Pr(Y = 1) \simeq 60\%$ show regular ratios, in all cases the ratio close to one.

3.4.4 Results of Case $(\exp(X_1), X_2)$

This discussion for the results of case when the true model is

$$\pi_i = \text{expit}(\alpha + \beta_1 \exp(X_1) + \beta_2 X_2)$$

and the standard model has been fitted

$$\pi_i = \text{expit}(\alpha + \beta_1 \exp(X_1)).$$

Table 3.9 and Table 3.10, shows the result of the second case of covariates $(\exp(X_1), X_2)$, in case $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$ respectively. We fitted the logistic regression model has log normal covariate $\exp(X_1)$. We can see, the approximation form of the least false value works well in all cases of combinations with difference variance and correlation. Moreover, the same behaviour results appeared in both cases of $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$. Reasonable ratio found in all cases, one case seems slightly low ratio $R_2 = 0.92$ in case of negative correlation, because may be the least false value and estimated parameters appeared a small value close to zero (0.055, 0.059), which maybe affect on the ratio value.

3.4.5 Results of Case $(\exp(X_1), \exp(X_1))$

The final part of results for the case when the true model is

$$\pi_i = \text{expit}(\alpha + \beta_1 \exp(X_1) + \beta_2 \exp(X_2))$$

and the standard model has been fitted

$$\pi_i = \text{expit}(\alpha + \beta_1 \exp(X_1)).$$

Table 3.11 and Table 3.12, shows the results of the case of $(\exp(X_1), \exp(X_2))$, in case $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$ respectively. The results appeared a good response in all cases and the ratio as well, which appeared close to one in all cases. Same behaviour results found in the two case of $\Pr(Y = 1) \simeq 10\%$ and $\Pr(Y = 1) \simeq 60\%$. However, some low ratio in case of negative correlation $R_2 = (0.85, 1.36, 0.91)$ this results maybe related to the same reason which discussed before, a small values of coefficient estimated.

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estimated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	-2.428	-2.428	1.00	0.027	0.029	0.96
-0.2	-2.426	-2.426	1.00	0.177	0.176	1.01
0.1	-2.428	-2.427	1.00	0.289	0.286	1.01
0.3	-2.427	-2.427	1.00	0.365	0.360	1.01
0.8	-2.430	-2.430	1.00	0.566	0.543	1.04
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	-2.377	-2.377	1.00	-0.215	-0.171	1.25
-0.2	-2.369	-2.370	0.99	0.093	0.108	0.86
0.1	-2.367	-2.369	0.99	0.325	0.317	1.02
0.3	-2.369	-2.371	0.99	0.492	0.457	1.07
0.8	-2.388	-2.382	1.00	0.902	0.809	1.11
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	-2.347	-2.347	1.00	0.016	0.040	0.40
-0.2	-2.332	-2.337	0.99	0.165	0.177	0.93
0.1	-2.333	-2.336	0.99	0.285	0.280	1.01
0.3	-2.336	-2.339	0.99	0.372	0.349	1.06
0.8	-2.368	-2.360	1.01	0.597	0.525	1.13

Table 3.7: Simulation results of last false values using different values of ρ_{ij} and variance when the model has $(X_1, \exp(X_2))$ covariates in case $\Pr(Y = 1) \simeq 10\%$

Overall, the simulation designed to investigate the behaviour of the least false value for logistic regression model with binary normal covariates and one of two covariates or both are skew distribution. Although we used different value of variance which is significant in determining the skewness of distribution, we have got reasonable results. However, there is slightly effect which appeared in low ratio in a few cases with negative correlation and $\Pr(Y = 1) \simeq 10\%$.

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estimated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	0.068	0.067	1.01	0.030	0.029	1.03
-0.2	0.067	0.068	0.99	0.288	0.286	1.01
0.1	0.067	0.068	0.99	0.542	0.543	0.99
0.3	0.067	0.068	0.99	0.362	0.359	1.01
0.8	0.067	0.068	0.99	0.542	0.543	0.99
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	0.104	0.105	0.99	-0.163	-0.171	0.95
-0.2	0.103	0.106	0.98	0.111	0.108	1.03
0.1	0.104	0.106	0.98	0.314	0.317	0.99
0.3	0.103	0.106	0.98	0.451	0.457	0.99
0.8	0.103	0.106	0.97	0.794	0.809	0.98
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	0.123	0.126	0.98	0.043	0.040	1.08
-0.2	0.121	0.125	0.97	0.179	0.177	1.01
0.1	0.121	0.125	0.96	0.277	0.280	0.99
0.3	0.122	0.126	0.97	0.343	0.349	0.98
0.8	0.119	0.126	0.94	0.511	0.525	0.97

Table 3.8: Simulation results of last false values using different values of ρ_{ij} and variance when the model has $(X_1, \exp(X_2))$ covariates in case $\Pr(Y = 1) \simeq 60\%$

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estimated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	-2.601	-2.601	1.00	0.061	0.059	1.03
-0.2	- -2.728	-2.728	1.00	0.186	0.187	0.99
0.1	-2.827	-2.827	1.00	0.280	0.281	0.99
0.3	-2.892	-2.894	0.99	0.342	0.344	0.99
0.8	-3.056	-3.063	0.99	0.493	0.502	0.98
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	-2.445	-2.444	1.00	-0.080	-0.078	1.02
-0.2	-2.670	-2.670	1.00	0.139	0.139	1.00
0.1	-2.839	-2.840	0.99	0.300	0.302	0.99
0.3	-2.948	-2.955	0.98	0.403	0.412	0.98
0.8	-3.230	-3.253	0.99	0.665	0.686	0.97
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	-2.682	-2.658	1.01	0.171	0.156	1.09
-0.2	-2.746	-2.741	1.01	0.223	0.217	1.02
0.1	-2.793	-2.797	0.99	0.261	0.263	0.99
0.3	-2.826	-2.840	0.99	0.286	0.294	0.97
0.8	-2.916	-2.953	0.99	0.350	0.371	0.94

Table 3.9: Simulation results of last false values using different values of ρ_{ij} and variance when the model has $(\exp(X_1), X_2)$ covariates in case $\Pr(Y = 1) \simeq 10\%$

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estimated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	-0.095	-0.100	0.95	0.055	0.059	0.92
-0.2	-0.236	-0.232	1.02	0.188	0.186	1.01
0.1	-0.331	-0.332	0.99	0.280	0.281	0.99
0.3	-0.400	-0.400	1.00	0.346	0.344	1.01
0.8	-0.567	-0.566	1.00	0.503	0.503	1.00
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	0.044	0.045	0.97	-0.076	-0.078	0.98
-0.2	-0.179	-0.183	0.98	0.135	0.139	0.97
0.1	-0.353	-0.355	0.99	0.301	0.302	0.99
0.3	-0.466	-0.470	0.99	0.410	0.412	0.99
0.8	-0.759	-0.759	1.00	0.686	0.687	0.99
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	-0.162	-0.173	0.94	0.146	0.156	0.94
-0.2	-0.251	-0.256	0.98	0.213	0.217	0.98
0.1	-0.321	-0.318	1.02	0.265	0.263	1.01
0.3	-0.369	-0.359	1.03	0.302	0.294	1.02
0.8	-0.496	-0.464	1.07	0.401	0.371	1.08

Table 3.10: Simulation results of last false values using different values of ρ_{ij} and variance when the model has $(\exp(X_1), X_2)$ covariates in case $\Pr(Y = 1) \simeq 60\%$

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estimated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	-2.216	-2.224	0.99	0.047	0.056	0.85
-0.2	-2.357	-2.357	1.00	0.184	0.183	1.01
0.1	-2.461	-2.461	1.00	0.284	0.283	1.01
0.3	-2.532	-2.533	0.99	0.350	0.351	0.99
0.8	-2.727	-2.721	1.01	0.532	0.527	1.01
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	-1.954	-1.997	0.98	-0.152	-0.112	1.36
-0.2	-2.228	-2.240	0.99	0.112	0.123	0.91
0.1	-2.436	-2.436	0.99	0.313	0.311	1.01
0.3	-2.592	-2.575	1.01	0.458	0.442	1.04
0.8	-2.996	-2.956	1.01	0.830	0.793	1.05
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	-2.222	-2.211	1.01	0.159	0.149	1.06
-0.2	-2.290	-2.289	1.00	0.212	0.210	1.01
0.1	-2.354	-2.362	0.99	0.263	0.265	0.99
0.3	-2.409	-2.419	0.99	0.303	0.306	0.99
0.8	-2.598	-2.604	0.99	0.430	0.431	0.99

Table 3.11: Simulation results of last false values using different values of ρ_{ij} and variance when the model has $(\exp(X_1), \exp(X_2))$ covariates in case $\Pr(Y = 1) \simeq 10\%$

$\sigma_1^2 = \sigma_2^2 = 0.1$	Parameters estimated, Least false values and Ratio					
ρ	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2
-0.6	0.274	0.271	1.01	0.052	0.056	0.94
-0.2	0.135	0.136	0.99	0.184	0.183	1.01
0.1	0.035	0.032	1.08	0.280	0.282	0.99
0.3	-0.037	-0.038	0.96	0.347	0.350	0.99
0.8	-0.223	-0.223	1.00	0.525	0.526	0.99
$\sigma_1^2 = 0.1, \sigma_2^2 = 0.3$						
-0.6	0.470	0.484	0.97	-0.101	-0.112	0.91
-0.2	0.225	0.236	0.95	0.130	0.123	1.05
0.1	0.037	0.038	0.96	0.310	0.311	0.99
0.3	-0.095	-0.098	0.97	0.436	0.442	0.99
0.8	-0.451	-0.466	0.97	0.777	0.793	0.98
$\sigma_1^2 = 0.6, \sigma_2^2 = 0.4$						
-0.6	0.267	0.256	1.04	0.138	0.149	0.92
-0.2	0.173	0.173	0.99	0.207	0.210	0.98
0.1	0.091	0.099	0.91	0.268	0.265	1.01
0.3	0.033	0.044	0.74	0.311	0.306	1.02
0.8	-0.142	-0.121	1.17	0.448	0.4317	1.03

Table 3.12: Simulation results of last false values using different values of ρ_{ij} and variance when the model has $(\exp(X_1), \exp(X_2))$ covariates in case $\Pr(Y = 1) \simeq 60\%$

3.5 Conclusion

We have applied the results defined in (2.14) and (2.15), which assumed covariates were multivariate normal, when the covariates do not follow this distribution. We consider five dimensional multivariate uniform and t-variables when only two covariates were fitted. The results showed that for these symmetric non-normal variables, the violation of the assumption of normality made little difference. We considered various two dimensional ways skewness could affect on results. Again the results derived in chapter 2 gave accurate answers. Some discrepancies were noticed when the value of coefficients were close to zero.

The effect of categorical variables has not been considered so far and we now continue with this case and we will discuss it in the next chapter.

Chapter 4

Least false values for logistic regression with one binary covariate and some multivariate normal covariates, some of which are omitted

4.1 Introduction

As we have discussed in the chapter 2, the properties of the skew-normal distribution can be used to find the least false values under a logistic regression model with missing covariates. The assumption on the covariates was that they come from a multivariate normal distribution. In this chapter we are interested to extend the work and examine the behaviour of MLE method and compute the least false values when one of covariates is binary. Suppose that the model we fit is

$$E(Y) = \text{expit}(\alpha + \gamma C + \beta_f^T X_f), \tag{4.1}$$

where $C \in \{0, 1\}$ is the binary covariate. Then X_f is a multivariate normal, p dimension variable. Suppose that

$$P(X_f|C = 0) \sim N(\mu_0, \Omega_0),$$

and

$$P(X_f|C = 1) \sim N(\mu_1, \Omega_1).$$

Let us consider

$$P(C = 0) = 1 - P(C = 1) = 1 - \pi_1 = \pi_0.$$

However, the correct model is

$$
\begin{aligned}
E(Y) &= \operatorname{expit}(\alpha + \gamma C + \beta_f^T X_f + \beta_a^T X_a) \\
&= \operatorname{expit}(\alpha + \gamma C + \beta^T X),
\end{aligned}
$$

say, where $X = (X_f \mid X_a)$, and X_a is a q-dimension multivariate normal variable of additional covariates that have not been included in the fitted model. Also, suppose that the partition of the Ω_j matrices corresponding to X_f and X_a is:

$$
\Omega_j = \begin{bmatrix} \Omega_{jff} & \Omega_{jfa} \\ \Omega_{jaf} & \Omega_{jaa} \end{bmatrix}, \ j = 0, 1.
$$

4.2 Computation of the Least False Values

As we have discussed in the chapter 2 about the MLE in case of the logistic model with missing covariates, we produced the ML equation which are used to find the least false values. In this chapter we consider the logistic regression model has some multivariate normal covariates and one binary covariate, so, if the model (4.1) has been fitted then the least false values, α^*, γ^* and β_f^* obey

$$
E(Y - \operatorname{expit}(\alpha^* + \gamma^* C + \beta_f^{*T} X_f)) = 0 \tag{4.2}
$$

$$
E(C(Y - \operatorname{expit}(\alpha^* + \gamma^* + \beta_f^{*T} X_f))) = 0 \tag{4.3}
$$

and

$$
E(X_{f\ell}(Y - \operatorname{expit}(\alpha^* + \gamma^* C + \beta_f^{*T} X_f))) = 0 \tag{4.4}
$$

for $\ell = 1, \cdots, p$. Now, we will work to analysis the (4.2), (4.3) and (4.4) to find the least false values, and we will compute it in the following subsections.

4.2.1 Calculation for Equation (4.2)

To analysis this equation we take the expectation of Y, given C and X gives,

$$
E_{X,C}(\operatorname{expit}(\alpha + \gamma C + \beta^T X)) = E_{X,C}(\operatorname{expit}(\alpha^* + \gamma^* C + \beta_f^{*T} X_f)).
$$

Now, suppose that the density of $X \mid C = j$ is g_j, $(j = 0, 1)$ and let us consider $Z_j = X - \mu_j$. Where Z_j is the centred version of $X \mid C = j$, μ_{jf} is the part of μ_j corresponding to X_f in $X = (X_f | X_a)$ when $C = j$. So,

$$
\pi_0 E_{g_0}(\operatorname{expit}(\alpha + \beta^T \mu_0 + \beta^T Z_0)) + \pi_1 E_{g_1}(\operatorname{expit}(\alpha + \gamma + \beta^T \mu_1 + \beta^T Z_1)) =
$$

$$
\pi_0 E_{g_0}(\operatorname{expit}(\alpha^* + \beta_f^{*T} \mu_{0f} + \beta_f^{*T} Z_{0f})) + \pi_1 E_{g_1}(\operatorname{expit}(\alpha^* + \gamma^* + \beta_f^{*T} \mu_{1f} + \beta_f^{*T} Z_{1f})) \tag{4.5}
$$

4.2.2 Calculation for Equation (4.3)

The form (4.3) include the variable C which is $C \in \{0, 1\}$, so, in case of $C = 0$ (4.3) will be zero, and when take the expectation of Y, given $C = 1$ and X, we obtain

$$E_{X,C=1}(\text{expit}(\alpha + \gamma + \beta^T X)) = E_{X,C=1}(\text{expit}(\alpha^* + \gamma^* + \beta_f^{*T} X_f)).$$

We use the same assumption which used before, considering Z_j and μ_{jf}, so, we can write the previous equation as

$$\pi_1 E_{g1}(\text{expit}(\alpha + \gamma + \beta^T \mu_1 + \beta^T Z_1)) = \pi_1 E_{g1}(\text{expit}(\alpha^* + \beta_f^{*T} \mu_{1f} + \gamma^* + \beta_f^{*T} Z_{1f})) \quad (4.6)$$

we can see clearly, if we use (4.5) in (4.6), we obtain

$$E_{g0}(\text{expit}(\alpha + \beta^T \mu_0 + \beta^T Z_0)) = E_{g0}(\text{expit}(\alpha^* + \beta_f^{*T} \mu_{0f} + \beta_f^{*T} Z_{0f})) \quad (4.7)$$

and

$$E_{g1}(\text{expit}(\alpha + \gamma + \beta^T \mu_1 + \beta^T Z_1)) = E_{g1}(\text{expit}(\alpha^* + \gamma^* + \beta_f^{*T} \mu_{1f} + \beta_f^{*T} Z_{0f})). \quad (4.8)$$

4.2.3 Calculation for Equation (4.4)

As we worked on the two previous subsections, now we have

$$E(X_{f\ell}(Y - \text{expit}(\alpha^* + \gamma^* C + \beta_f^{*T} X_f))) = 0,$$

where, $l = 1, \cdots, p$, and take the expectation of Y given C and X, gives

$$E(X_{f\ell}\text{expit}(\alpha + \gamma C + \beta^T X)) = E(X_{f\ell}(\text{expit}(\alpha^* + \gamma^* C + \beta_f^{*T} X_f)).$$

Now, we use the assumption on Z_j and μ_j where $j = 0, 1$ as before, then we can write the previous equation as

$$\pi_0 E_{g0}(Z_{f\ell}\text{expit}(\alpha + \beta^T \mu_0 + \beta^T Z_0)) + \pi_1 E_{g1}(Z_{f\ell}\text{expit}(\alpha + \beta^T \mu_1 + \beta^T Z_1)) =$$

$$\pi_0 E_{g_0}(Z_{f\ell}\text{expit}(\alpha^* + \beta_f^{*T} \mu_{0f} + \beta_f^{*T} Z_{0f})) + \pi_1 E_{g_1}(Z_{f\ell}\text{expit}(\alpha^* + \gamma^* + \beta_f^{*T} \mu_{1f} + \beta_f^{*T} Z_{1f})) \quad (4.9)$$

4.2.4 Solve Equations by Use the Approximation Form and Properties of the Skew-Normal Distribution

As we apply the assumption on (4.2), (4.3) and (4.4), we found results (4.5), (4.7), (4.8) and (4.9). Now, we need using approximation form $\text{expit}(u) \approx \Phi(ku)$, where $k = \frac{16\sqrt{3}}{15\pi}$ and the properties of the skew-normal distribution to solve these equations. If we approximate $\text{expit}(u)$ by $\Phi(ku)$ then we can use the results which discussed in chapter 2, i.e,

$$E(\text{expit}(\nu + \beta^T Z)) \approx E(\Phi(k(\nu + \beta^T Z))) = \Phi\left(\frac{k\nu}{\sqrt{1 + k^2 \beta^T \Omega \beta}}\right),$$

Then, if we apply this to (4.7) and (4.8) respectively we obtain

$$\frac{\alpha + \beta^T \mu_0}{\sqrt{1 + k^2 \beta^T \Omega_0 \beta}} = \frac{\alpha^* + \beta_f^{*T} \mu_{0f}}{\sqrt{1 + k^2 \beta_f^{*T} \Omega_{0ff} \beta_f^*}} \tag{4.10}$$

and

$$\frac{\alpha + \gamma + \beta^T \mu_1}{\sqrt{1 + k^2 \beta^T \Omega_1 \beta}} = \frac{\alpha^* + \gamma^* + \beta_f^{*T} \mu_{1f}}{\sqrt{1 + k^2 \beta_f^{*T} \Omega_{1ff} \beta_f^*}}. \tag{4.11}$$

Here, we assume g_j is a multivariate normal with mean μ_j and variance Ω_j. Subscript $f's$ denote the part of the vector or matrix corresponding to the X_f part of X. Now, the left hand side of (4.9) can be approximated by

$$\frac{\pi_0 k(\Omega_0 \beta)_l}{\sqrt{1 + k^2 \beta^T \Omega_0 \beta}} \phi\left(\frac{k(\alpha + \beta^T \mu_0)}{\sqrt{1 + k^2 \beta^T \Omega_0 \beta}}\right) + \frac{\pi_1 k(\Omega_1 \beta)_l}{\sqrt{1 + k^2 \beta^T \Omega_1 \beta}} \phi\left(\frac{k(\alpha + \gamma + \beta^T \mu_1)}{\sqrt{1 + k^2 \beta^T \Omega_1 \beta}}\right)$$

and this must be equal to

$$\frac{\pi_0 k(\Omega_{0ff} \beta_f^*)_l}{\sqrt{1 + k^2 \beta_f^{*T} \Omega_{0ff} \beta_f^*}} \phi\left(\frac{k(\alpha^* + \beta_f^{*T} \mu_{0f})}{\sqrt{1 + k^2 \beta_f^{*T} \Omega_{0ff} \beta_f^*}}\right) + \frac{\pi_1 k(\Omega_{1ff} \beta_f^*)_l}{\sqrt{1 + k^2 \beta_f^{*T} \Omega_{1ff} \beta_f^*}} \phi\left(\frac{k(\alpha^* + \gamma^* + \beta_f^{*T} \mu_{1f})}{\sqrt{1 + k^2 \beta_f^{*T} \Omega_{1ff} \beta_f^*}}\right).$$

Equations (4.10) and (4.11) mean that the above equations can be written

$$\frac{\pi_0 w_0}{A_0^*}(\Omega_{0ff} \beta_f^*) + \frac{\pi_1 w_1}{A_1^*}(\Omega_{1ff} \beta_f^*) = \frac{\pi_0 w_0}{A_0}(\Omega_0 \beta)_f + \frac{\pi_1 w_1}{A_1}(\Omega_1 \beta)_f$$

where,

$$A_j^* = \sqrt{1 + k^2 \beta_f^{*T} \Omega_{jff} \beta_f^*},$$

$$A_j = \sqrt{1 + k^2 \beta^T \Omega_j \beta}$$

and

$$w_0 = \phi\left(\frac{k(\alpha + \beta^T\mu_0)}{A_0}\right)$$

$$w_1 = \phi\left(\frac{k(\alpha + \gamma + \beta^T\mu_1)}{A_1}\right)$$

4.2.5 The Least false values in case $\Omega_0 = \Omega_1$

To find the least false values, we consider the case is $\Omega_0 = \Omega_1$: the case $\Omega_0 \neq \Omega_1$, appears to be more challenging. In this case $A_0^* = A_1^* = A^*$ and $A_0 = A_1 = A$, say. Therefore

$$A^{*-1}(\pi_0 w_0 + \pi_1 w_1)\Omega_{ff}\beta_f^* = A^{-1}(\pi_0 w_0 + \pi_1 w_1)(\Omega\beta)_f$$

then

$$\begin{aligned}\beta_f^* &= \frac{A^*}{A}\Omega_{ff}^{-1}(\Omega\beta)_f \\ &= \frac{A^*}{A}\Omega_{ff}^{-1}(\Omega_{ff}\beta_f + \Omega_{fa}\beta_a)\end{aligned}$$

As in the case of no binary variable, we proceed to eliminate β_f^* from the right hand side.

$$\begin{aligned}\beta_f^{*T}\Omega_{ff}\beta_f^* &= \frac{A^{*2}}{A^2}(\Omega_{ff}\beta_f + \Omega_{fa}\beta_a)^T\Omega_{ff}^{-1}(\Omega_{ff}\beta_f + \Omega_{fa}\beta_a) \\ &= \frac{A^{*2}}{A^2}\left[\beta_a^T\Omega_{af}\Omega_{ff}^{-1}\Omega_{fa}\beta_a + \beta_f^T\Omega_{ff}\beta_f + \beta_a^T\Omega_{af}\beta_f + \beta_f^T\Omega_{fa}\beta_a\right] \\ &= \frac{A^{*2}}{A^2}\left[\beta^T\Omega\beta - \beta_a^T\tilde{\Omega}\beta_a\right]\end{aligned}$$

where $\tilde{\Omega} = \Omega_{aa} - \Omega_{af}\Omega_{ff}^{-1}\Omega_{fa}$. So

$$\beta_f^{*T}\Omega_{ff}\beta_f^* = \frac{(1 + k^2\beta_f^{*T}\Omega_{ff}\beta_f^*)}{1 + k^2\beta^T\Omega\beta}(\beta^T\Omega\beta - \beta_a^T\tilde{\Omega}\beta_a)$$

and so, $\frac{A^{*2}}{A^2}$ can be found to be $(1 + k^2\beta_a^T\tilde{\Omega}\beta_a)^{-1}$ and hence the least false values are

$$\beta_f^* = \frac{1}{\sqrt{1 + k^2\beta_a^T\tilde{\Omega}\beta_a}}(\beta_f + \Omega_{ff}^{-1}\Omega_{fa}\beta_a). \tag{4.12}$$

Also:

$$\alpha^* + \beta_f^{*T}\mu_{0f} = \frac{1}{\sqrt{1 + k^2\beta_a^T\tilde{\Omega}\beta_a}}(\alpha + \beta^T\mu_0)$$

and

$$\alpha^* + \gamma^* + \beta_f^{*T}\mu_{1f} = \frac{1}{\sqrt{1 + k^2\beta_a^T\tilde{\Omega}\beta_a}}(\alpha + \gamma + \beta^T\mu_1).$$

From the first of these and (4.12) we get

$$\alpha^* = \frac{1}{\sqrt{1 + k^2 \beta_a^T \tilde{\Omega} \beta_a}} (\alpha + \beta^T \mu_0 - \beta_f^T \mu_{0f} - \beta_a^T \Omega_{af} \Omega_{ff}^{-1} \mu_{0f})$$

So, finally we can write the least false value α^* as

$$\alpha^* = \frac{1}{\sqrt{1 + k^2 \beta_a^T \tilde{\Omega} \beta_a}} (\alpha + \beta_a^T (\mu_{0a} - \Omega_{af} \Omega_{ff}^{-1} \mu_{0f})) \tag{4.13}$$

Also we can find γ^* as

$$\gamma^* = \frac{1}{\sqrt{1 + k^2 \beta_a^T \tilde{\Omega} \beta_a}} \left[\alpha + \gamma + \beta^T \mu_1 - (\alpha + \beta_a^T (\mu_{0a} - \Omega_{af} \Omega_{ff}^{-1} \mu_{0f})) - (\beta_f^T \mu_{1f} + \beta_a^T \Omega_{af} \Omega_{ff}^{-1} \mu_{1f})\right]$$

$$= \frac{1}{\sqrt{1 + k^2 \beta_a^T \tilde{\Omega} \beta_a}} \left[\gamma + \beta_a^T (\mu_{1a} - \Omega_{af} \Omega_{ff}^{-1} \mu_{1f}) - \beta_a^T (\mu_{0a} - \Omega_{af} \Omega_{ff}^{-1} \mu_{0f})\right]$$

finally the least false value γ^* is

$$\gamma^* = \frac{1}{\sqrt{1 + k^2 \beta_a^T \tilde{\Omega} \beta_a}} \left[\gamma + \beta_a^T ((\mu_{1a} - \mu_{0a}) - \Omega_{af} \Omega_{ff}^{-1} (\mu_{1f} - \mu_{0f}))\right] \tag{4.14}$$

If we comparing (4.12) and (4.13) with (2.14) and (2.15), we can see clearly, (4.12) and (2.14) have same expression. However, the expression (4.13) has been affected which is dependent upon μ_{0a} and μ_{0f}. Note that, (4.13) include μ_{0a}, such that $\alpha^* \neq \alpha$ even when $\Omega_{af} = 0$. Also (4.14) includes μ_{0a} and μ_{1a} and $\gamma^* \neq \gamma$ even when $\Omega_{af} = 0$. However, only when $\beta_a = 0$ we obtain $\alpha^* = \alpha$, $\beta_f^* = \beta_f$ and $\gamma^* = \gamma$, i.e. the fitted model is correct.

4.3 The Least False Values When the Fitted Model has Only One Binary Covariate

4.3.1 Introduction

The previous section discussed the behaviour of the MLE method and finds the least false values when some of the multivariate normal covariates are omitted from the fitted model. In this section we are interested to find the least false values when all the covariates of the multivariate normal are omitted and the fit model contain only one binary covariate. Suppose that we have correct model is

$$E(Y) = \text{expit}(\alpha + \gamma C + \beta X)$$

use the same steps applied in previous section, X distributed multivariate normal, β is vector has dimension $p \times 1$ and the model we fit is

$$E(Y) = \text{expit}(\hat{\alpha} + \hat{\gamma}C), \tag{4.15}$$

and we want to find the least false values α^* and γ^* in terms of the parameters of the true model.

4.3.2 The Least False Values of α^* and γ^*

As the same steps which used in the previous section, in this case the least false values, α^* and γ^* obey

$$E(Y - \text{expit}(\alpha^* + \gamma^* C)) = 0 \tag{4.16}$$

$$E(C(Y - \text{expit}(\alpha^* + \gamma^* C))) = 0 \tag{4.17}$$

$$E(X_j(Y - \text{expit}(\alpha^* + \gamma^* C))) = 0 \tag{4.18}$$

We transform X to multivariate normal distribution with zero mean as used before, then equation (4.16) becomes

$$\pi_0 E_{g_0}(\text{expit}(\alpha^*)) + \pi_1 E_{g_1}(\text{expit}(\alpha^* + \gamma^*))$$

$$= \pi_0 E_{g_0}(\text{expit}(\alpha + \beta^T \mu_0 + \beta^T Z_0)) + \pi_1 E_{g_1}(\text{expit}(\alpha + \gamma + \beta^T \mu_1 + \beta^T Z_1)),$$

and we can write the previous equation as

$$\pi_0 \text{expit}(\alpha^*) + \pi_1 \text{expit}(\alpha^* + \gamma^*) = \pi_0 \Phi\left(\frac{k(\alpha + \beta^T \mu_0)}{\sqrt{1 + k^2 \beta^T \Omega \beta}}\right) + \pi_1 \Phi\left(\frac{k(\alpha + \gamma + \beta^T \mu_1)}{\sqrt{1 + k^2 \beta^T \Omega \beta}}\right). \tag{4.19}$$

We can rewrite the equation (4.17) as

$$\pi_1 \text{expit}(\alpha^* + \gamma^*) = \pi_1 \Phi\left(\frac{k(\alpha + \gamma + \beta^T \mu_1)}{\sqrt{1 + k^2 \beta^T \Omega \beta}}\right),$$

so as approximation of $\Phi(ku) \approx \text{expit}(u)$ this equation can be written

$$\text{expit}(\alpha^* + \gamma^*) = \text{expit}\left(\frac{(\alpha + \gamma + \beta^T \mu_1)}{\sqrt{1 + k^2 \beta^T \Omega \beta}}\right),$$

then

$$\alpha^* + \gamma^* = \frac{(\alpha + \gamma + \beta \mu_1)}{\sqrt{1 + k^2 \beta^T \Omega \beta^T}}. \tag{4.20}$$

Use this result, then (4.19) becomes

$$\text{expit}(\alpha^*) = \Phi\left(\frac{k(\alpha + \beta^T \mu_0)}{\sqrt{1 + k^2 \beta^T \Omega \beta^T}}\right),$$

and we can write it as

$$\text{expit}(\alpha^*) = \text{expit}\left(\frac{\alpha + \beta^T \mu_0}{\sqrt{1 + k^2 \beta^T \Omega \beta}}\right),$$

then the least false value α^* is

$$\alpha^* = \frac{\alpha + \beta^T \mu_0}{\sqrt{1 + k^2 \beta^T \Omega \beta^T}}. \tag{4.21}$$

To find the least false value γ^*, using the result of (4.21) in (4.20) we get,

$$\frac{\alpha + \beta^T \mu_0}{\sqrt{1 + k^2 \beta^T \Omega \beta}} + \gamma^* = \frac{(\alpha + \gamma + \beta^T \mu_1)}{\sqrt{1 + k^2 \beta^T \Omega \beta}}$$

then the least false value γ^* is

$$\gamma^* = \frac{(\alpha + \gamma + \beta^T \mu_1)}{\sqrt{1 + k^2 \beta^T \Omega \beta}} - \frac{\alpha + \beta^T \mu_0}{\sqrt{1 + k^2 \beta^T \Omega \beta}}$$

$$\gamma^* = \frac{\gamma + \beta^T (\mu_1 - \mu_0)}{\sqrt{1 + k^2 \beta^T \Omega \beta}} \tag{4.22}$$

Finally, we can note that $\alpha^* \neq \alpha$ and $\gamma^* \neq \gamma$ even when $\Omega = 0$ or when $\mu_0 = \mu_1 = 0$, the expression still dependent on β. But does not matter when $\beta = 0$ which is because we have fitted the true model, so, in this case the least false values are $\alpha^* = \alpha$ and $\gamma^* = \gamma$.

4.4 Simulation Study

In this part we are interested to examine the expression of least false values which is computed by use the properties of skew-normal distribution in previous sections.

4.4.1 Design of Studies

This simulation designed to examine the behaviour of the expression for the least false values when the true logistic model contains six covariates c, x where, $c \epsilon \{0, 1\}$ is binary covariate and the rest of the covariates $x = (x_1, x_2, x_3, x_4, x_5)^T$ have normal distribution $N(\mu, \Omega)$

$$\Pr(Y = 1 | c, x_i) = \text{expit}(\alpha + \gamma c + \beta^T x_i),$$

where, x_i is the value of x on the i^{th} individual, $j = 1, \cdots, 5$, $i = 1, \cdots, n$ and the logistic regression model has only three covariates one of them is binary has been fitted. We designed the simulation as follows:

- $x_i|c$ has normal distribution with μ_0, Ω and μ_1, Ω when $c = 0$ and $c = 1$ respectively.
- Two cases of means have been chosen ($\mu_0 > \mu_1, \mu_0 < \mu_1$), so, that the effect of the opposite direction of bias that apply to γ are considered.
- We choose the parameters as follows $\gamma = 0.25, \beta_1 = 0.35, \beta_2 = 0.40, \beta_3 = 0.30, \beta_4 = 0.2, \beta_5 = 0.30$ and adjust $\alpha = -2.2, -5.2$, to give us two cases $\Pr(Y = 1) \simeq 60\%$ and $\Pr(Y = 1) \simeq 10\%$ respectively.
- We consider 5×5 covariance matrix Ω, which used in design of simulation in previous chapter, and use the same cases of correlation and variance.
- Large sample size has been used $n = 500$, $n = 10000$ and $N = 1000$ number of simulation.

4.5 Results and Discussion

The results of simulation have been reported in tables. It is shows comparison between the least false values which computed by expression form denoted by α^*, γ^*, β_1^*, β_2^* and parameters estimated by fitted model. The ratio between of them denoted by R_1, R_2 R_3 and R_4 respectively. Two cases of comparison have been considered, the first case consider ($\mu_0 = 2 > \mu_1 = 1$) and the second case, consider ($\mu_0 = 1 < \mu_1 = 2$). Moreover, these two cases are applied with six combinations of different correlation and three cases of variance (0.1, 0.5 and 1.5) which are the same as used in the simulation in previous chapter. Table 4.1 and Table 4.2, shows the results in case of $\Pr(Y = 1) \simeq 60\%$ with sample size $n = 500$ with $\mu_0 = 1, \mu_1 = 2$ and $\mu_0 = 2, \mu_1 = 1$ respectively .Table 4.3 and Tabel 4.4, shows the results in case of $\Pr(Y = 1) \simeq 10\%$. The same steps considered in case of sample size $n = 10000$, Table 4.5 and Table 4.6, shows the results in case of $\Pr(Y = 1) \simeq 60\%$. Table 4.7 and Table 4.8, shows the results in case of $\Pr(Y = 1) \simeq 10\%$ with $\mu_0 = 1, \mu_1 = 2$ and $\mu_0 = 2, \mu_1 = 1$ respectively.

We can see clearly, that the results in all tables for all combination of means and variances shows a moderate ratio which is close to one. That is mean the computed

$\sigma^2 = 0.1$			Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	1.00	1.02	1.01	0.98
0.2	0.2	0.4	1.01	0.98	0.98	1.03
0.7	0.8	0.7	1.00	0.99	1.03	0.98
0.8	0.7	0.9	1.01	1.00	1.01	1.01
0.1	-0.2	0.4	1.01	1.00	1.01	1.03
0.2	-0.2	-0.2	1.01	0.99	1.00	1.04
$\sigma^2 = 0.5$						
0.1	0.1	0.2	1.00	0.99	1.01	1.00
0.2	0.2	0.4	0.99	1.00	0.99	1.01
0.7	0.8	0.7	1.01	1.01	1.03	0.99
0.8	0.7	0.9	1.01	0.96	1.03	0.99
0.1	-0.2	0.4	1.00	1.00	1.00	1.00
0.2	-0.2	-0.2	0.99	1.01	0.98	0.97
$\sigma^2 = 1.5$						
0.1	0.1	0.2	0.99	0.99	1.00	0.99
0.2	0.2	0.4	0.99	0.98	0.99	0.99
0.7	0.8	0.7	1.01	0.99	1.02	1.01
0.8	0.7	0.9	1.00	1.01	1.01	1.01
0.1	-0.2	0.4	0.99	0.99	1.00	0.98
0.2	-0.2	-0.2	1.00	0.99	1.01	1.01

Table 4.1: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 1, \mu_1 = 2$ in case $\Pr(Y = 1) \simeq 60\%$, $n = 500$, R_i denote to the ratio of the least false values $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

form of the least false values works well, although the low percentage of probability $\Pr(Y = 1) \simeq 10\%$. On the other hand, the large values of variance have slightly affected. The effect of omitting some of the normal covariates on the estimate of β_f is the same attenuation as in Chapter 2. While the same attenuation is applied to γ, this is not the sole effect, because the difference in the means of the normal covariates between the populations with $C = 0$ and $C = 1$ has an effect.

$\sigma^2 = 0.1$			Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	1.02	0.99	1.04	0.99
0.2	0.2	0.4	1.05	0.92	1.03	1.03
0.7	0.8	0.7	1.01	0.99	1.02	1.00
0.8	0.7	0.9	1.01	1.05	0.99	1.02
0.1	-0.2	0.4	0.99	1.01	1.07	0.99
0.2	-0.2	-0.2	1.09	1.02	0.96	1.01
$\sigma^2 = 0.5$						
0.1	0.1	0.2	1.02	0.98	1.01	1.01
0.2	0.2	0.4	1.01	0.98	1.01	1.01
0.7	0.8	0.7	1.02	1.04	1.02	1.02
0.8	0.7	0.9	1.01	1.16	0.99	1.01
0.1	-0.2	0.4	0.98	1.00	1.02	0.97
0.2	-0.2	-0.2	0.98	1.01	1.01	1.02
$\sigma^2 = 1.5$						
0.1	0.1	0.2	1.01	0.96	0.99	1.01
0.2	0.2	0.4	1.01	0.95	1.00	0.99
0.7	0.8	0.7	1.01	1.02	1.01	1.01
0.8	0.7	0.9	1.01	1.17	1.01	0.99
0.1	-0.2	0.4	1.06	0.98	0.97	0.99
0.2	-0.2	-0.2	0.93	0.99	1.02	1.01

Table 4.2: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 2, \mu_1 = 1$ in case $\Pr(Y = 1) \simeq 60\%$, $n = 500$, R_i denote to the ratio of the least false values $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

$\sigma^2 = 0.1$			Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	1.03	1.11	1.03	1.01
0.2	0.2	0.4	1.03	1.16	1.01	0.98
0.7	0.8	0.7	1.02	1.23	0.98	1.04
0.8	0.7	0.9	1.03	1.20	1.01	1.01
0.1	-0.2	0.4	1.05	1.13	1.04	1.03
0.2	-0.2	-0.2	1.03	1.08	1.06	0.95
$\sigma^2 = 0.5$						
0.1	0.1	0.2	1.03	1.06	1.03	1.03
0.2	0.2	0.4	1.02	1.11	1.00	1.00
0.7	0.8	0.7	1.02	1.05	1.02	1.01
0.8	0.7	0.9	1.02	1.11	1.01	1.01
0.1	-0.2	0.4	1.04	1.11	0.97	1.05
0.2	-0.2	-0.2	1.04	1.09	1.01	1.01
$\sigma^2 = 1.5$						
0.1	0.1	0.2	1.03	1.06	1.03	1.03
0.2	0.2	0.4	1.03	1.08	1.04	1.04
0.7	0.8	0.7	1.02	1.05	1.01	1.03
0.8	0.7	0.9	1.02	1.04	1.02	1.02
0.1	-0.2	0.4	1.03	1.08	1.03	1.03
0.2	-0.2	-0.2	1.03	1.07	0.03	1.01

Table 4.3: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 1, \mu_1 = 2$ in case $\Pr(Y = 1) \simeq 10\%$, $n = 500$, R_i denote to the ratio of the least false values $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

	$\sigma^2 = 0.1$		Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	1.01	1.15	0.98	1.04
0.2	0.2	0.4	1.01	1.20	0.93	1.04
0.7	0.8	0.7	1.03	1.11	1.05	1.04
0.8	0.7	0.9	1.01	0.46	1.07	0.95
0.1	-0.2	0.4	0.99	1.10	0.81	1.03
0.2	-0.2	-0.2	0.99	1.09	0.88	0.91
	$\sigma^2 = 0.5$					
0.1	0.1	0.2	1.02	1.15	1.01	1.05
0.2	0.2	0.4	1.02	1.16	1.03	1.01
0.7	0.8	0.7	1.02	0.95	1.05	1.01
0.8	0.7	0.9	1.02	0.47	1.02	1.02
0.1	-0.2	0.4	1.02	1.09	1.01	1.03
0.2	-0.2	-0.2	1.03	1.06	1.11	1.04
	$\sigma^2 = 1.5$					
0.1	0.1	0.2	1.02	1.17	1.03	1.04
0.2	0.2	0.4	1.02	1.10	1.04	1.03
0.7	0.8	0.7	1.02	0.99	1.03	1.01
0.8	0.7	0.9	1.02	0.84	1.01	1.04
0.1	-0.2	0.4	1.02	1.05	1.07	1.05
0.2	-0.2	-0.2	1.01	1.06	1.05	1.01

Table 4.4: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 2, \mu_1 = 1$ in case $\Pr(Y = 1) \simeq 10\%$, $n = 500$, R_i denote to the ratio of the least false values $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

	$\sigma^2 = 0.1$		Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	1.00	1.02	1.01	1.00
0.2	0.2	0.4	1.00	1.02	1.00	1.01
0.7	0.8	0.7	1.00	1.01	0.99	1.01
0.8	0.7	0.9	1.00	0.98	1.00	1.01
0.1	-0.2	0.4	1.00	1.01	0.99	1.01
0.2	-0.2	-0.2	1.00	1.00	1.02	1.01
	$\sigma^2 = 0.5$					
0.1	0.1	0.2	1.01	1.02	1.00	1.01
0.2	0.2	0.4	1.00	1.02	1.01	1.01
0.7	0.8	0.7	1.00	1.00	1.00	1.00
0.8	0.7	0.9	1.00	1.03	0.99	1.01
0.1	-0.2	0.4	1.01	1.02	1.01	1.02
0.2	-0.2	-0.2	1.00	1.00	1.02	1.01
	$\sigma^2 = 1.5$					
0.1	0.1	0.2	1.01	1.03	1.01	1.01
0.2	0.2	0.4	1.01	1.04	1.01	1.01
0.7	0.8	0.7	1.00	1.01	1.00	1.00
0.8	0.7	0.9	1.01	1.01	1.01	1.01
0.1	-0.2	0.4	1.01	1.04	1.02	1.02
0.2	-0.2	-0.2	1.01	1.01	1.02	1.01

Table 4.5: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 1, \mu_1 = 2$ in case $\Pr(Y = 1) \simeq 60\%$, $n = 10000$, $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

$\sigma^2 = 0.1$			Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	1.00	1.01	0.98	1.02
0.2	0.2	0.4	1.00	0.99	1.02	1.00
0.7	0.8	0.7	1.00	0.99	1.00	0.99
0.8	0.7	0.9	1.00	1.08	1.03	0.99
0.1	-0.2	0.4	1.00	1.01	1.01	1.00
0.2	-0.2	-0.2	0.99	1.01	0.98	1.00
$\sigma^2 = 0.5$						
0.1	0.1	0.2	1.00	1.04	1.01	1.01
0.2	0.2	0.4	1.01	1.03	1.01	1.01
0.7	0.8	0.7	1.00	0.99	0.99	1.00
0.8	0.7	0.9	1.00	0.91	1.01	1.00
0.1	-0.2	0.4	1.00	1.02	1.02	1.01
0.2	-0.2	-0.2	1.00	1.00	0.99	1.02
$\sigma^2 = 1.5$						
0.1	0.1	0.2	1.01	1.04	1.02	1.02
0.2	0.2	0.4	1.01	1.05	1.01	1.02
0.7	0.8	0.7	1.00	0.98	1.00	1.00
0.8	0.7	0.9	1.01	0.95	1.00	1.01
0.1	-0.2	0.4	1.00	1.04	1.03	1.03
0.2	-0.2	-0.2	1.00	1.02	1.00	1.01

Table 4.6: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 2, \mu_1 = 1$ in case $\Pr(Y = 1) \simeq 60\%$, $n = 10000$, $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

$\sigma^2 = 0.1$			Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	0.99	1.00	0.99	0.99
0.2	0.2	0.4	0.99	1.00	0.99	0.99
0.7	0.8	0.7	0.98	0.99	1.00	1.00
0.8	0.7	0.9	0.99	1.00	1.00	0.99
0.1	-0.2	0.4	0.99	0.99	1.01	0.99
0.2	-0.2	-0.2	0.99	1.00	0.99	0.99
$\sigma^2 = 0.5$						
0.1	0.1	0.2	0.99	0.99	0.99	0.99
0.2	0.2	0.4	0.99	0.99	0.99	0.99
0.7	0.8	0.7	0.99	0.99	1.00	0.99
0.8	0.7	0.9	0.99	0.99	0.99	0.99
0.1	-0.2	0.4	0.99	0.99	1.00	0.99
0.2	-0.2	-0.2	1.00	0.99	1.01	1.00
$\sigma^2 = 1.5$						
0.1	0.1	0.2	0.99	0.98	0.99	0.99
0.2	0.2	0.4	0.99	0.98	0.99	0.99
0.7	0.8	0.7	0.99	0.99	0.99	1.00
0.8	0.7	0.9	0.99	0.99	0.99	0.99
0.1	-0.2	0.4	0.98	0.98	0.99	0.99
0.2	-0.2	-0.2	0.99	0.99	1.00	0.99

Table 4.7: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 1, \mu_1 = 2$ in case $\Pr(Y = 1) \simeq 10\%$, $n = 10000$, $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

$\sigma^2 = 0.1$			Least false values and Ratio			
Ω_{11}	Ω_{12}	Ω_{22}	R_1	R_2	R_3	R_4
0.1	0.1	0.2	1.00	0.99	1.00	1.00
0.2	0.2	0.4	0.99	0.99	0.99	0.99
0.7	0.8	0.7	1.00	1.00	0.99	1.00
0.8	0.7	0.9	0.99	0.97	0.99	0.99
0.1	-0.2	0.4	1.10	0.99	0.99	1.01
0.2	-0.2	-0.2	1.02	1.00	1.00	0.99
$\sigma^2 = 0.5$						
0.1	0.1	0.2	0.99	0.99	0.99	0.99
0.2	0.2	0.4	0.99	0.98	0.99	0.99
0.7	0.8	0.7	0.99	1.00	0.99	1.00
0.8	0.7	0.9	0.99	1.00	1.00	0.99
0.1	-0.2	0.4	0.98	0.99	0.99	0.99
0.2	-0.2	-0.2	0.98	0.99	1.00	1.00
$\sigma^2 = 1.5$						
0.1	0.1	0.2	0.99	0.98	0.99	0.99
0.2	0.2	0.4	0.99	0.98	0.98	0.98
0.7	0.8	0.7	0.99	1.00	1.00	0.99
0.8	0.7	0.9	0.99	1.00	0.99	0.99
0.1	-0.2	0.4	1.02	0.98	0.98	0.98
0.2	-0.2	-0.2	0.99	0.99	0.99	0.99

Table 4.8: Simulation results of last false values using different cases of σ^2 and $\mu_0 = 2, \mu_1 = 1$ in case $\Pr(Y = 1) \simeq 10\%$, $n = 10000$, $\alpha^*, \gamma^*, \beta_1^*, \beta_2^*$ respectively

4.6 Application to Randomized Trials

In a randomized trial where the outcome is binary, the treatment effect is often summarised by a log-odds ratio. The analysis will often carry out treatment groups while adjusting for covariates and this will usually be done using logistic regression. Suppose that the true model for the binary outcome Y in clinical trial is

$$\Pr(Y = 1 \mid T, X_f, X_a) = \text{expit}(\alpha + \tau T + \beta_f^T X_f + \beta_a^T X_a)$$

where $T \in \{0, 1\}$ denotes the treatment allocation, with $\Pr(T = j) = \pi_j$. If the model used for the analysis is

$$\Pr(Y = 1 \mid T, X_f) = \text{expit}(\alpha + \tau T + \beta_f^T X_f) \tag{4.23}$$

then the effect of this mis-specification can be studied using the results obtained in Section 4.2.5. Note that the treatment effect is now measured by the log-odds ratio, τ. In Section 4.2.5 the distribution of $X \mid C = j$ was taken to have mean μ_j and Ω_j, with mathematical tractability leading us to the assume $\Omega_0 = \Omega_1$. However, if we take $C = T$ and $\gamma = \tau$ and if treatments are allocated using randomization and X_f and X_a are baseline values, then the assumption $\Omega_0 = \Omega_1$ follows automatically, as does the equality of means $\mu_0 = \mu_1$. From (4.14) it follows that the least false value for τ obtained from fitting (4.23) is

$$\tau^* \approx \frac{\tau}{\sqrt{1 + k^2 \beta_a^T \widetilde{\Omega} \beta_a}}. \tag{4.24}$$

There are two features of this expression which should be noted.

1 As distinct from a general binary covariate, the fact that the distribution of the other covariates is independent of treatment allocation means that the effect of mis-specification is to appear to shrink the log odds ratio towards zero.

2 The attenuation depends on $\beta_a^T \widetilde{\Omega} \beta_a$. The size of the attenuation is governed not only by β_a but also by Ω_{aa}. However, the presence of $\widetilde{\Omega}$ in the factor means that the effect is reduced if there is a non-zero covariance Ω_{af} between the omitted and fitted covariates. In the extreme case, where the variation in X_a is wholly accounted by variation in X_f then $\widetilde{\Omega} = 0$ and, as would be expected, the attenuation vanishes.

If $\beta_a^T \widetilde{\Omega} \beta_a$ is small then (4.24) implies $\tau^* - \tau \approx \frac{1}{2} k^2 \tau \beta_a^T \widetilde{\Omega} \beta_a$. Gail et al. (1984), give an expression for the approximate asymptotic bias which is, using our notation,

$$\frac{1}{2} \beta_a^T \widetilde{\Omega} \beta_a \left[\exp\left(-\frac{1}{2}\tau\right) - \exp\left(\frac{1}{2}\tau\right) \right] \frac{e^{\alpha + \frac{1}{2}\tau}}{(1 + \exp^{\alpha})(1 + \exp^{\alpha+\tau})} \tag{4.25}$$

as mentioned in chapter 2. The apparent differences in dispersion matrix is because Gail and Colleagues fit a model which omits all covariates (which they do not assume to be multivariate Normal), so $X = X_a$ and in this case $\widetilde{\Omega} = \Omega_{aa}$. For small τ the factor in [] in (4.25) is approximately $-\tau$. The final factor in (4.25) is, if we neglect τ, of the form $p(1-p)$ so cannot exceed $\frac{1}{4}$ and as $k^2 = 0.346$ the two forms are broadly in agreement when the probability of response is not too extreme.

4.6.1 An Example: The Mayo Clinic Primary Biliary Cirrhosis Trial

No direct evaluation of the above results is possible as they are all expressed in terms of parameters values. However, some indication of size of the asymptotic bias, and how this changes with the included covariates, would be helpful. Purely by way of illustration, and so that realistic parameter values are chosen, we consider data from a trial of patients with primary biliary cirrhosis (PBC) conducted at the

Mayo Clinic over ten years from 1974. The trial randomized patients to placebo or penicillamine and is reported by Dickson et al. (1985), and the data are given in the book by Fleming and Harrington (2005). The data considered 312 patients in the trial with two groups, 158 patients takes D-penicillamine, 93 alive and 65 dead and 154 patients the placebo, 94 alive and 60 dead. We consider outcome as mortality and fit a model with a treatment indicator and five continuous baseline covariates, namely the serum values of bilirubin (mg/dl), cholesterol (mg/dl), and albumin (gm/dl), urinary copper (μg/day) and alkaline phosphatase (U/litre). All variables but albumin were log-transformed (base 10) to achieve Normality.

	log bilirubin	log cholesterol	albumin	log copper
log bilirubin	1			
log cholesterol	0.488	1		
albumin	-0.360	-0.038	1	
log copper	0.598	0.217	-0.278	1
log alkaline phosphatase	0.295	0.351	-0.146	0.277

Table 4.9: The correlations obtained from the dispersion matrix for the five continuous covariates chosen from the PBC trial

Included variables	$\tilde{q}$
Non	1.311
+ log bilirubin	1.072
+log cholesterol	1.068
+albumin	1.056
+log copper	1.039

Table 4.10: The values of $\tilde{q} = \sqrt{1 + k^2 \beta_a^T \tilde{\Omega} \beta_a}$ for a cumulative series of models

The dispersion matrix of the five baseline covariates, based on the 312 patients in the trial, was used as Ω and β was taken to be the estimated regression coefficients from the logistic regression. The values of $\widetilde{q} = \sqrt{1 + k^2 \beta_a^T \widetilde{\Omega} \beta_a}$ were then computed for a sequence of models in which the first model includes only the treatment indicator, the second also includes log bilirubin, and then, successively, log cholesterol, albumin and log copper are added. The correlations are shown in Table 4.9 and the $\tilde{q}$ values are in Table 4.10.

If we assume that the model with treatment indicator and all five variables is the correct model, then $\hat{\tau}$ from this model will be asymptotically unbiased. However, if a model with no covariates is fitted, $\hat{\tau}$ will tend to $\tau/\widetilde{q} \approx \tau/1.3$, i.e. a value about 75% of the correct value. Including log bilirubin reduced the bias and $\hat{\tau}$ will tend to $\tau/1.07$, a value in error by approximately 7%. As Table 4.10 shows, this can be reduced further by including more covariates, although the change is never as marked as when the first variable was introduced. Of course, different results would be obtained if terms were added in a different order.

τ^*, as given by (4.22), is the asymptotically biassed version of τ, i.e. it is the limiting value of $E(\hat{\tau})$ as the sample size increases without limit, when relevant covariates are omitted. While we see that $|\tau^*| < |\tau|$, it does not follow that $\hat{\tau}$ will be similarly shrunk relative to τ in any particular study.

Chapter 5

Information Matrix Test (IMT)

5.1 Introduction

The previous chapters discussed the behaviour of MLE and find the least false values for the logistic regression model under missing covariates. We considered multivariate normal, lognormal and binary covariates. We know that after fitting the logistic regression model, the next step is to examine how well the proposed model fits the observation data, this is called a goodness-of-fit test. Now we are interested to focus on one of the important global goodness-of-fit test, Information Matrix Tests (IMT). As we discussed in chapter 1, IMT appeared to give a reasonable results in a simulation which was discussed by Kuss (2002), who found it had good power for the logistic regression model. In this chapter we will give introduction to the IMT and a variant used by Kuss, the Diagonal Information Matrix Test IMT_{DIAG}. The IMT is a test for general mis-specification, produced by White (1982) who pointed out that the properties of the Maximum likelihood estimator and the information matrix can be exploited to yield a family of useful tests for model mis-specification. The idea of the IMT is to compare two different estimators of the information matrix to assess model fit. The IMT provides a unified framework for specification goodness of fit tests for a wide variety of distribution, multivariate or univariate, discrete or continuous.

5.2 Definition of IMT

The IMT is based on the information matrix equality that obtains when the model specification is correct. This equality implies the asymptotic equivalence of the Hessian and the score forms of Fisher's information matrix. As White (1982), points out, the IMT is designed to detect the failure of this equality and the failure implies the model mis-specification. The idea of the information matrix test is to compare

$E\left(\frac{-\partial^2\ell}{\partial\theta\partial\theta^T}\right)$ and $E\left(\frac{\partial\ell}{\partial\theta}\frac{\partial\ell}{\partial\theta^T}\right)$, as these differ when the model is mis-specified but not when the model is correct.

5.3 Fisher Information Matrix

Fisher information matrix essentially describes the amount of information data about an unknown parameter. Consider $X = (x_1, x_2, x_3, \ldots, x_n)^T$, a random sample, and consider the density function is $f(X \mid \theta)$ for some model of the data, where $\theta = (\theta_1, \ldots, \theta_p)^T$ is parameter vector and $\ell(\theta)$ is the log-likelihood function. So, the Fisher information matrix of sample size n, $I_n(\theta)$, is given by a $p \times p$ symmetric matrix whose rs^{th} element is given by the negative expected values of the second derivatives of the log-likelihood function $\ell(\theta)$:

$$I_n(\theta)_{r,s} = -E\left(\frac{\partial^2\ell}{\partial\theta_r\partial\theta_s}\right)$$

this definition corresponds to the expected Fisher information.

5.4 Basic Idea of the IMT

We are going in this section to simplify the general idea of the information matrix test as introduced by White (1982). Let us consider the density function $f(x_i, \theta)$ for individual observation and the data are independent, identically distribution so we have

$$\int f(x \mid \theta)dx = 1$$

and we consider $\ell(\theta) = \log f(x, \theta)$ to be the logarithm of a density function of x dependent upon p parameters θ, so the log-likelihood function in this case is

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta)$$

Now, as we defined the idea of the IMT to compare two different matrix of expected the first and second partial derivatives of the $\ell_n(\theta)$, we have

$$\frac{\partial\ell}{\partial\theta} = \int \frac{\partial f(x \mid \theta)}{\partial\theta}dx$$

$$= \int \frac{\partial \log f(x \mid \theta)}{\partial\theta} f(x \mid \theta)dx$$

$$= E\left(\frac{\partial \log(f(x \mid \theta))}{\partial\theta}\right) = 0 \qquad (5.1)$$

So, according to the ML method, we have

$$E\left(\frac{\partial \ell}{\partial \theta}\right) = 0.$$

Differentiating (5.1) again we get

$$0 = \int \frac{\partial^2 \log f(x \mid \theta)}{\partial \theta \partial \theta^T} f(x \mid \theta) dx + \int \frac{\partial \log f(x \mid \theta)}{\partial \theta} \frac{\partial \log f(x \mid \theta)}{\partial \theta^T} f(x \mid \theta) dx$$

So

$$E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}\right) + E\left(\frac{\partial \ell}{\partial \theta} \frac{\partial \ell}{\partial \theta^T}\right) = 0.$$

When the model is mis-specified, the above quantity will be not necessarily equal zero.

5.4.1 Asymptotic Distribution of $\hat{\theta}$

To more explain the idea of IMT we should looking for asymptotic distribution of estimated parameters. As we discussed the behaviour of the MLE under the wrong model in chapter 2, which Claeskens and Hjort (2008), pointed out the estimation the parameters of a given regression model. In the limit for each value of the parameter vector θ,

$$n^{-1} \ell_n(\theta) \to \int g(Y) \log f(Y \mid \theta) dY = E(\log f(Y \mid \theta))$$

where $g(Y)$ denoted to the true model and $f(Y \mid \theta)$ is the fitted model. Also, consider the Kullback-Leibler divergence (KL) from the true to the approximating model conditional on X, as (2.5). In this case $\hat{\theta} \to \theta^*$, where θ^* is the least false value (LF). Note that the least false value θ^* minimises the KL divergence (2.6), so, because the derivative of the KL is

$$E\left(\frac{\partial \log f(Y, \theta)}{\partial \theta}\right) = \int g(Y) \frac{\partial \log f(Y, \theta)}{\partial \theta} dY = 0.$$

Also, if we need define

$$J = -E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}\right)$$

and

$$K = \mathsf{var}\left(\frac{\partial \log f(Y, \theta)}{\partial \theta}\right) = E\left(\frac{\partial \ell}{\partial \theta} \frac{\partial \ell}{\partial \theta^T}\right)$$

these matrixes are identical when $g(Y) = \frac{\partial \log f(Y, \theta)}{\partial \theta}$ for all Y. As explained in Claeskens and Hjort (2008), the distribution of the $\hat{\theta}$, in this case from the central limit theorem there is convergence in distribution

$$\sqrt{n} \bar{U}_n \to U' \sim N_p(0, K)$$

where, $\bar{U} = n^{-1}\sum_{i=1}^{n} u(Y_i, \theta^*)$ and $u(Y_i, \theta^*) = \frac{\partial \log f(Y,\theta)}{\partial \theta} \mid_{\theta^*}$, which is leads to

$$\sqrt{n}(\hat{\theta} - \theta^*) \to J^{-1}U^{'} \sim N_p(0, J^{-1}KJ^{-1}).$$

So, we can say, the asymptotic MLE distribution under the null hypotheses H_0, in this case

$$\sqrt{n}\hat{\theta} \sim N(\theta_0, J^{-1})$$

where, θ_0 is the true value. And the asymptotic distribution of $\hat{\theta}$ under alternative hypotheses H_1 is

$$\sqrt{n}\hat{\theta} \sim N(\theta^*, J^{-1}KJ^{-1})$$

So, that is meaning $(J = K)$ if and only if when fitted the correct model (i.e. under H_0). This is the basis of the IMT which is looking for $J = -K$ if the model is correctly specified.

5.5 Fisher Information Matrix for Logistic Regression Model

We consider binary regression, where the outcome for individual i, $i = 1, \ldots, n$, is a random variable $Y_i = 1 \in \{0, 1\}$. Also $\Pr(Y_i \mid x_i) = \pi_i = \pi(\beta^T x_i)$ where x_i is a $p \times 1$ dimensional vector of covariates and β is a p-dimensional vector of parameters. It will be convenient to write $a_i = \beta^T x_i$ and ℓ_i to be the contribution to the log-likelihood ℓ from unit i.

We have

$$\ell(\beta) = \sum_{i=1}^{n} \ell_i(\beta) = \sum_{i=1}^{n} Y_i \log \pi_i + (1 - Y_i)\log(1 - \pi_i)$$

The p-dimensional likelihood equations $\partial\ell/\partial\beta = 0$ can be written:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{(Y_i - \pi_i)}{\pi_i(1 - \pi_i)}\right] \frac{\partial \pi_i}{\partial a_i} x_i = 0 \tag{5.2}$$

We can also derive the $p \times p$ matrix $\partial^2\ell/\partial\beta\partial\beta^T$ as:

$$\sum_{i=1}^{n} \left[\frac{(Y_1 - \pi_i)}{\pi_i(1 - \pi_i)} \frac{\partial^2 \pi_i}{\partial a_i^2} - \frac{(Y_1 - \pi_i)^2}{\pi_i^2(1 - \pi_i)^2} \left(\frac{\partial \pi_i}{\partial a_i}\right)^2\right] x_i x_i^T \tag{5.3}$$

In case of logistic regression model, let us consider the standard logistic regression model and for simplicity consider the case

$$\pi_i = \text{expit}(a_i),\ i = (1, 2, \ldots, n)$$

where $a_i = \alpha + \beta_1 x_{1i}$. To some writing in the following we write x_{1i} as x_i as the dimension of x_i is clear from the context. The first derivatives of the log likelihood are

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} (y_i - \pi_i),$$

and

$$\frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n} x_i (y_i - \pi_i)$$

So, we then have

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \frac{\partial}{\partial \alpha}\left[\frac{\partial \ell}{\partial \alpha}\right] = -\sum_{i=1}^{n}\left(\frac{\partial}{\partial \alpha}\left[\frac{\exp(\alpha+\beta_1 x_i)}{1+\exp(\alpha+\beta_1 x_i)}\right]\right)$$

$$= -\sum_{i=1}^{n}\left(\frac{\partial}{\partial i}\left(\text{expit}(a_i)\right)\frac{\partial i}{\partial \alpha}\right)$$

$$= -\sum_{i=1}^{n} \pi_i (1-\pi_i).$$

Similarly, the second derivative with β_1 is

$$\frac{\partial^2 \ell}{\partial \beta_1^2} = -\sum_{i=1}^{n} x_i^2 \pi_i (1-\pi_i)$$

and also, we have

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta_1} = -\sum_{i=1}^{n} x_i \pi_i (1-\pi_i).$$

Then, the Fisher's information matrix in this case is

$$I_n = -E\begin{bmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \beta_1} \\ \frac{\partial^2 \ell}{\partial \alpha \partial \beta_1} & \frac{\partial^2 \ell}{\partial \beta_1^2} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \pi_i(1-\pi_i) & \sum_{i=1}^{n} x_i \pi_i (1-\pi_i) \\ \sum_{i=1}^{n} x_i \pi_i (1-\pi_i) & \sum_{i=1}^{n} x_i^2 \pi_i (1-\pi_i) \end{bmatrix},$$

it is evaluated at the MLE $\hat{\beta}$.

5.6 Information Matrix Test (IMT) for Logistic Regression Model

5.6.1 The IM Test Procedure

The idea behind the information matrix test is that if the model is correctly specified then the quantity:

$$IM = \sum_{i=1}^{n} \left(\frac{\partial \ell_i}{\partial \beta} \frac{\partial \ell_i}{\partial \beta^T} \Big|_{\hat{\beta}} + \frac{\partial^2 \ell_i}{\partial \beta \partial \beta^T} \Big|_{\hat{\beta}} \right)$$

has zero mean. By using individual elements of the sums in (5.2) and (5.3) we can compute this quantity, for a general value of β, as the sum of:

$$\frac{\partial \ell_i}{\partial \beta} \frac{\partial \ell_i}{\partial \beta^T} + \frac{\partial^2 \ell_i}{\partial \beta \partial \beta^T} = \frac{(Y_i - \pi_i)}{\pi_i(1-\pi_i)} \frac{\partial^2 \pi_i}{\partial a_i^2} x_i x_i^T \tag{5.4}$$

where x_i is a $p \times 1$ dimensional vector and $a_i = \beta^T x_i$. We can test the null hypothesis that IM has zero mean by computing the variance of IM and then constructing a standard χ^2 statistic. The first step is to compute the variance of $n^{-\frac{1}{2}} \sum d_i$ where we write d_i for essentially the right hand side of (5.4):

$$\frac{(Y_i - \pi_i)}{\pi_i(1-\pi_i)} \frac{\partial^2 \pi_i}{\partial a_i^2} z_i$$

we will consider the logistic regression model, so $\pi = \text{expit}(\alpha + \beta^T x_i)$, and we have changed the $p \times p$ symmetric matrix $x_i x_i^T$ into a vector z_i in order to be able to use standard methods. As $x_i x_i^T$ is symmetric we do not wish to duplicate entries, so z_i is the q-dimensional vector of independent elements of $x_i x_i^T$. Usually this will be the $\frac{1}{2}p(p+1)$-dimensional vector

$$z_i^T = ([x_{11}, x_{21}, \ldots, x_{p1}], [x_{22}, x_{32}, \ldots, x_{p2}], \ldots, [x_{(p-1),(p-1)}, x_{p,(p-1)}], [x_{pp}])$$

where x_{st} is the $(s,t)^{th}$ element of $x_i x_i^T$, and we have supposed the subscript i for clarity.

In this case we have $\partial \pi_i / \partial a_i = \pi_i(1-\pi_i)$ and $\partial^2 \pi_i / \partial a_i^2 = \pi_i(1-\pi_i)(1-2\pi_i)$, and if we write:

$$\begin{aligned} d = n^{-\frac{1}{2}} \sum d_i &= n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{(Y_i - \pi_i)}{\pi_i(1-\pi_i)} \frac{\partial^2 \pi_i}{\partial a_i^2} z_i \\ &= n^{-\frac{1}{2}} \sum_{i=1}^{n} (Y_i - \pi_i)(1-2\pi) z_i \end{aligned}$$

then because the different terms are independent we obtain:

$$\Psi = \mathsf{var}(d) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\pi_i(1-\pi_i)}\left(\frac{\partial^2\pi_i}{\partial a_i^2}\right)^2 z_i z_i^T.$$

$$= \frac{1}{n}\sum_{i=1}^{n}\pi_i(1-\pi_i)(1-2\pi)^2 z_i z_i^T$$

which is a $q \times q$ dimensional matrix where $q = \frac{1}{2}p(p+1)$.

We should also note that if ∇D is defined as essentially the score, i.e.

$$\nabla D = n^{-\frac{1}{2}}\sum_{i=1}^{n}\frac{(Y_i-\pi_i)}{\pi_i(1-\pi_i)}\frac{\partial\pi_i}{\partial a_i}x_i$$

$$= n^{-\frac{1}{2}}\sum_{i=1}^{n}(Y_i-\pi_i)x_i$$

then the variance of ∇D is the $p \times p$ matrix Ω:

$$\Omega = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\pi_i(1-\pi_i)}\left(\frac{\partial\pi_i}{\partial a_i}\right)^2 x_i x_i^T$$

$$= \frac{1}{n}\sum_{i=1}^{n}\pi_i(1-\pi_i)x_i x_i^T$$

and the covariance of d and ∇D is the $q \times p$ matrix

$$\Delta = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\pi_i(1-\pi_i)}\left(\frac{\partial\pi_i}{\partial a_i}\right)\left(\frac{\partial^2\pi_i}{\partial a_i^2}\right) z_i x_i^T$$

$$= \frac{1}{n}\sum_{i=1}^{n}\pi_i(1-\pi_i)(1-2\pi)z_i x_i^T$$

Central limit arguments suggest that asymptotically $(d^T, (\nabla D)^T)$ is a $q+p$ dimensional normal variable. However, the IM test requires d to be evaluated at $\hat{\beta}$, $\hat{d}$, say, and at this value we know that $\nabla D = 0$. Consequently the variance of $\hat{d}$ is the variance of d conditional on $\nabla D = 0$ which is $\Psi - \Delta\Omega^{-1}\Delta^T$.

5.7 Information Matrix test (IMT_{DIAG})

The idea of the IM_{DIAG} test and IM test are the same,the only difference is that for the former the elements of z_i are just the diagonal elements of $x_i {x_i}^T$, so z_i is the p dimensional vector:

$$z_i^T = (x_{i1}^2, x_{i2}^2, \ldots, x_{ip}^2).$$

To explain the difference in size of vector z_i in the two cases of IM test and IM_{DIAG} test, let us consider a simple example. Suppose we have a symmetric matrix with elements $x_i x_i^T$ and 3×3 dimension as:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix},$$

where, $x_{rs} = x_{ri} x_{si}$. Then in the case of the IM test, the dimension of vector z_i^T is 1×6 and elements are :

$$z_i^T = [x_{11}, x_{12}, x_{13}, x_{22}, x_{23}, x_{33}],$$

whereas in the case of IM_{DIAG} test, z_i is the 3×1 dimensional vector:

$$z_i^T = [x_{11}, x_{22}, x_{33}].$$

5.8 Dimensional Matrix of IMT

As we discussed in previous section about the basic idea of the information matrix test, the main point is examine if $(J - K)$ is possibly 0. White (1982), discussed that, one of the procedures is change the dimensional of symmetric matrix $x_i x_i^T$ in (5.4) from $p \times p$ symmetric matrix to $q = \frac{1}{2} p(1 + p)$ vector. The idea is we do not wish to duplicate the elements of the matrix which allowed using the standard method. So, the q vector is $vec(x_i x_i^T)$. To more explain the behaviour of the dimensional of $x_i x_i^T$, let us consider simple example, we have $X = (1, x_1, x_2, x_3)$ as a covariates of the logistic regression model. Then the 4×4 symmetric matrix (XX^T)

$$XX^T = \begin{bmatrix} 1 & x_1 & x_2 & x_3 \\ x_1 & x_1^2 & x_1 x_2 & x_1 x_3 \\ x_2 & x_2 x_1 & x_2^2 & x_2 x_3 \\ x_3 & x_3 x_1 & x_3 x_2 & x_3^2 \end{bmatrix},$$

changed to 10 dimensional vector,

$$q = vec(XX^T) = (1, x_1, x_2, x_3, x_1^2, x_2 x_1, x_3 x_1, x_2^2, x_3 x_2, x_3^2)^T.$$

The issues of the dimension of XX^T has been the attention of many researchers: the elements of this matrix may have an effect on the covariance matrix of the IMT and may be some components are linear combinations of others leading to singularity of the estimated covariance matrix, this pointed out by White (1982) and Lin and Wel (1991).

The idea is to remove the duplicate elements in XX^T, and this require more elements to be removed. As we found above we must reduce the $p \times p$ symmetric matrix XX^T at least to a $q = \frac{1}{2}p(p+1)$ dimensional vector. However, in some models we need to do more giving $q < \frac{1}{2}p(p+1)$. For example, in case of polynomial regression if consider we have the function,

$$E(Y \mid X = x) = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

In this case we have $X = (1, x, x^2, x^3)$, and the matrix XX^T is

$$XX^T = \begin{bmatrix} 1 & x & x^2 & x^3 \\ x & x^2 & x^3 & x^4 \\ x^2 & x^3 & x^4 & x^5 \\ x^3 & x^4 & x^5 & x^6 \end{bmatrix},$$

and

$$vec(XX^T) = (1, x, x^2, x^3, x^2, x^3, x^4, x^4, x^5, x^6)^T.$$

Clearly, we can see there are more than one covariates are replicated, *i.e.* (x^2, x^3, x^4). Then, we need to reduce $vec(XX^T)$ in case of polynomial regression (i.e $q < \frac{1}{2}p(p+1)$), so, in this case the dimension is (5×1). Kuss (2002), discussed a new approach which reduced the elements of $vec(XX^T)$ to a vector, which contained only the diagonal elements of (XX^T) matrix and the comparison between IMT and IMT_{DIAG} has been appeared IMT_{DIAG} has reasonable behaviour. There adjustments are to allow for *exact* redundancy in the vectorised form of XX^T. Issues related to approximate redundancy also arise and will be addressed in the following chapters.

Chapter 6

Distribution of Moments of the IMT Statistic

6.1 The IMT Under missing covariates

We are interested in the distribution of IMT and hence the moments of this statistic. White (1982) introduced the test statistic as

$$d_g(y,\theta) = \frac{\partial \ell(y)}{\partial \theta_r}\frac{\partial \ell(y)}{\partial \theta_s} + \frac{\partial^2 \ell(y)}{\partial \theta_r \partial \theta_s}$$

where g ranges over appropriately chosen elements of the matrix and y will stand in place of the data: $g = 1, \ldots, q \leq \frac{1}{2}p(p+1)$, where $p = dim(\theta)$ and $r, s = 1, \ldots, p$. The IMT statistic is based on the q-vector

$$D_g(\hat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_g(y_i, \hat{\theta}_n); 1 \leq g \leq q$$

where $\hat{\theta}_n$ is the MLE under $\ell(\cdot)$, where $y_1, y_2, \ldots, y_n$ are the data. We assume that the y_i are independent and identically distribution.

6.2 The IMT Under missing covariates for Logistic Regression Model

In this part we will apply the procedure of the IMT statistic under missing covariates for a logistic regression model. If X_i is a p-dimensional vector of covariates drawn from normal distribution and Y_i is binary with

$$P(Y_i = 1 \mid X_i) = \text{expit}(\alpha + \beta^T X_i).$$

In the following we treat the simple case where the fitted model is

$$P(Y_i = 1 \mid X_i) = \text{expit}(\alpha + \beta_1 X_{1i})$$

for a scalar X_1 and that the true model has

$$P(Y_i = 1 \mid X_i) = \text{expit}(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i}),$$

where X_2 is also a scalar. We have the log-likelihood function contribution for the i^{th} element (Y_i, X_i) is

$$\ell(Y_i, X_i) = Y_i(\alpha + \beta^T X_i) - \log(1 + \exp(\alpha + \beta^T X_i))$$

and so,

$$\frac{\partial \ell_i}{\partial \alpha} = Y_i - \pi_i; \; \frac{\partial \ell_i}{\partial \beta_1} = (Y_i - \pi_i) X_{1i}$$

and note that we only consider fitting the model with X_1, even if the true model also includes X_2 ($i.e. \beta_2 \neq 0$). From this we get:

$$\frac{\partial^2 \ell_i}{\partial \theta \partial \theta^T} = \begin{bmatrix} -\pi_i(1-\pi_i) & -\pi_i(1-\pi_i)X_i \\ -\pi_i(1-\pi_i)X_i & -\pi_i(1-\pi_i)X_i^2 \end{bmatrix}$$

Also,

$$\frac{\partial \ell_i}{\partial \theta} \frac{\partial \ell_i}{\partial \theta^T} = \begin{bmatrix} (Y_i-\pi_i)^2 & (Y_i-\pi_i)^2 X_i \\ (Y_i-\pi_i)^2 X_i & (Y_i-\pi_i)^2 X_i^2 \end{bmatrix}$$

using,

$$(Y_i - \pi_i)^2 - \pi_i(1-\pi_i) = (Y_i - \pi_i)(1 - 2\pi_i),$$

as Y_i^2 is Y_i, and so we get that

$$d_g(y_i, \theta) = (Y_i - \pi_i)(1 - 2\pi_i) \begin{bmatrix} 1 \\ X_i \\ X_i^2 \end{bmatrix}.$$

6.3 An Alternative Formulae of Variance

In this section we work out the variance of d. In this part we are interested to find a formulae of the variance of d statistic, even when the model is mis-specified. To perform the IMT we need to find the mean and variance of

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_{gi}$$

Under H_0 $E(d_{gi}) = 0$, and so the IMT could be written as

$$T^T \mathsf{var}(T)^{-1} T$$

which will have a χ^2-distribution on rank($\mathsf{var}(T)$) d.f. as T is asymptotically Normal. However, the test statistic has to be evaluated at the MLE $\hat{\theta}$ and this introduces a complication.
The MLE $\hat{\theta}$ is the solution to

$$S = \frac{1}{\sqrt{n}} \nabla \ell = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla \ell_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - \pi_i) \begin{bmatrix} 1 \\ X_i \end{bmatrix} = 0.$$

The expression for T is

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (y_i - \pi_i)(1 - 2\pi_i) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix}$$

and this is clearly going to be highly correlated with S. Therefore, the appropriate variance for the IMT is $\mathsf{var}(T \mid S = 0)$. As T and S are sums of independent elements, the Central limit Theorem implies that $(T, S)^T$ is asymptotically Normal and so we can use

$$\mathsf{var}(T \mid S = 0) = \mathsf{var}(T) - \mathsf{cov}(T, S)\mathsf{var}(S)^{-1}\mathsf{cov}(T, S)^T. \tag{6.1}$$

To work out $\mathsf{var}(T \mid S = 0)$, so, in this case we can write

$$\mathsf{var}(T) = \mathsf{var}([d_{g1} + d_{g2} + \cdots + d_{gn}]/\sqrt{n}) = \mathsf{var}(d_{g1}),$$

and similarly

$$\mathsf{var}(S) = \mathsf{var}(\nabla \ell_1),$$

and

$$\mathsf{cov}(T, S) = \mathsf{cov}(d_{g1}, \nabla \ell_1).$$

6.3.1 The Variance of IMT Under Missing Covariates for Logistic Regression Model

We now need to find expressions for $\mathsf{var}(d_{g1})$, $\mathsf{var}(\nabla \ell_1)$ and $\mathsf{cov}(d_{g1}, \nabla \ell_1)$

We already have that

$$d_g = (y_i - \pi_i)(1 - 2\pi_i) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix}$$

and

$$\nabla \ell_i = (y_i - \pi_i) \left[\begin{array}{c} 1 \\ x_i \end{array} \right]$$

so, the variance is

$$\mathsf{var}(d_g) = E(d_g d_g^T) - E(d_g)E(d_g^T)$$

and we have

$$d_g d_g^T = (y - \pi)^2 (1 - 2\pi)^2 \left[\begin{array}{ccc} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{array} \right]$$

taking expectation $E_{Y|X}$ we obtain

$$E(d_{g1}) = E_X \left[(\pi_t - \pi)(1 - 2\pi) \left[\begin{array}{c} 1 \\ x_i \\ x_i^2 \end{array} \right] \right] \tag{6.2}$$

and,

$$E(d_{g1} d_{g1}^T) = E_X \left[(\pi_t(1 - 2\pi) + \pi^2)(1 - 2\pi)^2 \left[\begin{array}{ccc} 1 & X & X^2 \\ X & X^2 & X^3 \\ X^2 & X^3 & X^4 \end{array} \right] \right]. \tag{6.3}$$

Now we need to compute $\mathsf{cov}(d_g, \nabla \ell)$. In fact $E(\nabla \ell) = 0$, not only if the model is correct but also when evaluated at the least false value θ^*, so in this case

$$\mathsf{cov}(d_{g1}, \nabla \ell_1) = E(d_g \nabla \ell)^T.$$

and we have

$$d_{g_1} \nabla \ell_1^T = (y - \pi)(1 - 2\pi) \left[\begin{array}{c} 1 \\ x_i \\ x_i^2 \end{array} \right] (y - \pi) \left[\begin{array}{cc} 1 & x_i \end{array} \right]$$

$$= (y - \pi)^2 (1 - 2\pi) \left[\begin{array}{cc} 1 & x_i \\ x_i & x_i^2 \\ x_i^2 & x_i^3 \end{array} \right]$$

then,

$$E(d_{g_1} \nabla \ell_1^T) = E_X \left[(\pi_t(1 - 2\pi) + \pi^2)(1 - 2\pi) \left[\begin{array}{cc} 1 & X \\ X & X^2 \\ X^2 & X^3 \end{array} \right] \right]. \tag{6.4}$$

Now we will work out $\mathsf{var}(\nabla \ell)$, as before, since $E(\nabla \ell) = 0$, so

$$\mathsf{var}(\nabla \ell_1) = E(\nabla \ell \nabla \ell^T)$$

$$= E_X E_{Y|X} \left[\begin{array}{cc} (Y - \pi)^2 & (Y - \pi)^2 X \\ (Y - \pi)^2 X & (Y - \pi)^2 X^2 \end{array} \right]$$

and note that

$$E_{Y|X}(Y-\pi)^2 = E_{Y|X}(Y(1-2\pi)+\pi^2) = \pi_t(1-2\pi)+\pi^2,$$

where, π_t is $E(Y)$ under the true model. So,

$$E(\nabla\ell\nabla\ell^T) = E_X\left[\begin{array}{cc} \pi_t(1-2\pi)+\pi^2 & (\pi_t(1-2\pi)+\pi^2)X \\ (\pi_t(1-2\pi)+\pi^2)X & (\pi_t(1-2\pi)+\pi^2)X^2 \end{array}\right]. \tag{6.5}$$

Hence, the required variance (6.1)

$$E(d_g d_g^T) - E(d_g)E(d_g^T) - E(d_g\nabla\ell^T)E(\nabla\ell\nabla\ell^T)^{-1}E((\nabla\ell)d_g^T) \tag{6.6}$$

and we have expressions for each component from (6.2), (6.3), (6.4) and (6.5) We need to evaluate these components by simulation.

6.3.2 The Dispersion Matrix Under Wrong Model

We are interested to compute the $\mathsf{var}(T \mid S=0)$, even when the wrong model has been fitted. We will compute each of the components of this variance separately. We see from section 6.3.1 that we need to evaluate, e.g

$$E(d) = E_X\left((\pi_t-\pi)(1-2\pi)\left[\begin{array}{c} 1 \\ X \\ X^2 \end{array}\right]\right)$$

and also,

$$E(dd^T) = E_X\left([\pi_t(1-2\pi)+\pi^2](1-2\pi)^2\left[\begin{array}{ccc} 1 & X & X^2 \\ X & X^2 & X^3 \\ X^2 & X^3 & X^4 \end{array}\right]\right).$$

This cannot be done analytically so we simulate 5000 values of X and replace the $E(d)$ by the mean of these 5000 values. In evaluating π_t we use the values of the parameters α_t, β_{1t} and β_{2t}. What do we use for π? We need to evaluate $\pi(\alpha,\beta_1)$ at the least false values α^* and β_1^* for α and β_1. So, e.g, the first element of $E(d)$ is found by simulation from

$$E_X[(\text{expit}(\alpha_t+\beta_{t1}X_1+\beta_{t2}X_2) - \text{expit}(\alpha^*+\beta_1^*X_1))(1-2\text{expit}(\alpha^*+\beta_1^*X_1))]$$

where,

$$\alpha^* = \frac{\alpha_t+\beta_{t2}(\mu_2-\rho\mu_1)}{\sqrt{1+k^2\beta_{t2}^2\sigma^2(1-\rho^2)}},$$

$$\beta_1^* = \frac{\beta_{t1} + \rho\beta_{t2}}{\sqrt{1 + k^2\beta_{t2}^2\sigma^2(1-\rho^2)}}$$

and X draw from bivariate normal distribution with $\mu = (\mu_1, \mu_2)$, and $\sigma_1^2 = \sigma_2^2$. The formulae of the least false values α^* and β_1^* have calculated in chapter 2.

6.4 Empirical Variance of IMT

The expression in (6.6) is the variance V of d at $\hat{\theta}$ but we need an estimate, $\hat{V}$. If we have a sample $\{(y_i, x_{i1}) \mid i = 1, \ldots, n\}$ how can we estimate V consistently? One candidate would be to compute

$$d_i = (y_i - \hat{\pi}_i)(1 - 2\hat{\pi}_i) \begin{bmatrix} 1 \\ x_i \\ x_i^2 \end{bmatrix} \quad i = 1, \ldots, n$$

and

$$\nabla \ell_i = (y_i - \hat{\pi}_i) \begin{bmatrix} 1 \\ x_i \end{bmatrix} \quad i = 1, \ldots, n$$

where, $\hat{\pi}_i$ is the fitted value from the model with just x_1. Now compute

$$\hat{W}_n = \frac{1}{n}\sum_{i=1}^{n} d_i d_i^T - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} d_i^T\right)$$

and

$$\hat{B}_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\pi})^2 \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix},$$

$$\hat{C}_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\pi})^2(1 - 2\hat{\pi}_i) \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \\ x_i^2 & x_i^3 \end{bmatrix}$$

Then use

$$\hat{V} = \hat{W}_d - \hat{C}_n \hat{B}_n^{-1} \hat{C}_n^{\,T} \tag{6.7}$$

as an estimate of V, we will assess this by simulation.

6.5 Simulation Study

This simulation examines the correctness of the form of the dispersion matrix V in (6.6) and (6.7).

6.5.1 Design of Simulation

To achieve the aim of this simulation, we will consider a logistic regression model which has two covariates draw from bivariate normal distribution with mean zero and covariance matrix Σ as:

$$\pi_t = \text{expit}(\alpha_t + \beta_{t1}x_1 + \beta_{t2}x_2)$$

and the fitted model is

$$\pi = \text{expit}(\alpha + \beta_1 x_1)$$

- Apply in two cases of logistic model,

- The fitted is the true logistic model (i.e $\beta_{t2} = 0$)

- The fitted model is mis-specified (i.e $\beta_{t2} \neq 0$).

- Use two cases of variance ($\sigma_1^2 = \sigma_2^2 = 0.2$), ($\sigma_1^2 = \sigma_2^2 = 2$) and correlation $\rho = 0.1$.

- We choose some different components of parameters α_t, β_{t1} and β_{t2} to calculate π_t.

- We compute the least false values α^* and β_1^* by formulae to calculate π.

- We compute the true variance by simulateing d_i and take the variance to be $\mathsf{var}(\sqrt{n}\bar{d}) = V_{tr}$.

- We compute the theoretical variance $\mathsf{var}(d) = V_T$ at the least false value and calculate $E(d_1)$ and $E(d_1 d_1^T)$ as described in section 6.3.2.

- Finally, for each simulation we compute the empirical variance V_E and take the mean over the simulations.

- We make comparison between the diagonal elements of dispersion matrix V_E, V_T vs. V_{tr} respectively.

- Apply on different sample size $n = 500, 1000, 5000$ and $N = 5000$ number of simulations.

6.5.2 Results and Discussion

Results are reported in tables, which shows the diagonal elements of the variance matrix: V_E denotes the empirical variance, V_T denotes the theoretical variance and V_{tr} denotes the true variance. The true parameters appear as α_t, β_{t1}, and β_{t2}; Rn_E and Rn_T denote to the rank of the covariance matrix empirical and theoretical respectively. The Ratio R_E and R_T are $\sqrt{\frac{V_E}{V_{tr}}}$, $\sqrt{\frac{V_T}{V_{tr}}}$ respectively. $S.D(\pi_t)$ denotes the standard deviation over a sample where π_t is the true model. In our simulation we consider two covariates, so in this case the dispersion matrix of d is a 3×3 dimensional matrix.

6.5.3 Results Under True Model

Table 6.1 and table 6.2, shows the results of simulation, which appeared the diagonal elements of matrix V, the empirical version and theoretical form comparing with true variance, which use $\rho = 0.1$ in case of $\sigma_1^2 = \sigma_2^2 = 0.2$ and $\sigma_1^2 = \sigma_2^2 = 2$ respectively by sample size $n = 500$. Table 6.3 and Table 6.4, reported the results by sample size $n = 1000$, with equal variance $\sigma_1^2 = \sigma_2^2$, 0.2 and 2 respectively. Table 6.5 and Table 6.6, shows the results in case of sample size $n = 5000$ and with variance 0.2 and 2. We can see clearly, that all diagonal elements appeared small in value in all different cases of sample size and variance. The first element was much closer to zero than of the rest. In almost cases the results appeared reasonable ratio which is meaning the theoretical variance and empirical variance are close to the true value. There are some slightly strange ratio almost in case of sample size $n = 500$, the reason may be affected by small value of standard deviation of π_t $S.D(\pi_t)$, otherwise the ratio is close to one. In case of sample size $n = 1000$ and $n = 5000$, the behaviour of results shows almost the same pattern, with the ratio close to one and that is meaning the formulae of the variance works well. In a few cases with small values of $S.D(\pi_t)$ which affected on the ratio where the first two elements were more sensitive. Overall, we have reasonable results to say that, the alternative formulae of variance works well and the two first elements still more sensitive which appeared tend to zero.

Diagonal component of variance IMT and Ratio										
α_t	β_{t1}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.61	0.02	3	2	$6.8095e^{-07}$	$1.0641e^{-09}$	$7.0710e^{-07}$	0.98	0.39
-	-	-	-	-	-	$4.3965e^{-04}$	$2.9824e^{-04}$	$4.6150e^{-04}$	0.97	0.80
-	-	-	-	-	-	$6.5530e^{-04}$	$6.0731e^{-04}$	$6.9025e^{-04}$	0.97	0.94
0.80	0.50	0.69	0.04	3	3	$1.1742e^{-05}$	6.4247^{-06}	$1.1641e^{-05}$	1.01	0.74
-	-	-	-	-	-	$8.2395e^{-04}$	$7.6936e^{-04}$	$8.1842e^{-04}$	1.01	0.97
-	-	-	-	-	-	$2.1784e^{-03}$	$2.3987e^{-03}$	$2.2081e^{-03}$	0.99	1.04
1.20	2.20	0.73	0.16	3	3	$6.0873e^{-04}$	$6.5146e^{-04}$	$6.5197e^{-04}$	0.97	0.99
-	-	-	-	-	-	$4.5568e^{-03}$	$4.9824e^{-03}$	$4.8376e^{-03}$	0.97	1.01
-	-	-	-	-	-	$1.7132e^{-03}$	$1.7935e^{-03}$	$1.8288e^{-03}$	0.97	0.99
2.30	0.20	0.90	0.007	3	2	$2.2587e^{-06}$	$4.6206e^{-08}$	$2.5916e^{-06}$	0.93	0.13
-	-	-	-	-	-	$3.4470e^{-05}$	$7.0526e^{-06}$	$3.8482e^{-05}$	0.95	0.42
-	-	-	-	-	-	$3.6847e^{-03}$	$4.0920e^{-03}$	$4.0201e^{-03}$	0.96	1.01
0.20	2.30	0.53	0.22	3	3	$3.1828e^{-04}$	$3.3163e^{-04}$	$3.2639e^{-04}$	0.99	1.01
-	-	-	-	-	-	$5.7410e^{-03}$	$6.1171e^{-03}$	$6.0210e^{-03}$	0.98	1.01
-	-	-	-	-	-	$1.7934e^{-03}$	$1.9713e^{-03}$	$1.8774e^{-03}$	0.98	1.02

Table 6.1: Simulation results of the variance (V_{tr}) comparing with empirical (V_E) and theoretical variance (V_T) in case of fitted true model, using different values of true parameters by generated variables from bivariate normal distribution with sample size $n = 500$ and $\sigma_1^2 = \sigma_2^2$=0.2

Diagonal component of variance IMT and Ratio										
α_t	β_{t1}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.57	0.08	3	3	$1.2933e^{-05}$	$9.2752e^{-06}$	$1.3667e^{-05}$	0.97	0.82
-	-	-	-	-	-	$2.2703e^{-02}$	$2.3813e^{-02}$	$2.3300e^{-02}$	0.99	1.01
-	-	-	-	-	-	$1.3504e^{-01}$	$1.4015e^{-01}$	$1.3681e^{-01}$	0.99	1.01
0.80	0.50	0.68	0.14	3	3	$2.4323e^{-04}$	$2.4817e^{-04}$	$2.4979e^{-04}$	0.99	0.99
-	-	-	-	-	-	$4.2714e^{-02}$	$4.6277e^{-02}$	$4.3906e^{-02}$	0.99	1.02
-	-	-	-	-	-	$1.9446e^{-01}$	$2.1171e^{-01}$	$2.0659e^{-01}$	0.97	1.01
1.20	2.20	0.63	0.36	3	3	$1.4666e^{-03}$	$1.6904e^{-03}$	$1.6126e^{-03}$	0.95	1.02
-	-	-	-	-	-	$1.8138e^{-02}$	$2.0199e^{-02}$	$1.9872e^{-02}$	0.96	1.01
-	-	-	-	-	-	$2.5199e^{-02}$	$2.9136e^{-02}$	$2.8595e^{-02}$	0.94	1.01
2.30	0.20	0.90	0.02	3	3	$1.4526e^{-05}$	$4.5251e^{-06}$	$1.5026e^{-05}$	0.98	0.55
-	-	-	-	-	-	$1.3613e^{-03}$	$7.8328e^{-04}$	$1.4260e^{-03}$	0.98	0.74
-	-	-	-	-	-	$3.2184e^{-01}$	$3.4163e^{-01}$	$3.5742e^{-01}$	0.95	0.98
0.20	2.30	0.52	0.38	3	3	$1.4890e^{-03}$	$1.6581e^{-03}$	$1.5810e^{-03}$	0.97	1.02
-	-	-	-	-	-	$1.6339e^{-02}$	$1.7567e^{-02}$	$1.7492e^{-02}$	0.97	1.01
-	-	-	-	-	-	$1.5041e^{-02}$	$1.6676e^{-02}$	$1.5721e^{-02}$	0.98	1.02

Table 6.2: Simulation results of the variance (V_{tr}) comparing with empirical (V_E) and theoretical variance (V_T) in case of fitted true model, using different values of true parameters by generated variables from bivariate normal distribution with sample size $n = 500$ and σ_1^2=σ_2^2=2

Diagonal component of variance IMT and Ratio										
α_t	β_{t1}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.57	0.02	3	2	$3.4865e^{-07}$	$9.4241e^{-08}$	3.7802^{-07}	0.96	0.50
-	-	-	-	-	-	$3.6625e^{-04}$	$2.5895e^{-04}$	$3.6895e^{-04}$	0.99	0.84
-	-	-	-	-	-	$6.0920e^{-04}$	$5.6444e^{-04}$	$6.2892e^{-04}$	0.98	0.95
0.80	0.50	0.69	0.05	3	3	$9.1613e^{-06}$	5.6977^{-06}	8.9828^{-06}	1.01	0.80
-	-	-	-	-	-	$7.9258e^{-04}$	$6.9865e^{-04}$	$8.0062e^{-04}$	0.99	0.93
-	-	-	-	-	-	$2.2189e^{-03}$	$2.0362e^{-03}$	$2.2954e^{-03}$	0.98	0.94
1.20	2.20	0.73	0.17	3	3	$6.2529e^{-04}$	$6.4373e^{-04}$	$6.2290e^{-04}$	1.01	1.02
-	-	-	-	-	-	$4.7051e^{-03}$	$4.9757e^{-03}$	$4.5926e^{-03}$	1.01	1.04
-	-	-	-	-	-	$1.7770e^{-03}$	$1.8703e^{-03}$	$1.8523e^{-03}$	0.98	1.01
2.30	0.20	0.90	0.007	3	2	$8.5391e^{-07}$	$5.1853e^{-08}$	$7.9321e^{-07}$	1.03	0.26
-	-	-	-	-	-	$2.1094e^{-05}$	$7.9454e^{-06}$	$2.0874e^{-05}$	1.01	0.62
-	-	-	-	-	-	$4.0230e^{-03}$	$4.5576e^{-03}$	$4.2742e^{-03}$	0.97	1.03
0.20	2.30	0.53	0.21	3	3	$3.2437e^{-04}$	$3.4145e^{-04}$	$3.1786e^{-04}$	1.01	1.03
-	-	-	-	-	-	$5.9352e^{-03}$	$6.3223e^{-03}$	$6.1501e^{-03}$	0.98	1.01
-	-	-	-	-	-	$1.9006e^{-03}$	$2.0843e^{-03}$	$1.9091e^{-03}$	0.99	1.04

Table 6.3: Simulation results of the variance (V_{tr}) comparing with empirical (V_E) and theoretical variance (V_T) in case of fitted true model, using different values of true parameters by generated variables from bivariate normal distribution with sample size $n = 1000$ and σ_1^2=σ_2^2=0.2

Diagonal component of variance IMT and Ratio										
α_t	β_{t1}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.57	0.08	3	3	$1.1287e^{-05}$	$8.7035e^{-06}$	$1.1311e^{-05}$	0.99	0.88
-	-	-	-	-	-	$2.2973e^{-02}$	$2.3067e^{-02}$	$2.2678e^{-02}$	1.01	1.01
-	-	-	-	-	-	$1.3841e^{-01}$	$1.4018e^{-01}$	$1.3834e^{-01}$	1.00	1.01
0.80	0.50	0.67	0.14	3	3	$2.4038e^{-04}$	$2.4441e^{-04}$	$2.4915e^{-04}$	0.98	0.99
-	-	-	-	-	-	$4.3783e^{-02}$	$4.4257e^{-02}$	$4.4706e^{-02}$	0.99	0.99
-	-	-	-	-	-	$1.9858e^{-01}$	$1.9486e^{-01}$	$1.0478e^{-01}$	0.98	0.98
1.20	2.20	0.64	0.35	3	3	1.5709^{-03}	$1.6876e^{-03}$	$1.6469e^{-03}$	0/98	1.01
-	-	-	-	-	-	$1.9049e^{-02}$	$2.0225e^{-02}$	$2.0199e^{-02}$	0.97	1.00
-	-	-	-	-	-	$2.6726e^{-02}$	$2.9664e^{-02}$	$2.7877e^{-02}$	0.98	1.03
2.30	0.20	0.90	0.02	3	3	$9.5367e^{-06}$	$4.8114e^{-06}$	$9.9900e^{-06}$	0.98	0.69
-	-	-	-	-	-	$1.1285e^{-03}$	$8.4254e^{-04}$	$1.1651e^{-04}$	0.98	0.85
-	-	-	-	-	-	$3.4869e^{-01}$	$3.5731e^{-01}$	$3.5686e^{-01}$	0.99	1.00
0.20	2.30	0.51	0.37	3	3	$1.5825e^{-03}$	$1.6740e^{-03}$	$1.6475e^{-03}$	0.98	1.01
-	-	-	-	-	-	$1.7080e^{-02}$	$1.7845e^{-02}$	$1.7148e^{-02}$	0.98	1.02
-	-	-	-	-	-	$1.6076e^{-02}$	$1.7015e^{-02}$	$1.6955e^{-02}$	0.97	1.01

Table 6.4: Simulation results of the variance (V_{tr}) comparing with empirical (V_E) and theoretical variance (V_T) in case of fitted true model, using different values of true parameters by generated variables from bivariate normal distribution with sample size $n = 1000$ and σ_1^2=σ_2^2=2

Diagonal component of variance IMT and Ratio										
α_t	β_{t1}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.57	0.02	3	2	$1.4467e^{-07}$	$1.0852e^{-07}$	$1.4564e^{-07}$	0.99	0.86
-	-	-	-	-	-	$3.0244e^{-04}$	2.0783^{-04}	3.0551^{-04}	0.99	1.00
-	-	-	-	-	-	$6.2497e^{-04}$	$5.7726e^{-04}$	$5.9654e^{-04}$	0.99	1.02
0.80	0.50	0.68	0.04	3	3	$6.6964e^{-06}$	6.4494^{-06}	6.7141^{-06}	0.99	0.98
-	-	-	-	-	-	$7.5095e^{-04}$	$7.4655e^{-04}$	7.5835^{-04}	0.99	0.99
-	-	-	-	-	-	$2.2465e^{-03}$	$2.5739e^{-03}$	$2.2774e^{-03}$	0.99	1.06
1.20	2.20	0.73	0.16	3	3	$6.4308e^{-04}$	$6.3806e^{-04}$	$6.4243e^{-04}$	0.99	1.00
-	-	-	-	-	-	$4.8294e^{-03}$	$4.8693e^{-03}$	$4.8611e^{-03}$	0.99	1.00
-	-	-	-	-	-	$1.8397e^{-03}$	$1.8400e^{-03}$	$1.8755e^{-03}$	0.99	0.99
2.30	0.20	0.91	0.007	3	2	$1.4528e^{-07}$	$4.7059e^{-08}$	$4.7059e^{-08}$	0.99	0.56
-	-	-	-	-	-	$9.7290e^{-06}$	$7.2109e^{-06}$	$7.2109e^{-06}$	0.99	0.85
-	-	-	-	-	-	$4.3046e^{-03}$	$4.1416e^{-03}$	$4.1416e^{-03}$	1.00	0.98
0.20	2.30	0.53	0.21	3	3	$3.2313e^{-04}$	$3.0604e^{-04}$	$3.1606e^{-04}$	1.01	0.98
-	-	-	-	-	-	$6.0494e^{-03}$	$5.9242e^{-03}$	$5.9321e^{-03}$	1.01	0.99
-	-	-	-	-	-	$1.9320e^{-03}$	$1.7989e^{-03}$	$1.9265e^{-03}$	1.00	0.97

Table 6.5: Simulation results of the variance (V_{tr}) comparing with empirical (V_E) and theoretical variance (V_T) in case of fitted true model, using different values of true parameters by generated variables from bivariate normal distribution with sample size $n = 5000$ and σ_1^2=σ_2^2=0.2

Diagonal component of variance IMT and Ratio										
α_t	β_{t1}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.57	0.08	3	3	$9.6733e^{-06}$	$8.8645e^{-06}$	9.4912^{-06}	1.01	0.97
-	-	-	-	-	-	$2.3043e^{-02}$	$2.2925e^{-02}$	$2.2485e^{-02}$	1.01	1.01
-	-	-	-	-	-	$1.4050e^{-01}$	$1.3758e^{-01}$	$1.3916e^{-01}$	1.00	0.99
0.80	0.50	0.67	0.14	3	3	$2.3936e^{-04}$	$2.4375e^{-04}$	$2.4375e^{-04}$	0.99	0.99
-	-	-	-	-	-	$4.4519e^{-02}$	$4.5007e^{-02}$	$4.4951e^{-02}$	0.99	1.00
-	-	-	-	-	-	$2.0393e^{-01}$	$2.0124e^{-01}$	$2.0554e^{-01}$	0.99	0.99
1.20	2.20	0.63	0.36	3	3	1.6684^{-03}	$1.6768e^{-03}$	$1.7178e^{-03}$	0.99	0.99
-	-	-	-	-	-	$1.9874e^{-02}$	$2.0214e^{-02}$	$2.0423e^{-02}$	0.99	0.99
-	-	-	-	-	-	$2.8469e^{-02}$	$2.7480e^{-02}$	$2.8812e^{-02}$	0.99	0.98
2.30	0.20	0.91	0.02	3	3	$6.0942e^{-06}$	$5.0654e^{-06}$	$6.2226e^{-06}$	0.99	0.90
-	-	-	-	-	-	$9.7134e^{-04}$	$8.7521e^{-04}$	$9.8237e^{-04}$	0.99	0.94
-	-	-	-	-	-	$3.7084e^{-01}$	$3.8834e^{-01}$	$3.7508e^{-01}$	0.99	1.01
0.20	2.30	0.53	0.37	3	3	$1.6593e^{-03}$	$1.6640e^{-03}$	$1.6922e^{-03}$	0.99	0.99
-	-	-	-	-	-	$1.7670e^{-02}$	$1.7645e^{-02}$	$1.8321e^{-02}$	0.98	0.98
-	-	-	-	-	-	$1.6895e^{-02}$	$1.6909e^{-02}$	$1.7285e^{-02}$	0.99	0.99

Table 6.6: Simulation results of the variance (V_{tr}) comparing with empirical (V_E) and theoretical variance (V_T) in case of fitted true model, using different values of true parameters by generated variables from bivariate normal distribution with sample size $n = 5000$ and σ_1^2=σ_2^2=2

Diagonal component of variance IMT and Ratio											
α_t	β_{t1}	β_{t2}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.2	0.57	0.03	3	3	$7.5599e^{-07}$	$1.3361e^{-07}$	$7.8920e^{-07}$	0.98	0.41
-	-	-	-	-	-	-	$4.7254e^{-04}$	$3.3482e^{-04}$	$4.9968e^{-04}$	0.97	0.82
-	-	-	-	-	-	-	$6.6070e^{-04}$	$5.5405e^{-04}$	$6.8594e^{-04}$	0.98	0.90
0.80	0.50	0.4	0.68	0.06	3	3	$1.4161e^{-05}$	$8.2973e^{-06}$	$1.3967e^{-05}$	1.01	0.77
-	-	-	-	-	-	-	$9.3914e^{-04}$	$9.2617e^{-04}$	$9.2278e^{-04}$	1.01	1.00
-	-	-	-	-	-	-	$2.1530e^{-03}$	$2.2723e^{-03}$	$2.1456e^{-03}$	1.00	1.03
1.20	2.20	0.8	0.73	0.18	3	3	$6.0838e^{-04}$	$6.4471e^{-04}$	$6.2340e^{-04}$	0.99	1.02
-	-	-	-	-	-	-	$4.6345e^{-03}$	$4.9022e^{-03}$	$4.6574e^{-03}$	0.99	1.02
-	-	-	-	-	-	-	$1.7018e^{-03}$	$1.8214e^{-03}$	$1.79524e^{-03}$	0.97	1.01
2.30	0.20	1	0.90	0.04	3	2	$3.4057e^{-06}$	$2.4924e^{-07}$	$3.4776e^{-07}$	0.99	0.27
-	-	-	-	-	-	-	$5.0822e^{-05}$	$1.9222e^{-05}$	$5.2125e^{-05}$	0.99	0.61
-	-	-	-	-	-	-	$3.7366e^{-03}$	$4.3500e^{-03}$	$3.9682e^{-03}$	0.97	1.04
0.20	2.30	1.2	0.53	0.24	3	3	$3.1353e^{-04}$	$3.1952e^{-04}$	$3.1275e^{-04}$	1.00	1.01
-	-	-	-	-	-	-	$5.7344e^{-03}$	$6.1732e^{-03}$	$5.8063e^{-03}$	0.99	1.03
-	-	-	-	-	-	-	$1.7897e^{-03}$	$1.9324e^{-03}$	$1.7948e^{-03}$	0.99	1.03

Table 6.7: Simulation results of the variance (V_{tr}) comparing with empirical variance (V_E) and theoretical variance (V_T) in case of fitted missing covariates model, using different values of parameters by generated variables from bivariate normal distribution with sample size $n = 500$ and σ_1^2=σ_2^2=0.2

6.5.4 Results Under Missing Covariate Model

In this part we consider the results when the missing covariate logistic model has been fitted. That is meaning when the variance of IMT computed under H_1 and uses the least false values. The results of different case of sample size and variance showed in several tables. Table (6.7) and Table (6.8) shows the results of sample size $n = 500$. Table (6.9) and Table (6.10), shows the results of sample size $n = 1000$. Lastly, Table (6.11) and Table (6.12), shows the results of sample size 5000. In general, the behaviour of ratio appeared the same behaviour which found in case of $\beta_{2t} = 0$, all cases of different variance and sample size appeared reasonable ratio which is close to one. A few cases shows low ratio, the reason is as discussed before concerning to the small value of $S.D(\pi t)$.

Diagonal component of variance IMT and Ratio												
α_t	β_{t1}	β_{t2}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.2	0.57	0.1	3	3	$1.5955e^{-05}$	$1.0720e^{-05}$	$1.6529e^{-05}$	0.98	0.81
-	-	-	-	-	-	-	$2.4823e^{-02}$	$2.4484e^{-02}$	$2.5906e^{-02}$	0.98	0.97
-	-	-	-	-	-	-	$1.4327e^{-01}$	$1.3571e^{-01}$	$1.4843e^{-01}$	0.98	0.96
0.80	0.50	0.4	0.66	0.18	3	3	$2.3400e^{-04}$	$2.5316e^{-04}$	$2.3730e^{-04}$	0.99	1.03
-	-	-	-	-	-	-	$4.3984e^{-02}$	$5.1198e^{-02}$	$4.5628e^{-02}$	0.98	1.05
-	-	-	-	-	-	-	$1.9280e^{-01}$	$2.3559e^{-01}$	$2.0038e^{-01}$	0.98	1.08
1.20	2.20	0.8	0.62	0.36	3	3	$1.2620e^{-03}$	$1.4252e^{-03}$	$1.3850e^{-03}$	0.95	1.01
-	-	-	-	-	-	-	$2.1819e^{-02}$	$2.3741e^{-02}$	$2.3411e^{-02}$	0.97	1.01
-	-	-	-	-	-	-	$3.2250e^{-02}$	$3.5432e^{-02}$	$3.6766e^{-02}$	0.94	0.98
2.30	0.20	1	0.83	0.17	3	2	$3.8371e^{-05}$	$1.3283e^{-05}$	$4.5381e^{-05}$	0.92	0.54
-	-	-	-	-	-	-	$4.9410e^{-03}$	$3.7476e^{-03}$	$5.6578e^{-03}$	0.93	0.81
-	-	-	-	-	-	-	$3.2318e^{-01}$	$3.2733e^{-01}$	$3.4807e^{-01}$	0.96	0.97
0.20	2.30	1.2	0.50	0.39	3	3	$1.1036e^{-03}$	$1.2138e^{-03}$	$1.1557e^{-03}$	0.98	1.02
-	-	-	-	-	-	-	$2.5064e^{-02}$	$2.6421e^{-02}$	$2.6421e^{-02}$	0.97	0.99
-	-	-	-	-	-	-	$2.9204e^{-02}$	$3.2711e^{-02}$	$3.1094e^{-02}$	0.97	1.02

Table 6.8: Simulation results of the variance (V_{tr}) comparing with empirical variance (V_E) and theoretical variance (V_T) in case of fitted missing covariates model, using different values of parameters by generated variables from bivariate normal distribution with sample size $n = 500$ and σ_1^2=σ_2^2=2

Diagonal component of variance IMT and Ratio												
α_t	β_{t1}	β_{t2}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.2	0.56	0.03	3	3	$4.1405e^{-07}$	$1.1487e^{-07}$	$1.1487e^{-07}$	0.97	0.51
-	-	-	-	-	-	-	$4.0648e^{-04}$	$2.8918e^{-04}$	$4.1885e^{-04}$	0.99	0.83
-	-	-	-	-	-	-	$6.340e^{-04}$	$5.0622e^{-04}$	$6.4030e^{-04}$	0.99	0.90
0.80	0.50	0.4	0.69	0.05	3	3	$1.1003e^{-05}$	$7.8269e^{-06}$	$1.1677e^{-05}$	0.97	0.82
-	-	-	-	-	-	-	$8.9669e^{-04}$	$8.1291e^{-04}$	$9.1548e^{-04}$	0.99	0.94
-	-	-	-	-	-	-	$2.1692e^{-03}$	$2.3403e^{-03}$	$2.2328e^{-03}$	0.99	1.02
1.20	2.20	0.8	0.74	0.16	3	3	$6.2763e^{-04}$	$6.4955e^{-04}$	$6.5224e^{-04}$	0.98	0.99
-	-	-	-	-	-	-	$4.7946e^{-03}$	$4.9422e^{-03}$	$4.9252e^{-03}$	0.99	1.00
-	-	-	-	-	-	-	$1.7828e^{-03}$	$1.8524e^{-03}$	$1.6958e^{-03}$	1.02	1.04
2.30	0.20	1	0.90	0.04	3	3	$3.4057e^{-06}$	$2.4924e^{-07}$	$3.4776e^{-06}$	0.99	0.27
-	-	-	-	-	-	-	$5.0822e^{-05}$	$1.9222e^{-05}$	$5.2125e^{-05}$	0.99	0.61
-	-	-	-	-	-	-	$3.7366e^{-03}$	$4.3500e^{-03}$	$3.9682e^{-03}$	0.99	1.04
0.20	2.30	1.2	0.53	0.23	3	3	$3.0911e^{-04}$	$3.0836e^{-04}$	$3.1398e^{-04}$	0.99	0.99
-	-	-	-	-	-	-	$5.8800e^{-03}$	$5.9649e^{-03}$	$5.9842e^{-03}$	0.99	0.99
-	-	-	-	-	-	-	$1.8528e^{-03}$	$1.8680e^{-03}$	$1.8859e^{-03}$	0.99	0.99

Table 6.9: Simulation results of the variance (V_{tr}) comparing with empirical variance (V_E) and theoretical variance (V_T) in case of fitted missing covariates model, using different values of parameters by generated variables from bivariate normal distribution with sample size $n = 1000$ and σ_1^2=σ_2^2=0.2

Diagonal component of variance IMT and Ratio											
α_t	β_{t1}	β_{t2}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.2	0.57	0.11	3	3	$1.3585e^{-05}$	$1.1718e^{-05}$	$1.3273e^{-05}$	1.01	0.93
-	-	-	-	-	-	-	$2.4996e^{-02}$	$2.5316e^{-02}$	$2.5229e^{-02}$	0.99	1.00
-	-	-	-	-	-	-	$1.4486e^{-01}$	$1.6289e^{-01}$	$1.4526e^{-01}$	0.99	1.05
0.80	0.50	0.4	0.66	0.18	3	3	$2.3253e^{-04}$	$2.6095e^{-04}$	$2.2984e^{-04}$	1.01	1.06
-	-	-	-	-	-	-	$4.5129e^{-02}$	$4.8176e^{-02}$	$4.4637e^{-02}$	1.01	1,03
-	-	-	-	-	-	-	$1.9837e^{-01}$	$2.3374e^{-01}$	$2.0511e^{-01}$	0.98	1.06
1.20	2.20	0.8	0.59	0.36	3	3	$1.3404e^{-03}$	$1.4576e^{-03}$	$1.4116e^{-03}$	0.97	1.01
-	-	-	-	-	-	-	$2.2914e^{-02}$	$2.4611e^{-02}$	$2.3911e^{-02}$	0.98	1.01
-	-	-	-	-	-	-	$3.3481e^{-02}$	$3.7438e^{-02}$	$3.4621e^{-02}$	0.98	1.03
2.30	0.20	1	0.84	0.16	3	3	$3.1898e^{-05}$	$2.0825e^{-05}$	$3.0775e^{-05}$	1.01	0.82
-	-	-	-	-	-	-	$4.7158e^{-03}$	$3.8265e^{-03}$	$4.6557e^{-03}$	1.01	0.91
-	-	-	-	-	-	-	$3.3975e^{-01}$	$3.7696e^{-01}$	$3.4582e^{-01}$	0.99	1.04
0.20	2.30	1.2	0.49	0.34	3	3	$1.1553e^{-03}$	$1.2469e^{-03}$	$1.2100e^{-03}$	0.98	1.01
-	-	-	-	-	-	-	$2.6105e^{-02}$	$2.7428e^{-02}$	$2.7580e^{-02}$	0.97	0.99
-	-	-	-	-	-	-	$3.0897e^{-02}$	$3.4167e^{-02}$	$3.3522e^{-02}$	0.96	1.01

Table 6.10: Simulation results of the variance (V_{tr}) comparing with empirical variance (V_E) and theoretical variance (V_T) in case of fitted missing covariates model, using different values of parameters by generated variables from bivariate normal distribution with sample size $n = 1000$ and σ_1^2=σ_2^2=2

Diagonal component of variance IMT and Ratio											
α_t	β_{t1}	β_{t2}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.2	0.57	0.03	3	3	$1.8198e^{-07}$	$1.3264e^{-07}$	$1.8121e^{-07}$	1.00	10.85
-	-	-	-	-	-	-	$3.4533e^{-04}$	$3.1822e^{-04}$	$3.4883e^{-04}$	0.99	0.96
-	-	-	-	-	-	-	$5.9300e^{-04}$	$5.8219e^{-04}$	$5.8529e^{-04}$	1.01	0.99
0.80	0.50	0.4	0.68	0.06	3	3	$8.4613e^{-06}$	$7.9628e^{-06}$	$8.4471e^{-06}$	1.00	0.97
-	-	-	-	-	-	-	$8.5793e^{-04}$	$8.8372e^{-04}$	$8.5238e^{-04}$	1.00	1.01
-	-	-	-	-	-	-	$2.2046e^{-03}$	$2.2535e^{-03}$	$2.1730e^{-03}$	1.01	1.02
1.20	2.20	0.8	0.72	0.18	3	3	$6.3600e^{-04}$	$6.4124e^{-04}$	$6.2714e^{-04}$	1.01	1.01
-	-	-	-	-	-	-	$4.8989e^{-03}$	$5.0763e^{-03}$	$4.8235e^{-03}$	1.01	1.02
-	-	-	-	-	-	-	$1.8277e^{-03}$	$2.0614e^{-03}$	$2.8117e^{-03}$	1.00	1.06
2.30	0.20	1	0.90	0.04	3	3	$5.0523e^{-07}$	$2.4625e^{-07}$	$5.2674e^{-07}$	0.98	0.68
-	-	-	-	-	-	-	$2.3393e^{-05}$	$1.9714e^{-05}$	$2.4428e^{-05}$	0.98	0.90
-	-	-	-	-	-	-	$4.3268e^{-03}$	$4.5158e^{-03}$	$4.5136e^{-03}$	0.98	1.00
0.20	2.30	1.2	0.54	0.23	3	3	$3.1146e^{-04}$	$3.1615e^{-04}$	$3.1427e^{-04}$	0.99	1.00
-	-	-	-	-	-	-	$5.9995e^{-03}$	$6.0021e^{-03}$	$6.1376e^{-03}$	0.99	0.99
-	-	-	-	-	-	-	$1.9081e^{-03}$	$1.9165e^{-03}$	$1.9210e^{-03}$	0.99	0.99

Table 6.11: Simulation results of the variance (V_{tr}) comparing with empirical variance (V_E) and theoretical variance (V_T) in case of fitted missing covariates model, using different values of parameters by generated variables from bivariate normal distribution with sample size $n = 5000$ and σ_1^2=σ_2^2=0.2

Diagonal component of variance IMT and Ratio											
α_t	β_{t1}	β_{t2}	π_t	S.D (π_t)	Rn_E	Rn_T	V_E	V_T	V_{tr}	R_1	R_2
0.30	0.25	0.2	0.57	0.10	3	3	$1.1739e^{-05}$	$1.1750e^{-05}$	$1.1168e^{-05}$	1.02	1.02
-	-	-	-	-	-	-	$2.5148e^{-02}$	$2.5546e^{-02}$	$2.4530e^{-02}$	1.01	1.02
-	-	-	-	-	-	-	$1.4799e^{-01}$	$1.4767e^{-01}$	$1.4309e^{-01}$	1.00	1.01
0.80	0.50	0.4	0.66	0.18	3	3	$2.3311e^{-04}$	$2.4578e^{-04}$	$2.3088e^{-04}$	1.00	1.03
-	-	-	-	-	-	-	$4.6057e^{-02}$	$4.7721e^{-02}$	$4.5607e^{-02}$	1.00	1.03
-	-	-	-	-	-	-	$2.0328e^{-01}$	$2.2383e^{-01}$	$2.0399e^{-01}$	0.99	1.04
1.20	2.20	0.8	0.62	0.36	3	3	$1.4003e^{-03}$	$1.4503e^{-03}$	$1.4825e^{-03}$	0.97	0.99
-	-	-	-	-	-	-	$2.3669e^{-02}$	$2.4580e^{-02}$	$2.3994e^{-02}$	0.99	1.01
-	-	-	-	-	-	-	$3.5334e^{-02}$	$3.8109e^{-02}$	$3.3710e^{-02}$	0.98	1.01
2.30	0.20	1	0.84	0.10	3	3	$2.5661e^{-05}$	$1.8176e^{-05}$	$2.5049e^{-05}$	1.01	0.85
-	-	-	-	-	-	-	$4.4638e^{-03}$	$4.0200e^{-03}$	$4.3535e^{-03}$	1.01	0.96
-	-	-	-	-	-	-	$3.4891e^{-01}$	$3.8443e^{-01}$	$3.3750e^{-01}$	1.01	1.06
0.20	2.30	1.2	0.51	0.39	3	3	$1.1970e^{-03}$	$1.1607e^{-03}$	$1.2109e^{-03}$	0.99	0.98
-	-	-	-	-	-	-	$2.6962e^{-02}$	$2.5974e^{-02}$	$2.6619e^{-02}$	1.01	0.99
-	-	-	-	-	-	-	$3.2469e^{-02}$	$3.0895e^{-02}$	$3.3152e^{-02}$	0.99	0.97

Table 6.12: Simulation results of the variance (V_{tr}) comparing with empirical variance (V_E) and theoretical variance (V_T) in case of fitted missing covariates model, using different values of parameters by generated variables from bivariate normal distribution with sample size $n = 5000$ and σ_1^2=σ_2^2=2

6.6 The IMT_{DIAG} Under missing covariates

As we considered the behaviour of the IMT under missing covariates logistic model in previous sections, now we will consider the calculation of IMT_{DIAG}. We know that the IM_{DIAG} approach has the same idea as the Information matrix test, but compares just the diagonal elements of the two form of the information matrix. So, z_i is $(p+1)\times 1$ -dimensional vector of the diagonal elements $x_i x_i^T$ matrix. Therefore, the IMT_{DIAG} has the same argument which discussed in previous case IMT statistic, but the vector z_i has different dimension and different elements. As we used in case of IMT, consider we have true model with two covariates X_1 and X_2 and fitting the model with X_1 then,

$$d_g(y_i, \theta) = (Y_i - \pi_i)(1 - 2\pi_i) \begin{bmatrix} 1 \\ X_i^2 \end{bmatrix}$$

6.6.1 An Alternative Formulae of Variance IMT_{DIAG}

As we consider in case of IMT, we wish to compute $\mathsf{var}(T \mid S = 0)$, so

$$\mathsf{var}(T) = E(d_g d_g^T) - E(d_g)E(d_g^T).$$

As before

$$E(d_g d_g^T) = E_x\left([\pi_t(1 - 2\pi) + \pi^2](1 - 2\pi)^2 \begin{bmatrix} 1 & x_i^2 \\ x_i^2 & x_i^4 \end{bmatrix}\right)$$

and,

$$E(d_g) = E_X \left((\pi_t - \pi)(1-2\pi) \begin{bmatrix} 1 \\ x_i^2 \end{bmatrix} \right).$$

Also,

$$\mathsf{cov}(d_g, \nabla\ell) = E(d_g \nabla\ell) - E(d_g)E(\nabla\ell)$$

$$= E_x \left((\pi_t(1-2\pi) + \pi^2)(1-2\pi) \begin{bmatrix} 1 & X \\ X^2 & X^3 \end{bmatrix} \right).$$

This is as before $[E(d_g \nabla\ell)]$ become at the least false values, $E(\nabla\ell) = 0$, so the second term is zero. As the same

$$\mathsf{var}(\nabla\ell) = E(\nabla\ell\nabla\ell^T) = E_X \left((\pi_t(1-2\pi) + \pi^2) \begin{bmatrix} 1 & X \\ X & X^2 \end{bmatrix} \right)$$

As we discussed for IMT, we use the least false values for the parameters in π, so

$$\pi_t = \text{expit}(\alpha_t + \beta_{t1}X_1 + \beta_{t2}X_2),$$

and

$$\pi = \text{expit}(\alpha^* + \beta_1^* X_1).$$

6.6.2 The Variance of IMT_{DIAG} for Logistic Regression Model

We need to use the same assumption which used in case of IMT, but, in this case we have

$$d_g = (y_i - \pi_i)(1-2\pi_i) \begin{bmatrix} 1 \\ x_i^2 \end{bmatrix}$$

So to calculate the variance V, we need to calculate $\mathsf{var}(d_g)$ and $\mathsf{cov}(d_g, \nabla\ell)$, we can see $\mathsf{var}(\nabla\ell)$ has the same expression which used in case of IMT. Firstly, we will work out $\mathsf{var}(d_g)$, we have

$$dd^T = (y-\pi)^2(1-2\pi)2 \begin{bmatrix} 1 & x_i^2 \\ x_i^2 & x_i^4 \end{bmatrix}$$

taking expectation $E_{Y|X}$ we obtain

$$E(dd^T) = E_X \left[(\pi_t(1-2\pi) + \pi^2)(1-2\pi)^2 \begin{bmatrix} 1 & X^2 \\ X^2 & X^4 \end{bmatrix} \right].$$

Secondly,we need to calculate $\mathsf{cov}(d_g, \nabla\ell)$, as we discussed in case of IMT $E(\nabla\ell) = 0$ at the least false value, and we have

$$d_g \nabla\ell^T = (y-\pi)(1-2\pi) \begin{bmatrix} 1 \\ x_i^2 \end{bmatrix} (y-\pi) \begin{bmatrix} 1 & x_i \end{bmatrix}$$

$$= (y-\pi)^2(1-2\pi)\left[\begin{array}{cc} 1 & x_i \\ x_i^2 & x_i^3 \end{array}\right]$$

So,

$$\mathsf{cov}(d_g, \nabla\ell) = E(d\nabla\ell^T) = E_X\left[(\pi_t(1-2\pi)+\pi^2)(1-2\pi)\left[\begin{array}{cc} 1 & X \\ X^2 & X^3 \end{array}\right]\right].$$

Then, we use (6.6) to find V.

6.6.3 Empirical Variance of IMT_{DIAG}

As before one candidate would be to compute

$$d_i = (y_i - \hat{\pi}_i)(1-2\hat{\pi}_i)\left[\begin{array}{c} 1 \\ x_i^2 \end{array}\right] \quad i = 1, \ldots, n$$

and

$$\nabla\ell_i = (y_i - \hat{\pi}_i)\left[\begin{array}{c} 1 \\ x_i \end{array}\right] \quad i = 1, \ldots, n$$

where, $\hat{\pi}_i$ is the fitted value from the model with just x_1. Now compute

$$\hat{W}_n = \frac{1}{n}\sum_{i=1}^{n} d_i d_i^T - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} d_i^T\right)$$

and

$$\hat{B}_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\pi})^2\left[\begin{array}{cc} 1 & x_i \\ x_i & x_i^2 \end{array}\right],$$

$$\hat{C}_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\pi})^2(1-2\hat{\pi}_i)\left[\begin{array}{cc} 1 & x_i \\ x_i^2 & x_i^3 \end{array}\right]$$

Then use (6.7) as an estimate of V.

Chapter 7

Asymptotic Distribution of IMT statistic and Power Calculation

7.1 Introduction

The behaviour of the asymptotic distribution of goodness-of-fit tests is an important statistical problem. In this chapter we are interested to investigate the behaviour of the distribution of IMT statistic. Kuss (2002), discussed methods to examine goodness-of-fit tests, where he shows the IMT_{DIAG} statistic has reasonable power even for a logistic model with very sparse data. We know that the parameters estimators, under the null hypotheses H_0 where there is no mis-specification, will be consistent, asymptotically normal and asymptotically efficient estimators. Under the alternative hypothesis H_1 when the model is mis-specified, however, this estimator will be biased and inconsistent. The constructing of the IMT is based on $\hat{d}$, so, to develop the test the probability limit of d required, and the mean and the variance of the asymptotic distribution of $n\hat{d}^T\hat{V}^{-1}\hat{d}$, should also be examined. For more information about asymptotic distribution of statistics see Hausman (1978).

7.2 Behaviour of the IMT Statistic Distribution

As we discussed in the previous chapters the IMT statistic is distributed asymptotically as central χ^2 distribution under H_0 when the model is correctly specified, and is non-central χ^2 under H_1 when the model mis-specificed. However, the behaviour of the IMT statistic in practice seems affected the near singularity of covariance matrix V related to the first two elements was much close to zero as shown in results in previous chapter. This problem means that in some circumstances properties of the distribution of the IMT (e.g mean and variance) are far away from the properties of χ^2 distribution. Even the use of a generalised inverse $\hat{V}^{-}$ instead $\hat{V}^{-1}$ does not improve matters of V.

7.3 The Behaviour of the Covariance Matrix of the IMT Statistic

As we discussed in the previous chapter the computed formulae of the variance V as $\hat{V}$ works well especially if the $S.D$ of π_t not small and when the sample size is large. However, because, the statistic IMT is $nd^T V^{-1} d$, which depends upon V^{-1} not V,the near singularity of V^{-1} affects the behaviour of the properties of the statistic. Indeed, the issue which may be causes the singularity problem is the zero elements of V. As

$$\sum_{i=1}^{n} \nabla \ell_i = 0 \tag{7.1}$$

and some elements of $\sum_{i=1}^{n} d_i$ are close to elements of (7.1), then the dispersion matrix V is close to singular. In fact, the dispersion matrix depends upon $E(d)$, and in this case

$$E(d) = E_X(\pi_t - \pi)(1 - 2\pi) \begin{bmatrix} 1 \\ X \\ X^2 \end{bmatrix}.$$

We can see the first two elements will be zero if the factor $(1 - 2\pi)$ is constant, corresponding to the log-likelihood functions

$$E(Y - \pi) = 0 \; , E((Y - \pi)X) = 0,$$

and close to 0 if $(1 - 2\pi)$ varies little between cases. To illustrate this problem we should focus on the elements of the eigenvalues and the eigenvectors of the covariance matrix V by simulation example.

7.3.1 Simulation Example for Eigenvector of the Covariance Matrix V

We consider this simulation example to illustrate the behaviour of the covariance matrix of d by investigating it is eigenvectors and eigenvalues. If we have a true logistic regression model has been fitted

$$\pi = \text{expit}(\alpha + \beta_1 X_1)$$

where X is drawn from a normal distribution with zero mean and $\sigma^2 = 2$ and using different cases of parameters α and β_1, with sample size $n = 500$ and $N = 5000$ simulations. Our example designed for computing the eigenvalues and eigenvectors of the empirical covariance matrix $V = \hat{\mathsf{var}}(\hat{d})$. Three cases used different parameters under H_0, (α, β_1) are (0,1), (0.3,0.25), (0.8,0.5) respectively. The result of simulation shows the average of the empirical covariance matrix and the eigenvalues and eigenvectors for three cases of parameters are, respectively;

In case $\alpha = 0$, $\beta_1 = 1$ the empirical covariance is

$$V_{emp} = \begin{bmatrix} 6.535 & 0.044 & -87.78 \\ 0.044 & 522.5 & -3.418 \\ -87.78 & -3.418 & 1247 \end{bmatrix} \times 10^{-4},$$

and the eigenvalues of V_{emp} are

$$Evalues = \begin{bmatrix} 1254.01 & 522.519 & 0.35722 \end{bmatrix} \times 10^{-4}$$

and the eigenvalues are the columns of

$$Evectors = \begin{bmatrix} -0.070198948 & -0.0007014321 & 0.9975327642 \\ -0.004665729 & 0.9999890452 & 0.0003748199 \\ 0.997522099 & 0.0046279054 & 0.0702014515 \end{bmatrix}.$$

In case $\alpha = 0.3, \beta_1 = 0.25$ the empirical covariance is

$$V_{emp} = \begin{bmatrix} 0.1345 & -3.964 & -10.52 \\ -3.964 & 230.2 & 167.8 \\ -10.52 & 167.8 & 1369 \end{bmatrix} \times 10^{-4},$$

and the eigenvalues,eigenvectors are

$$Evalues = \begin{bmatrix} 1393.85 & 206.087 & 0.01958 \end{bmatrix} \times 10^{-4}$$

$$Evactors = \begin{bmatrix} 0.007878297 & 0.01175401 & 0.999899883 \\ -0.142772221 & -0.98967333 & 0.012758707 \\ -0.989724217 & 0.14285844 & 0.006118794 \end{bmatrix}.$$

In case $\alpha = 0.8, \beta_1 = 0.5$ the empirical covariance is

$$V_{emp} = \begin{bmatrix} 2.368 & -26.55 & -39.02 \\ -26.55 & 423.7 & 92.72 \\ -39.02 & 92.72 & 1908 \end{bmatrix} \times 10^{-4},$$

and the eigenvalues,eigenvectors are

$$Evalues = \begin{bmatrix} 1914.75 & 419.326 & 0.11908 \end{bmatrix} \times 10^{-4}$$

$$
Evectors = \begin{bmatrix} 0.02123125 & 0.05749870 & 0.99811980 \\ -0.06243347 & -0.99632006 & 0.05872306 \\ -0.99782328 & 0.06356285 & 0.01756328 \end{bmatrix}.
$$

Covariance matrix appeared small variance more close to zero in all cases of different parameters. The first element of eigenvalues concerns to the first element of rows in eigenvector and the second elements concerning to the second element of rows in eigenvector and the same of the third element. Also, we can see clearly, that the third elements of the eigenvalues appeared very close to zero comparing with the two first elements in all cases of simulation. So, we can see that the smallest eigenvalue is very small, demonstrating the near singularity of V. To more illustrate this point Figure 7.1, shows the first element of the third columns of the eigenvector of the empirical covariance matrix. We can see that nearly all the elements which are the third element in the first row related to the third element of the eigenvalues in all cases of different parameters appeared close to one. This suggest that there is little variation in d_1. Similarly there is much less variation in d_2 compared with d_3.

Generally, as we discussed the behaviour of IMT distribution, we focus on the results which are shown by $Kuss$ (2002). These results show the comparison between various goodness-of-fit statistics with sparse data, which appeared IMT and IMT_{DIAG} have reasonable power. Kuss's does not mention the behaviour of the asymptotic distribution of IMT. Firstly, we are interested to re-simulate of Kuss's example to show how the behaviour of the distribution of IMT is affected, before going to present our idea to solve this problem. So, we will consider the example given by $Kuss$ (2002) in the last section.

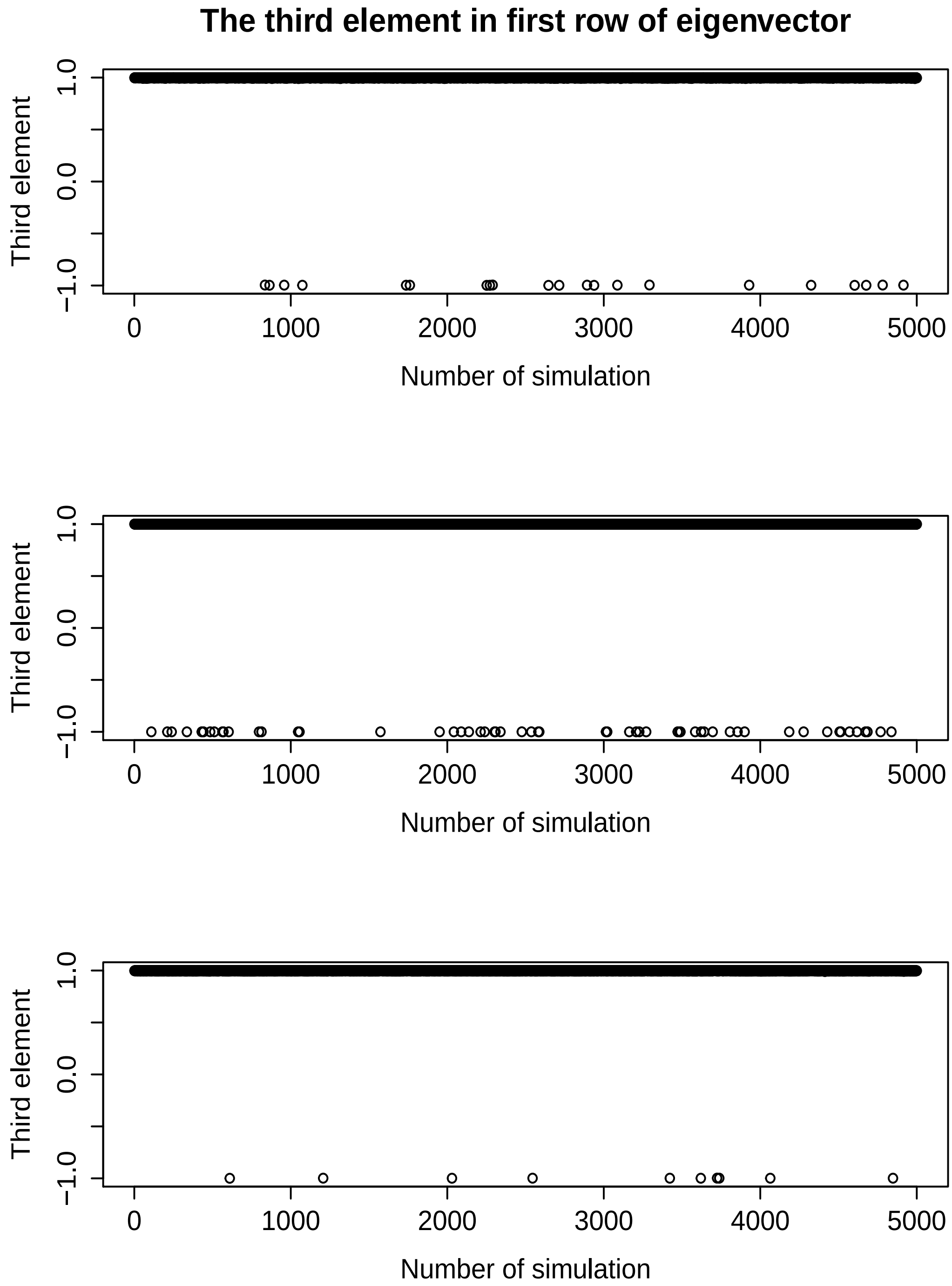

Figure 7.1: Plot of the third element in the first row of the eigenvector of empirical covariance matrix V, under H_0 for three different cases of parameters (α, β_1), $(0, 1)$, $(0.30, 0.25)$ and $(0.80, 0.50)$ respectively, with $\sigma_1^2 = \sigma_2^2 = 2$ and $\rho = 0.1$, sample size $n = 500$ and $N = 5000$ number of simulation.

7.4 Information Matrix Test Reduced ($IMTR$)

In this section, our purpose is to develop the form of IMT statistic which is asymptotically distributed χ^2_R distribution under H_0, when the model is correctly specified, and non-central $\chi^2_R(\lambda)$ distribution under H_1, when the model is mis-specified, in this case $d \sim N(\mu, V)$, then $d^T V^{-1} d \sim \chi^2$ and

$$E(d^T V^{-1} d) = E(V^{-1} d d^T)$$
$$= E(V^{-1}(d-\mu)(d-\mu)^T) + \mu^T V^{-1} \mu$$
$$= rank(V) + \mu^T V^{-1} \mu$$

Note that in this case χ^2 has mean $R + \lambda$ and variance $2(R + 2\lambda)$, where R is the rank of V and $\lambda = \mu^T V^{-1} \mu$. So, the main point is avoid the singularity problem that discussed in previous sections, which is related with the log likelihood function. The basic idea is to consider a version of the IMT based on a reduced set of the elements of d. Therefore, we removed the elements which are related to the log likelihood function. To illustrate our idea let us consider an example as we discussed in previous if we have fitted the model with one covariate then, we have

$$E(d) = E_X \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = E_X \begin{bmatrix} (\pi_t - \pi)(1-2\pi) \\ (\pi_t - \pi)(1-2\pi)x_i \\ (\pi_t - \pi)(1-2\pi)x_i^2 \end{bmatrix}$$

So, as we discussed we need to remove the elements d_1 and d_2 from d, and then we will use only just d_3 to compute the statistic. in this case $d = d_3$ and the statistic is $n d_3^2 V^{-1}$. This approach we calle the $IMTR$, and we will evaluate the $IMTR$ statistic by simulation to examine the behaviour of its asymptotic distribution.

7.5 Simulation Study

In this part of simulation, we are interested to examine the asymptotic distribution of IMT statistic in case when all the elements of (d) are used, and also we need to investigate the properties of the $IMTR$ and how the reduced elements improve and the asymptotic distribution of the $IMTR$ as chi-square distribution with mean [rank (V)] and variance [2 rank (V)], if the fitted model is correct. Also, we investigate the asymptotic distribution of IMT under mis-specified model to focus on the behaviour of the asymptotic distribution of IMT, which is in this case is distributed non central chi-square distribution with mean is [rank $(V) + \lambda$] and variance [2 rank (V)+4 λ] where $\lambda = E(d)^T V^{-1} E(d)$. Moreover, examine the effect of elements of variance matrix by likelihood function.

7.5.1 Design of Simulation

This simulation designed to examine the asymptotic distribution of IMT and $IMTR$, we will consider two cases of simulation under true model and under mis-specified model. If we have true logistic regression model with two covariates

$$\pi_i = \text{expit}(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2}).$$

Firstly, we will focus on asymptotic distribution of IMT when the true model is fitted. Secondly, investigate the asymptotic distribution of IMT when the missing covariate logistic model has been fitted:

$$\pi_i = \text{expit}(\alpha + \beta_1 x_{i1}).$$

- we consider x_{i1} and x_{i2} as a draw from bivariate normal distribution $X \sim N_2(0, \Omega)$.
- We consider the 2×2 covariance matrix is.

$$\Omega = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

- Use two cases of variance $\sigma^2 = 0.2, 2$ and $\rho = 0.1$.
- We choose different component of parameters under fitted true model as (α_t, β_t): (0,1), (0.3,0.25), (0.8, 0.5), (1.2,2.2), (3.5,2.3).
- Under fitted missing covariates model,we choose different component of parameters α_t=(0,0.8,0.9,1.2,0.7), β_{t1}=(1,0.7,1.3,2.2,1.5) and $\beta t2$=(0.6,0.4,1.2,1.8,2).
- Three cases of sample size uses $n = 500, 1000, 5000$ and $N = 50000$ number of simulation.

7.5.2 Results and Discussion in Case of Correctly Specified Model

In this simulation we consider to compute the IMT with two cases of dispersion matrix V and V_E as we discussed in the previous chapter. To investigate the behavior of IMT and $IMTR$ under effects of theoretical variance which computed by alternative formulae and empirical variance, and comparing the results. The results of simulation reported in several tables. These tables show the mean and the variance of IMT and $IMTR$ by each found the theoretical and empirical variance. That is The $IMTE$

denote to the statistic computed by empirical variance and $IMTV$ denote to use theoretical variance,

$$IMTE = \bar{d}^T \hat{\mathsf{var}}(\bar{d})^{-1}\bar{d}$$

and

$$IMTV = \bar{d}^T \mathsf{var}(\bar{d})^{-1}\bar{d}$$

where, $\hat{d} = (\hat{d}_1, \hat{d}_2, \hat{d}_3)^T$ and $d = (d_1, d_2, d_3)^T$ i.e. full matrix. Also, $IMTE1$ and $IMTV1$ denote to the statistic when, $\hat{d} = (\hat{d}_2, \hat{d}_3)^T$ and $d = (d_2, d_3)^T$, i.e reduced the first element. Finally, $IMTE2$ and $IMTV2$ denoted to the statistic when,$\hat{d} = (\hat{d}_3)^T$, and $d = (d_3)^T$, i.e. reduced the two first elements. Also, α_t and β_{t1} denote to the true parameters of π_t and $S.D(\pi_t)$ is the standard deviation of π_t over the distribution of the covariates. Table 7.1 and Table 7.2 shows the results in case of sample size $n = 500$ and $\sigma_1^2 = \sigma_2^2 = 0.2, 2$ respectively. Table 7.3 and Table 7.4 shows the results in case of sample size $n = 1000$, Table 7.5 and Table 7.6 shows the results in case of sample size $n = 5000$.

If we maintain the IMT is asymptotically distributed as χ^2_R distribution, with $df = R$ where, R is the rank of V, so, the statistics $IMTE$ or $IMTV$ should have mean $R = 3$ and variance $2R = 6$, the statistics $IMTE1$ or $IMTV1$ has mean $R = 2$ and variance $2R = 4$ and the last statistics, $IMTE2$ or $IMTV2$ have mean $R = 1$ and variance $2R = 2$. Generally, we can see clearly, that the properties of χ^2 distribution do not apply for both $IMTE$ and $IMTV$ for most sets of parameters, different σ^2 and different sample sizes. The variance shows by far the more erratic behaviour. If we look at the second proposed statistic $IMTE1$ or $IMEV1$, the properties of χ^2 still do not apply, but, the departures are less than problem for $IMTE$, $IMTV$ and it is looks better. The final proposed statistic, which is our proposed $IMTR$, the new form of the IMT denoted in this simulation by $IMTE2$ and $IMTV2$, shows reasonable properties, the mean and the variance appeared very close to the properties of χ^2 distribution across all cases.

If we consider the results by the sample size, we can see that, when the sample size is larger, the results appear much better. In case of sample size $n = 500$, $IMTR$ in some cases appeared slightly affected, especially when using the empirical variance, and for small values of the $S.D$ of π_t. If we make a comparison between the IMT, computed by empirical variance and theoretical variance, the results reported that, in large sample size $n = 5000$ have the same behaviour. Finally, we can say although

				Mean and variance of the IMT						
α_t	β_{t1}	π_t	$S.D\pi_t$	-	$IMTE$	$IMTE1$	$IMTE2$	$IMTV$	$IMTV1$	$IMTV2$
0	1	0.48	0.07	Mean	3.575	2.292	1.150	1874.6	1.991	1.008
				var	8.966	5.913	2.965	39629687	4.813	2.422
0.3	0.25	0.57	0.01	Mean	2.191	2.080	1.055	4.887	4.887	1.180
				var	4.472	4.145	2.186	98.414	98.381	3.935
0.8	0.5	0.48	0.07	Mean	2.955	2.188	1.122	18028.2	3.228	0.992
				var	7.693	5.069	2.856	6196039094	32.236	2.115
1.2	2.2	0.72	0.21	Mean	5.448	2.614	1.269	13.705	1.974	1.002
				var	47.064	10.114	4.313	917.927	4.496	2.307
3.5	2.3	0.95	0.02	Mean	12.056	4.044	2.053	20.455	2.206	1.168
				var	613.99	50.889	23.290	2581.01	10.060	4.145

Table 7.1: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is correctly specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 500$, σ_1^2=σ_2^2=0.2, $\rho = 0.1$ and $\mu_1 = \mu_2 = 0$

				Mean and variance of the IMT						
α_t	β_{t1}	π_t	$S.D\pi_t$	-	$IMTE$	$IMTE1$	$IMTE2$	$IMTV$	$IMTV1$	$IMTV2$
0	1	0.49	0.25	Mean	7.151	2.823	1.386	11.951	1.960	0.98
				var	82.956	13.393	6.156	539.85	4.365	2.213
0.3	0.25	0.48	0.07	Mean	2.483	2.204	1.122	2.030	2.030	0.999
				var	6.343	5.178	2.782	5.829	5.829	2.468
0.8	0.5	0.66	0.14	Mean	4.560	2.458	1.249	47.621	1.975	0.998
				var	22.865	8.007	4.115	15031.2	4.592	2.318
1.2	2.2	0.63	0.35	Mean	15.856	3.666	1.618	4.287	1.973	0.993
				var	854.47	28.515	9.430	51.095	5.128	2.594
3.5	2.3	0.83	0.24	Mean	22.847	4.240	1.388	3.885	1.957	0.976
				var	2217.8	39.473	5.355	48.626	5.127	2.087

Table 7.2: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is correctly specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 500$, σ_1^2=σ_2^2=2, $\rho = 0.1$ and $\mu_1 = \mu_2 = 0$

there are slight effects in some cases related to the small value of $S.D(\pi_t)$, the new form of statistic $IMTR$ works well and has reasonable behaviour in most of the cases investigated. Moreover, we can say that the $IMTR$ statistic appeared to have an asymptotic χ^2 distribution without strange behaviour, at least with request to the mean and variance.

Mean and variance of the IMT										
α_t	β_{t1}	π_t	$S.D\pi_t$	-	$IMTE$	$IMTE1$	$IMTE2$	$IMTV$	$IMTV1$	$IMTV2$
0	1	0.50	0.12	Mean	3.524	2.168	1.087	489.7	1.905	0.922
				var	8.859	5.194	2.636	1930003	4.017	1.859
0.30	0.25	0.56	0.03	Mean	2.087	2.039	1.030	3.448	3.447	1.044
				var	4.155	4.013	2.099	36.880	36.866	2.564
0.8	0.5	0.67	0.03	Mean	2.879	2.105	1.058	8890.9	2.564	1.019
				var	7.124	4.596	2.437	1126068493	12.820	2.153
1.2	2.2	0.72	0.18	Mean	4.570	2.353	1.168	8.285	2.015	1.028
				var	31.828	7.353	3.400	235.13	4.324	2.238
3.5	2.3	0.94	0.06	Mean	6.738	2.875	1.507	11.146	2.102	1.071
				var	145.32	14.751	7.699	612.50	7.120	2.971

Table 7.3: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is correctly specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 1000$, σ_1^2=σ_2^2=0.2, $\rho = 0.1$ and $\mu_1 = \mu_2 = 0$

Mean and variance of the IMT										
α_t	β_{t1}	π_t	$S.D\pi_t$	-	$IMTE$	$IMTE1$	$IMTE2$	$IMTV$	$IMTV1$	$IMTV2$
0	1	0.49	0.26	Mean	5.728	2.451	1.217	7.099	1.990	0.995
				var	59.68	8.599	4.041	137.50	4.225	2.122
0.3	0.25	0.56	0.09	Mean	2.327	2.132	1.074	2.008	2.008	0.967
				var	5.734	4.867	2.541	4.929	4.929	2.125
0.8	0.5	0.65	0.13	Mean	4.092	2.272	1.152	26.793	1.979	0.984
				var	18.246	6.292	3.254	4349.58	4.270	2.079
1.2	2.2	0.63	0.35	Mean	9.610	2.831	1.324	3.637	1.981	0.997
				var	359.52	14.428	5.681	26.351	4.490	2.182
3.5	2.3	0.83	0.26	Mean	12.485	3.107	1.203	3.471	1.983	0.995
				var	804.510	18.161	3.771	28.060	4.883	2.093

Table 7.4: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is correctly specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 1000$, σ_1^2=σ_2^2=2, $\rho = 0.1$ and $\mu_1 = \mu_2 = 0$

Mean and variance of the IMT										
α_t	β_{t1}	π_t	$S.D\pi_t$	-	$IMTE$	$IMTE1$	$IMTE2$	$IMTV$	$IMTV1$	$IMTV2$
0	1	0.50	0.10	Mean	3.324	2.055	1.028	127.34	1.989	0.992
				var	8.555	4.482	2.268	104530	4.079	2.051
0.3	0.25	0.57	0.02	Mean	2.018	2.018	1.010	2.247	2.247	0.972
				var	4.080	4.079	2.094	8.201	8.199	2.027
0.8	0.5	0.70	0.04	Mean	2.906	2.020	1.016	1626.3	2.165	1.037
				var	6.462	4.121	2.124	25053397	5.819	2.177
1.2	2.2	0.73	0.17	Mean	3.458	2.077	1.035	4.117	1.995	0.994
				var	11.521	4.788	2.371	24.312	4.112	2.050
3.5	2.3	0.95	0.05	Mean	3.577	2.164	1.109	5.073	2.106	1.041
				var	13.550	5.298	2.909	56.482	5.102	2.277

Table 7.5: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is correctly specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 5000$, σ_1^2=σ_2^2=0.2, $\rho = 0.1$ and $\mu_1 = \mu_2 = 0$

Mean and variance of the IMT										
α_t	β_{t1}	π_t	$S.D\pi_t$	-	$IMTE$	$IMTE1$	$IMTE2$	$IMTV$	$IMTV1$	$IMTV2$
0	1	0.49	0.26	Mean	3.802	2.106	1.053	3.872	1.995	0.994
				var	17.762	4.961	2.493	18.494	4.025	1.989
0.3	0.25	0.58	0.07	Mean	2.075	2.026	1.010	1.946	1.946	0.955
				var	4.353	4.227	2.132	3.940	3.940	1.879
0.8	0.5	0.68	0.14	Mean	3.368	2.061	1.033	7.923	2.033	1.026
				var	9.912	4.566	2.326	228.722	4.214	2.146
1.2	2.2	0.63	0.35	Mean	4.351	2.178	1.067	3.127	2.001	0.990
				var	33.614	5.723	2.661	9.642	4.135	2.053
3.5	2.3	0.48	0.07	Mean	4.630	2.237	1.042	3.098	1.998	0.997
				var	43.533	6.198	2.372	9.118	4.117	2.019

Table 7.6: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 5000$, σ_1^2=σ_2^2=2, $\rho = 0.1$ and $\mu_1 = \mu_2 = 0$

Mean and variance of the IMT by Theoretical and Empirical Variance											
α_t	β_{t1}	β_{t2}	π_t	$S.D\pi_t$	-	$IMTV$	$IMTV1$	$IMTV2$	$IMTE$	$IMTE1$	$IMTE2$
0	1	0.6	0.80	0.07	Mean	84.71	2.010	0.973	4.721	2.492	1.172
					var	52231	5.634	1.960	29.73	8.554	3.373
0.8	0.7	0.4	0.86	0.06	Mean	752.8	2.510	0.973	4.806	2.593	1.237
					var	68376	14.04	1.988	30.94	9.676	4.078
0.9	1.3	1.2	0.94	0.05	Mean	35.70	2.168	0.958	10.94	3.820	1.791
					var	11022	11.31	2.001	424.1	39.46	12.66
1.2	2.2	1.8	0.98	0.02	Mean	14.09	1.802	0.897	76.01	17.21	6.215
					var	1104	9.907	2.505	3664	4742	61.25
0.7	1.5	2	0.97	0.05	Mean	23.96	2.557	0.983	18.85	5.178	2.362
					var	3322	18.85	2.297	1906	115.0	26.32

Table 7.7: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is mis-specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 500$, σ_1^2=σ_2^2=0.2, $\rho = 0.6$ and $\mu_1 = \mu_2 = 0$

7.5.3 Results and Discussion in Case of Mis-specified Model

In this part we will discuss the results under H_1, when the model is mis-specified. We used the same assumptions which we discussed in previous section, but in this case $\beta_{t2} \neq 0$, and we choose different cases of parameters ($\beta_{t2} = 0.4, 0.6, 1.2, 1.8, 2$). Table 7.7 and Table 7.8 shows the results in two case of $\sigma_1^2 = \sigma_2^2 = 0.2, 2$ respectively and sample size $n = 500$, Table 7.9 and Table 7.10, shows the results in case of sample size $n = 1000$ and Table 7.11 and Table 7.12, shows the results in case of sample size $n = 5000$.

We see from the tables that $IMTV$ and $IMTE$ generally do not have means and variance that are close to those expected from a χ^2 distribution. This is due to the instability resulting from the close relation between the expressions for $IMTV$ and $IMTE$ and the corresponding log-likelihood. As was the case under H_0, our alternative $IMTR$ gave more stable results.

However, the assumption that its distribution closely follows a non-central χ^2 is not well supported. If X is $\chi^2_\nu(\lambda)$, for non-central parameter λ, then $\Re = \frac{\mathsf{var}(X)-2\nu}{E(X)-\nu} = 4$. The sample version of this quantity, for $n = 5000$, are shown in Table 7.13. Although not highly discrepant, the agreement is disappointing.

Mean and variance of the IMT by Theoretical and Empirical Variance											
α_t	β_{t1}	β_{t2}	π_t	$S.D\pi_t$	-	$IMTV$	$IMTV1$	$IMTV2$	$IMTE$	$IMTE1$	$IMTE2$
0	1	0.6	0.53	0.31	Mean	7.525	1.987	1.006	9.327	3.177	1.600
					var	169.5	4.286	2.226	173.8	18.40	9.035
0.8	0.7	0.4	0.64	0.25	Mean	9.930	1.948	0.969	7.086	2.850	1.372
					var	367.1	4.140	2.085	87.51	13.28	5.704
0.9	1.3	1.2	0.59	0.32	Mean	6.616	2.203	1.111	14.89	4.126	1.977
					var	108.2	4.645	2.352	507.3	33.14	13.63
1.2	2.2	1.8	0.56	0.36	Mean	7.411	2.623	1.358	24.89	5.447	2.561
					var	105.8	6.117	3.262	1402	55.45	22.45
0.7	1.5	2	0.55	0.34	Mean	7.374	2.368	1.311	17.31	4.666	2.498
					var	113.0	4.865	2.852	599.4	38.57	19.44

Table 7.8: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is mis-specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 500$, σ_1^2=σ_2^2=2, $\rho = 0.6$ and $\mu_1 = \mu_2 = 0$

Mean and variance of the IMT by Theoretical and Empirical Variance											
α_t	β_{t1}	β_{t2}	π_t	$S.D\pi_t$	-	$IMTV$	$IMTV1$	$IMTV2$	$IMTE$	$IMTE1$	$IMTE2$
0	1	0.6	0.79	0.08	Mean	46.89	2.114	1.003	4.040	2.271	1.088
					var	15058	5.413	2.045	19.88	6.349	2.659
0.8	0.7	0.4	0.86	0.06	Mean	325.2	2.224	0.979	4.247	2.339	1.121
					var	11002	8.191	1.976	22.45	6.972	3.063
0.9	1.3	1.2	0.96	0.04	Mean	21.55	2.216	0.972	6.619	2.858	1.372
					var	3211	8.860	1.974	120.4	14.34	5.935
1.2	2.2	1.8	0.97	0.06	Mean	11.53	2.229	0.960	18.85	4.997	2.478
					var	665.2	12.41	2.314	3090	116.3	32.06
0.7	1.5	2	0.96	0.04	Mean	12.32	2.045	0.983	8.718	3.294	1.645
					var	789.5	6.986	2.001	324.6	23.95	9.512

Table 7.9: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is mis-specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 1000$, σ_1^2=σ_2^2=0.2, $\rho = 0.6$ and $\mu_1 = \mu_2 = 0$

	Mean and variance of the IMT by Theoretical and Empirical Variance										
α_t	β_{t1}	β_{t2}	π_t	$S.D\pi_t$	-	$IMTV$	$IMTV1$	$IMTV2$	$IMTE$	$IMTE1$	$IMTE2$
0	1	0.6	0.52	0.29	Mean	5.339	2.031	1.039	7.250	2.742	1.407
					var	56.47	4.064	2.107	115.7	11.46	5.809
0.8	0.7	0.4	0.66	0.21	Mean	6.745	2.018	1.011	5.771	2.511	1.232
					var	124.5	4.249	2.147	62.82	8.980	4.099
0.9	1.3	1.2	0.62	0.33	Mean	5.228	2.391	1.225	11.40	3.645	1.808
					var	42.32	5.033	2.614	339.7	21.56	9.677
1.2	2.2	1.8	0.62	0.34	Mean	6.431	3.076	1.642	19.72	5.109	2.502
					var	51.73	7.455	4.154	1020	39.09	16.72
0.7	1.5	2	0.56	0.33	Mean	6.570	2.809	1.664	14.47	4.458	2.554
					var	64.10	6.395	4.069	472.7	29.57	16.06

Table 7.10: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is mis-specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 1000$, σ_1^2=σ_2^2=2, $\rho = 0.6$ and $\mu_1 = \mu_2 = 0$

	Mean and variance of the IMT by Theoretical and Empirical Variance										
α_t	β_{t1}	β_{t2}	π_t	$S.D\pi_t$	-	$IMTV$	$IMTV1$	$IMTV2$	$IMTE$	$IMTE1$	$IMTE2$
0	1	0.6	0.81	0.08	Mean	11.57	2.052	1.013	3.299	2.068	1.020
					var	618.9	4.353	2.024	8.901	4.452	2.124
0.8	0.7	0.4	0.86	0.05	Mean	73.43	2.057	0.993	3.361	2.071	1.019
					var	44634	4.802	1.997	9.779	4.609	2.210
0.9	1.3	1.2	0.95	0.04	Mean	6.138	1.828	0.988	3.673	2.208	1.116
					var	131.4	3.744	1.926	14.37	5.515	2.879
1.2	2.2	1.8	0.98	0.03	Mean	5.646	2.250	1.048	4.551	2.528	1.332
					var	72.80	6.869	2.217	36.79	8.885	4.982
0.7	1.5	2	0.96	0.05	Mean	6.091	2.194	1.067	3.981	2.338	1.223
					var	98.12	5.684	2.179	20.08	6.687	3.646

Table 7.11: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is mis-specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 5000$, σ_1^2=σ_2^2=0.2, $\rho = 0.6$ and $\mu_1 = \mu_2 = 0$

Mean and variance of the IMT by Theoretical Variance											
α_t	β_{t1}	β_{t2}	π_t	$S.D\pi_t$	-	$IMTV$	$IMTV1$	$IMTV2$	$IMTE$	$IMTE1$	$IMTE2$
0	1	0.6	0.49	0.29	Mean	3.742	2.182	1.191	4.641	2.470	1.361
					var	13.25	4.457	2.577	31.30	7.381	4.365
0.8	0.7	0.4	0.63	0.24	Mean	3.796	2.037	1.011	3.847	2.180	1.085
					var	16.21	4.099	2.001	17.33	5.459	2.680
0.9	1.3	1.2	0.60	0.32	Mean	5.332	3.718	1.992	5.573	4.537	2.375
					var	18.56	10.05	5.573	108.3	20.68	10.44
1.2	2.2	1.8	0.60	0.36	Mean	9.211	6.771	3.693	17.05	8.265	4.322
					var	39.29	22.35	12.40	398.1	48.54	23.00
0.7	1.5	2	0.56	0.33	Mean	8.539	5.970	4.231	13.45	7.029	4.845
					var	38.76	19.31	14.72	226.4	37.38	25.91

Table 7.12: Simulation results of mean and variance of IMT by theoretical and empirical variance, when the model is mis-specified and $df = 3, 2, 1$ related to the three cases of IMT respectively, variables generated from bivariate Normal distribution with sample size $n = 5000$, σ_1^2=σ_2^2=2, $\rho = 0.6$ and $\mu_1 = \mu_2 = 0$

		$\sigma_1^2 = \sigma_2^2 = 0.2$			Values of $\Re$	
α_t	β_{t1}	β_{t2}	π_t	$S.D\pi_t$	$\Re$-Theoretical	$\Re$-Empirical
0	1	0.6	0.81	0.08	1.85	6.20
0.8	0.7	0.4	0.86	0.05	0.43	11.05
0.9	1.3	1.2	0.95	0.04	6.17	7.58
1.2	2.2	1.8	0.98	0.03	4.52	8.98
0.7	1.5	2	0.96	0.05	1.76	7.38
		$\sigma_1^2 = \sigma_2^2 = 2$				
0	1	0.6	0.49	0.29	3.02	6.55
0.8	0.7	0.4	0.63	0.24	0.09	8.00
0.9	1.3	1.2	0.60	0.32	3.60	6.14
1.2	2.2	1.8	0.60	0.36	3.86	6.32
0.7	1.5	2	0.56	0.33	3.94	6.22

Table 7.13: Compute the value $\Re = \mathsf{var}(X) - 2\nu/E(X) - \nu$ of $IMTR$ by theoretical and empirical variance, with sample size $n = 5000$, σ_1^2=σ_2^2=0.2 and 2, $\rho = 0.6$ and $\mu_1 = \mu_2 = 0$

7.6 Conclusion

We have investigated the new form of the information matrix test $IMTR$ by simulation which reduced the elements of d to remove overlap with elements of the log-likelihood function. In fact, although there is slightly different results when using the empirical covariance matrix with sample size $n = 500$, the $IMTR$ appeared reasonable asymptotic distribution behaviour and the properties very close to the χ^2-distribution under H_0. However, the form of the distribution under H_1 is less clear. According to these results, it would be helpful to try an alternative approach. In the next chapter we will investigate the application of the bootstrap to this problem.

7.7 Simulation Study of Kuss (2002)

Kuss (2002), discussed and compared various goodness-of-fit tests in logistic regression with sparse data. The idea of the comparison is to evaluate goodness-of-fit tests and also examine the behaviour of the tests. We will focus on four goodness-of-fit tests $(\hat{C}_g, RSS, IM, IM_{DIAG})$. The simulation has been designed to examine the behaviour of various goodness-of-fit tests under the alternative hypotheses of a missing covariate, or wrong function a form of the covariate. In our work, we focus on behaviour of goodness-of-fit tests under alternative hypotheses in case of missing covariate model and the behaviour of the asymptotic distribution of goodness-of-fit statistics, because in these cases we could not reproduce Kuss's results. Therefore, we will examine in more depth the behaviour of the tests and determine more information about asymptotic MLE distribution in case of the missing covariate model

$$\pi_i = \text{expit}(0.405x_i + 0.223u_i),$$

where $X, U \sim U(-6, 6)$, X and U independent.

7.7.1 Design of studies

We designed the simulation study as Kuss's example follows:

The sample sizes are n=100 and n=500;

the number of simulations is 1000;

distribution of the predictor variables X, U is $U(-6, 6)$, X and U independent, chosen to conform with Kuss's work.

Use four of goodness-of-fit tests from the simulation study under two different alternative hypotheses:

- True covariates fitted.
- Missing covariate.

- Fitted model in all cases is a standard logistic model with an intercept and one covariate X.

7.7.2 Results and discussion of Tests Under Correct Model

In Table 7.15, we report some results, the mean ,variance and the empirical power of four goodness-of-fit tests from simulation study under correct model, namely

$$\pi_i = \text{expit}(0.693x_i).$$

The distribution of value of π_i given $X \sim U(-6, 6)$ are shown in the histogram in Figure 7.2.

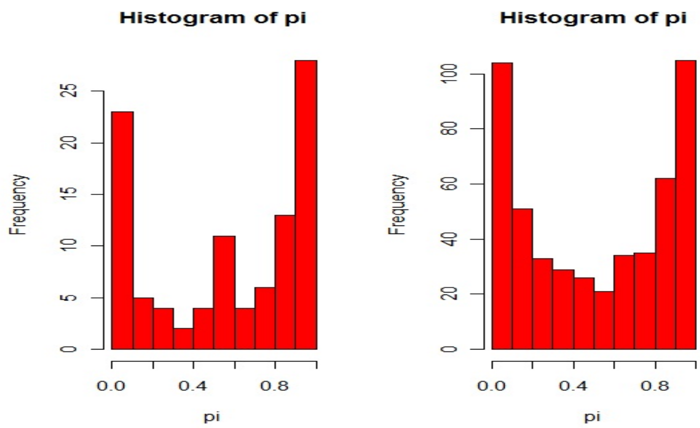

Figure 7.2: Histogram plots of the value of π_i given $X \sim U(-6, 6)$ with two samples size n=100 and n=500 respectively.

Statistics used in the simulation as goodness-of fit tests are: Hosmer - Lemeshow ($\hat{C}_g$), Information matrix (IM), Information matrix Diagonal (IM_{DIAG}) and residual sum of squares (RSS). The asymptotic distribution of statistics is χ^2_{df} distribution, where the mean and variance equal df and $2df$ respectively. In case of ($\hat{C}_g$) statistic we have chosen the number of group is $g = 10$ so, degree of freedom is $df = g - 2$. we can see the results shown in Table 7.14, the mean and variance of all statistics

$n = 100$					$n = 500$		
$-$	df	Mean	Var	%Rej	Mean	Var	%Rej
$\hat{C}_g$	8	8.06	20.47	4.6	7.96	17.12	5.70
IM	3	3.06	7.23	5.10	3.00	6.33	4.70
IM_{DIAG}	2	2.04	3.97	5.50	1.94	3.63	4.20
RSS	1	0.98	1.81	4.60	0.99	1.83	4.10

Table 7.14: Results of N=1000 simulation with sample size n=100 and n=500 under correct model

appeared close to *df* and 2*df*. Moreover, we found the results are better when fit the model with sample size $n = 500$. However, there is slightly large variance of ($\hat{C}_g$) in case of sample size $n = 100$. Overall, the empirical power and type I error looks good .

7.7.3 Results and Discussion of Tests Under Missing Covariate Model

In this part we will report the results of power to detect a misspecified model for same goodness-of-fit tests under missing covariate model, when the model is:

$$\text{logit}(\pi_i) = \text{expit}(0.405x_i + 0.223u_i)$$

and fit standard logistic regression model with x_i.

n=100					n=500		
$-$	df	Mean	Var	%Rej	Mean	Var	%Rej
$\hat{C}_g$	8	7.44	11.13	1.50	7.35	12.62	3.20
IM	3	3.01	6.05	5.50	2.38	4.15	1.90
IM_{DIAG}	2	1.82	3.06	3.3	2.05	3.46	4.80
RSS	1	0.92	1.51	4.10	0.99	1.73	4.50

Table 7.15: Results of N=1000 simulation with sample size n=100 and n=500 under missing covariate model

Table 7.15, shows results from simulation study under alternative hypotheses missing covariate model.We can see that the mean and variance of all statistics close to *df* and 2*df*, but we have slightly smaller variance in case of $\hat{C}_g$. However, we have low power in all cases when used IM statistics in case of sample size $n = 500$, IM_{DIAG} statistic and RSS in case of sample size $n = 100$ and $\hat{C}_g$ statistic in both cases of sample size.

Chapter 8

Bootstrap Version and Power

8.1 Introduction

In the previous chapter we have produced a new form test $IMTR$ statistic, to remove the strange behaviour in the distribution of the full version of the statistic, but, it is based on large samples and may be inaccurate and misleading in small samples. In this chapter we will investigate the results for the new form of statistic $IMTR$, as introduced in previous chapter, using the bootstrap. We consider strategy to compute and investigate the p-value of the bootstrap. Inferences and accurate standard errors for parameters and mean functions require distribution assumptions and, often, large sample size. In small samples standard statistical method can be misleading, in this case a bootstrap can be used for test. Bootstrap methods discussed by Efron (1979), which shows in principle, bootstrap methods are more widely applicable and require few assumptions. As such, the bootstrap should provide a valid test of the null hypotheses that the model is correctly specified.

8.2 The Basic Idea of the Bootstrap Method

Suppose we have a sample $x_1, x_2, \ldots, x_n$ draw from any distribution , say, such as Normal distribution. The sample values are thought of as the outcomes of independent and identically distributed random variables $X_1, X_2, \ldots, X_n$. The sample is to be used to make inferences about a population characteristic, the equation here is, what is confidence interval for the population median? We can find an approximate answer to this by compute the median of the random sample $x_1^*, x_2^*, \ldots, x_n^*$, which is simulated from the known distribution say G, and repeat this simulation B times to find confidence interval for the median. In most cases, G will not be actually known , and so, this simulation is not available. Efron (1979), pointed out bootstrap method, that the observed data can be used to estimate G, and can sample from the estimate $\hat{G}$. Obtaining a random sample $x_1^*, x_2^*, \ldots, x_n^*$, from amounts to sampling with replacement from the $x_1, \ldots, x_n$ observed values and then, repeat this B times, see Efron and Tibshirani (1993), Davison and Hinkley (1997) and as applied example see Weisberg (2005, p.244).

8.3 Hypothesis Testing with the Bootstrap

The bootstrap methods are most naturally used to compute confidence intervals but can be adapted for hypothesis testing. It is important, when using the bootstrap for perform hypothesis tests, that at same stage of the calculation, the assumption incorporated in to the arithmetic. We will consider, as an example the one sample problem, where the single unknown probability distribution G produces the data set X by random sampling $X = (x_1, \ldots, x_n)$. We have calculate a statistic of interest from X say the mean $E(X) = \mu$.

In this situation we wish to test the null hypothesis

$$H_0 : E(X) = \mu_0,$$

where, $E(X) = \mu_x$ is unknown. Now we need to investigate this hypotheses and we could use a t-statistic i.e.

$$t_{obs} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

where s is the standard deviation of the set of observations x. Instead, we will work to evaluate the significance of t_{obs} by bootstrapping the data set, so, in each bootstrap sample compute the statistic

$$t_b^* = \frac{\tilde{x}^* - \mu_0}{s^*/\sqrt{n}},$$

where, $b = 1, \ldots, B$. Note that the empirical distribution $\hat{G}$ is not an appropriate estimate for G, because it does not obey H_0. Consequently, some care is needed in the definition of $\tilde{x}^*$ and s^*. We translate the empirical distribution $\hat{G}$, so, that it has the desired mean: we use as our estimated null distribution the empirical distribution of the values:

$$\tilde{x}_i = x_i - \bar{x} + \mu_0 \tag{8.1}$$

for each bootstrap sample, as this is guaranteed to have a mean μ_0. $\tilde{x}^*$ is the mean of (8.1) for the b^{th} bootstrap sample and s^* in the corresponding SD.

Finally, we have observed statistic t_{obs}, the achieved significance level of the test in this case is defined to be the probability of observing at least that large a value when the H_0 is true, $Prob_{H_0}\{|t^*| \geq |t_{obs}|\}$, and the probability of the bootstrap is

$$P_{Boot} = \frac{\#(|t_b^*| \geq |t_{obs}|)}{B}.$$

8.4 Bootstrap the IMT Statistic

In this part we are interested to use the idea of the bootstrap test on the IMT statistic, to confirm the behaviour distribution of the IMT statistic. As we found in the previous chapter that there was poor asymptotic behaviour for the distribution of some forms of IMT and we proposed a new form to avoid this problem. In fact, we have applied this in large sample sizes, which appeared to give reasonable results for use of the new form of statistic. To confirm of these results we use the bootstrap test to investigate the behaviour of IMT distribution.

8.4.1 Design the Test

So, as we discussed in the previos chapter we have a sample $D = (d_1, d_2, \ldots, d_n \in R^q)$, of differences between the two vectorized forms of the information matrix and we wish to test the null hypotheses:

$$H_0 : E(d_i) = 0$$

Let consider the IMT_{obs} statistic for the observed sample, i.e.

$$IMT_{obs} = n\bar{d}^T V_E^{-1}\bar{d},$$

where, V_E is the empirical variance of the statistic. So, to evaluate the significance of the IMT_{obs} by bootstrapping the data set D, we obtain bootstrap samples $D^* = (d_1^*, d_2^*, \ldots, d_n^*)$, and in each bootstrap sample we calculate,

$$IMT_b^* = n(\bar{d}^* - \bar{d})^T V_E^{*-1}(\bar{d}^* - \bar{d}),$$

where, $b = 1, 2, \ldots, B$ and V_E^* is the empirical variance of the bootstrap sample. We must subtract $\bar{d}$ from each $\bar{d}^*$ in order to ensure that we are sampling under H_0. Now, the probability of the test under H_0 is true, $Prob_{H_0}\{IMT^* \geq IMT_{obs}\}$, and then the p-value of bootstrap in this case is

$$P_{Boot} = \frac{IMT_b^* \geq IMT_{obs}}{B}$$

8.5 Simulation study of Bootstrap

Our central goal of this simulation is to assess the power of the two version of the statistic, namely IMT and $IMTR$ by the bootstrap test, under using the empirical covariance matrix. To confirm the results previously found on the behaviour of the distribution of IMT, the strategy of this simulation is first to compute the IMT and $IMTR$ when the model is correctly specified and find P_{Boot}. If we repeat this procedure in N simulation then the P_{Boot} values obtained should come from a $U[0,1]$ distribution.

8.5.1 Design of Simulation

We design this simulation to investigate the statistic IMT and $IMTR$ and the bootstrap test, P_{Boot}. To achieve the goal of this simulation we will consider two cases, when we fitted the true logistic regression model and when missing covariates model has been fitted. So, we designed this simulation in two partes

In the first part we consider the simulatuion under H_0 when the model is correctly specified:

- We consider three cases of true model with different chosen parameters:

$$\pi_i = \text{expit}(0.5 + 0.8x_{i1} + 0.6x_{i2} + 1.2x_{i3}),$$

$$\pi_i = \text{expit}(0.9 + 1.3x_{i1} + 1.1x_{i2}),$$

$$\pi_i = \text{expit}(0.2 + 0.3x_{i1}).$$

  and fitted the model under H_0 in each cases.

- We choose covariates x as draw from normal distribution $X \sim N_3(0, \Sigma)$.
- We consider the covariance matrix Σ with $\sigma_1^2 = \sigma_3^2 = 4, \sigma_2^2 = 9$ and the correlation is $\rho = 0.5$.
- We compute d_i and calculate observed statistic IMT_{obs}.
- Calculate bootstrap statistic IMT^* in each bootstrap sample.
- Calculate p-value P_{Boot}, and investigate its distribution using a histogram.
- The sample size and bootstrap sample are $n = B = 500$ and N=1000 is the number of simulations.

In the second part of this simulation we consider to investigate the behaviour distribution of the $IMTR$ under under H_1 when the model is mis-specified.

- We have consider three cases of the true logistic model with two covariates:

$$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 2x_{i2})$$

$$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 1.5x_{i2})$$

$$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 0.5x_{i2})$$

  and we fit the model when omitted the last covariates x_{i2}.

- Calculate the P_{Boot} in case of $IMTR$.

- The sample size and bootstrap sample are $n = B = 500$ and $N = 5000$ is the number of simulations.

8.5.2 Results and Discussion Under True Model

The first part of this simulation designed to compute the P_{Boot} for fitted the logistic model under H_0, as shown in Figure 8.1. In this case the true logistic model is

$$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 1.5x_{i2} + 2.5x_{i3}).$$

The above histogram is the histogram of P_{Boot} of IMT statistic computed by full elements of empirical variance. The below histogram is the histogram of P_{Boot} of new form statistic $IMTR$. It is clear that the histogram of P_{Boot} for IMT are not uniformly distribution on the interval $[0, 1]$, indicating that the χ^2 properties to the H_0 is very poor. However, the results are shown in the below histogram, the value of P_{Boot} for the new form of statistic $IMTR$ is uniformly distribution on the interval $[0, 1]$, which is indicating that the behaviour distribution of the $IMTR$ statistic is χ^2 distributed. The second proposed fitted the logistic model under H_0 and we considers the true model is

$$\pi_i = \text{expit}(0.9 + 1.3x_{i1} + 1.1x_{i2}).$$

The results of fitted the above model shown in Figure 8.2, the above histogram concerning to the values of the P_{Boot} for IMT statistic and the below histogram concerning to the values of the P_{Boot} for $IMTR$ statistic. We can see clearly that the values of P_{Boot} for IMT are not uniformly distribution and P_{Boot} for $IMTR$ are uniformly distribution. The final proposed fitted the logistic model, when the true model is

$$\pi_i = \text{expit}(0.4 + 0.8x_{i1}).$$

The results shown in Figure 8.3, the above histogram reported the P_{Boot} for IMT which appeared are not uniformly distribution and the second histogram reported P_{Boot} for $IMTR$ which is appeared are uniformly distribution.

Over all, we can say in all cases of different parameters for fitted true logistic model, the values of the P_{Boot} for the statistic $IMTR$ has a distribution much clear to uniform than that for IMT. These results confirm the results of previous chapter and consider the statistic $IMTR$ works well as approaches to avoid the singular problem.

To more illustrate the behaviour of p-value, we have in the results N of p-values where N is the number of simulations, and those are supposedly from a uniform $[0, 1]$ distribution. The test of this can be found from

$$DU = \sum_{g=1}^{10} \frac{(p_g - \frac{1}{10}N)^2}{\frac{1}{10}N}$$

where p_g is the number of p-value in the g^{th} group, i.e. $[0, 0.1), [0.1, 0.2), \ldots, (0.9, 1]$. So this should be χ^2_9 if the p-values are uniform. We calculate DU in all cases of models in simulation and the results are shown in Table 8.1. In this case the $E(DU) = 9$, and we can see clearly the results gave reasonable values of DU in case of $IMTR$. However, in case of IMT statistic the values of DU appeared far away from the normal behaviour for a uniform distribution.

	DU	
The true model	IMT	$IMTR$
$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 1.5x_{i2} + 2.5x_{i3})$	323.4	6.04
$\pi_i = \text{expit}(0.9 + 1.3x_{i1} + 1.1x_{i2})$	354 .21	13.33
$\pi_i = \text{expit}(0.4 + 0.8x_{i1})$	7020.48	20.96

Table 8.1: Results of the calculation of DU values in three cases of models.

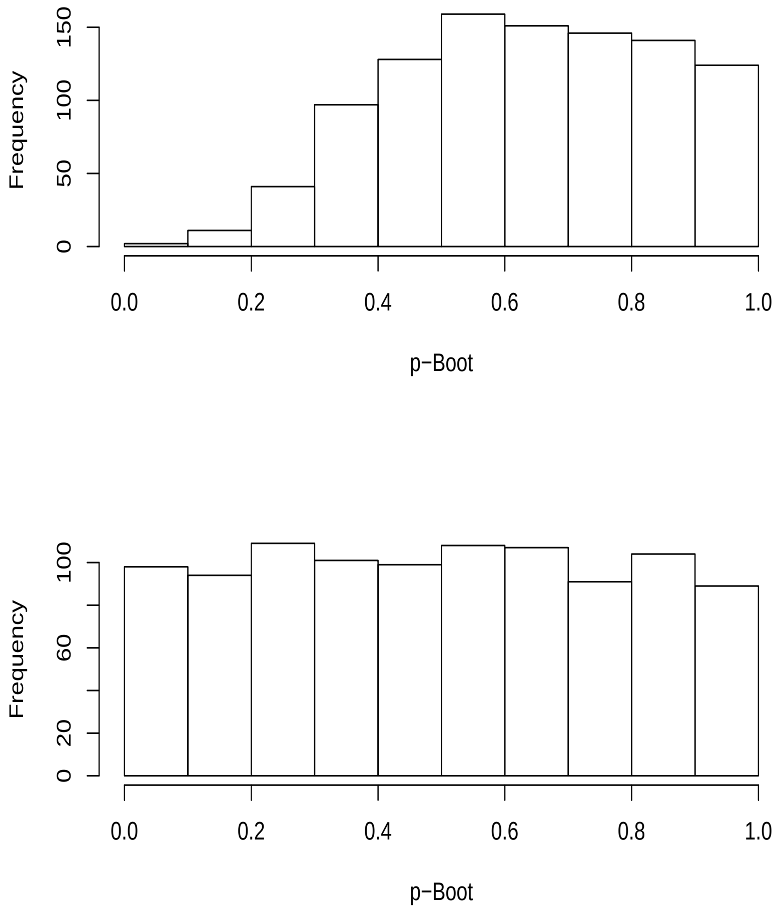

Figure 8.1: Histogram bootstrap of the P_{Boot} of IMT and $IMTR$ respectively, under true model given by $\alpha = 0.5, \beta_1 = 0.8, \beta_2 = 0.6, \beta_3 = 1.2$, sample size and bootstrap sample is n=B=500 and N=1000 number of simulation.

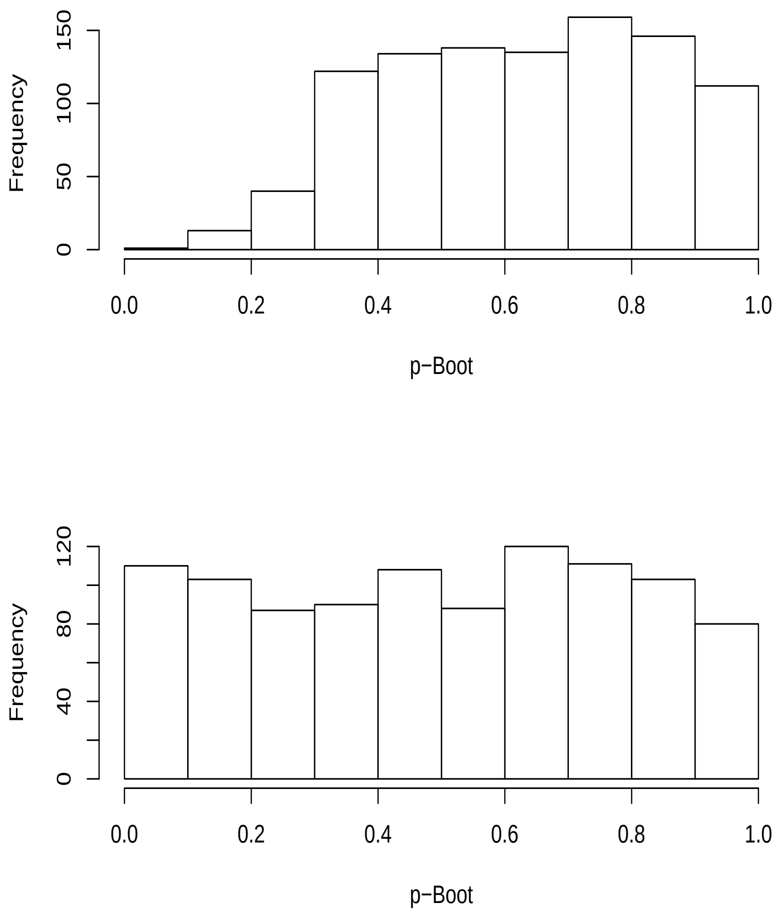

Figure 8.2: Histogram bootstrap of theP_{Boot} of IMT and $IMTR$ respectively, under true model given by $\alpha = 0.9, \beta_1 = 1.3, \beta_2 = 1.1$, sample size and bootstrap sample is n=B=500 and N=1000 number of simulation.

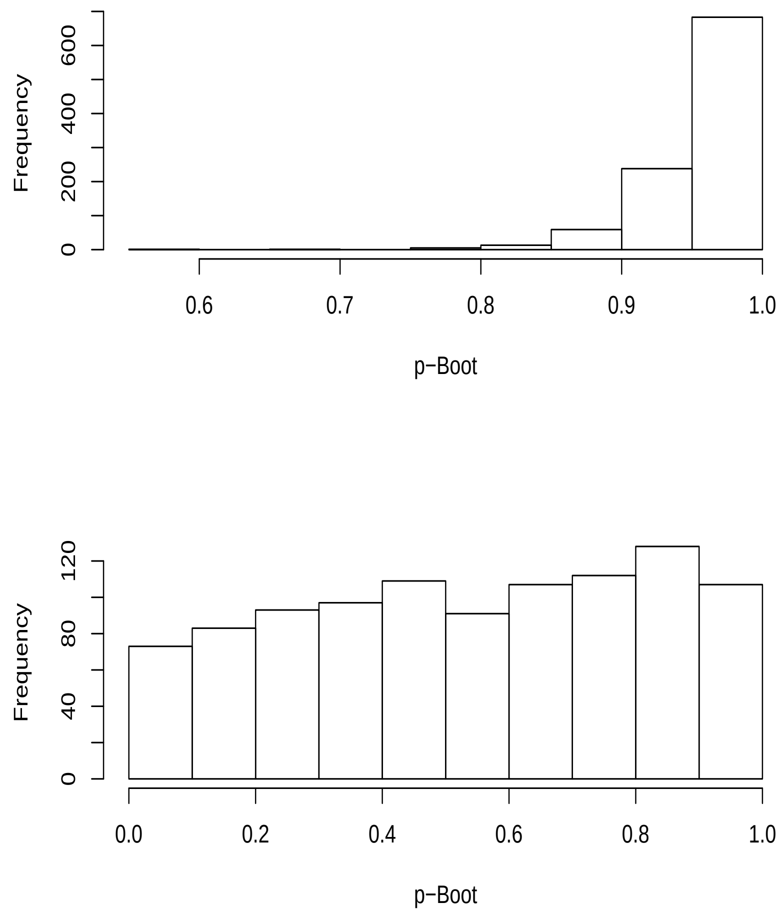

Figure 8.3: Histogram bootstrap of the P_{Boot} of IMT and $IMTR$ respectively, under true model given by $\alpha = 0.2, \beta_1 = 0.3$, sample size and bootstrap sample is n=B=500 and N=1000 number of simulation.

8.5.3 Results and Discussion Under Mis-spesification

The second part of simulation concerning to computed the P_{Boot} for $IMTR$ only under fitting missing covariates logistic model. To more investigate the behaviour of $IMTR$, we made comparison the histogram of P_{Boot} for $IMTR$ between true model and missing covariates model. We considered three cases of the true logistic model with two covariates and we fitted the logistic model when x_{i2} omitted. The results shown in Figure 8.4, 8.5 and 8.6 the above histogram denoted to the P_{Boot} for $IMTR$ under H_0 as before we can see P_{Boot} are uniformly distribution in all cases. However, the histogram below concerning to the P_{Boot} for $IMTR$ under H_1 mis-specified model, it is clear that the value of P_{Boot} tend quick to zero which meaning we reject the null hypotheses.

Moreover, if we calculate the value of DU in each case, the results shows in Tabel 8.2.

In fact, we support the behaviour distribution of P_{Boot} under mis-specification which the value is very small related to reject H_0. Finally, from these results it seems clearly to say the $IMTR$ has reasonable behaviour.

	DU of the $IMTR$	
The true model	Under H_0	Under H_1
$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 2x_{i2})$	50.9	239.12
$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 1.5x_{i2})$	15.10	404.57
$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 0.5x_{i2})$	8.52	135.88

Table 8.2: Results of the calculation of DU values in three cases of models under H_0 and H_1.

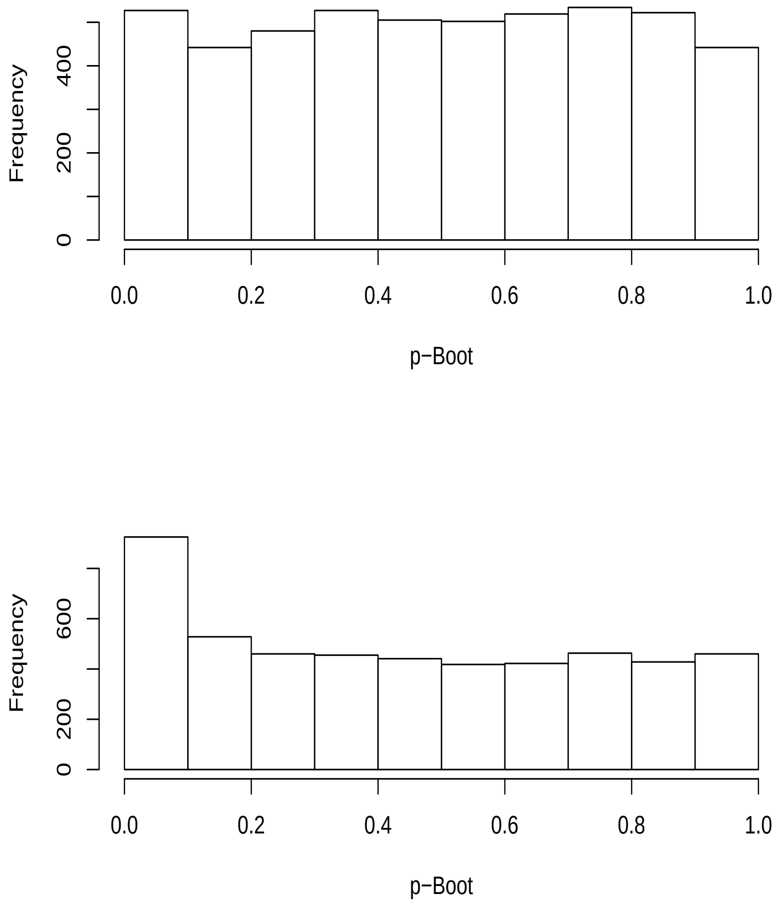

Figure 8.4: Histogram bootstrap of the P_{Boot} of $IMTR$, for fitted true logistic model with two covariates with $\alpha = 0.4, \beta_1 = 0.8, \beta_2 = 1.5$, and missing one covariate model respectively, sample size and bootstrap smple is n=B=500 and N=5000 number of simulation.

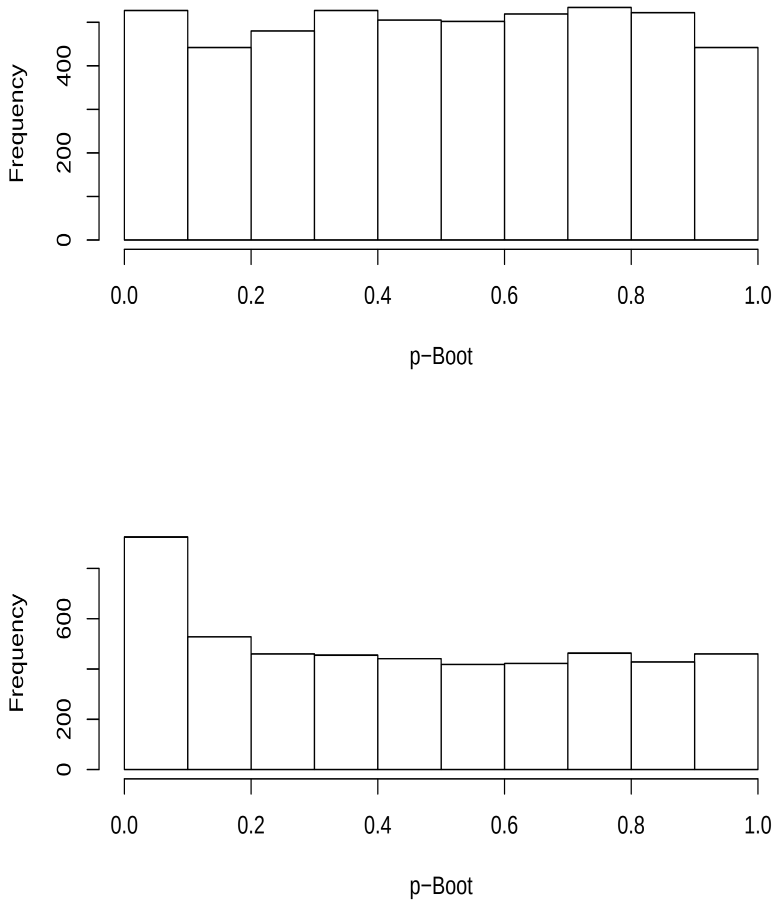

Figure 8.5: Histogram bootstrap of the P_{Boot} of $IMTR$, for fitted true logistic model with two covariates with $\alpha = 0.4, \beta_1 = 0.8, \beta_2 = 0.5$, and missing one covariate model respectively, sample size and bootstrap smple is n=B=500 and N=5000 number of simulation.

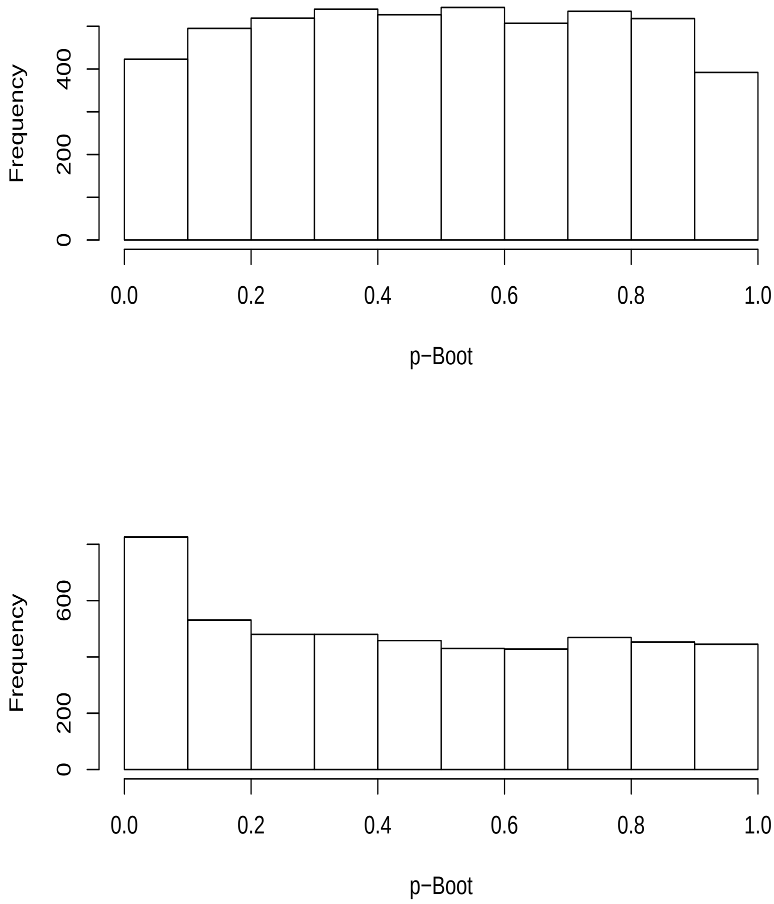

Figure 8.6: Histogram bootstrap of the P_{Boot} of $IMTR$, for fitted true logistic model with two covariates with $\alpha = 0.4, \beta_1 = 0.8, \beta_2 = 2$, and missing one covariate model respectively, sample size and bootstrap smple is n=B=500 and N=5000 number of simulation.

8.6 Power of Tests Consideration

In the early part of this chapter we examined the proposed new test statistic $IMTR$ by bootstrap test. In Figure 8.4, there appeared to be large numbers of small p-values less than 0.1. We need to the power of the bootstrap test by calculation the power under a mis-specified model with using asymptotic distribution for $IMTR$. We also attempt to calculate the power theoretically and compare with the empirical values.

8.6.1 Definition

When the null hypothesis H_0 is true and all assumptions are met, the chance of incorrectly declaring H_0 to be false at level α, is only α. If $\alpha = 0.1$, then in 10%, of tests we will get a $p - values \leq 0.1$. When H_0 is false, we will expect to see small p-values more often. The power of a test is defined to be the probability of rejecting H_0 under a given alternative. This definition pointed out by Weisberg (2005, p.31).

8.6.2 Simulation of Power Calculation

In this part we are interested to show the design of the simulations and calculation of the bootstrap, empirical and theoretical power. We consider the mis-specified model which we fitted in the previous simulation 8.5.

Bootstrap Power

The result of the previous simulation gives 5000 values of $IMTR$, so, we used these values to calculate the power of the bootstrap test. In this case we will calculate the bootstrap power as

$$Power_B = \frac{\#(P_{Boot} < 0.1)}{N}$$

where N the number of simulation.

Empirical Power

To calculate the empirical power, we know that the statistic $IMTR$ is asymptotically central χ^2 on 1 df when H_0 is true. So we need to calculate the power as

$$Power_E = p(IMTR > criticalregion_{H_0})$$

where critical region point under $\alpha = 0.1$ is $IMTR > 2.71$.

Theoretical Power

To calculate the theoretical power we assume that under H_1, $IMTR$ is a non-central χ^2 on ν df, with non-centrality parameters λ. Thus

$$E(IMTR) = \nu + \lambda$$

and the statistic is

$$IMTR = nd^T V_E^{-1} d.$$

The mean and variance of d are

$$E(d) = \mu \;, \mathsf{var}(d) = \frac{V}{n}$$

and

$$E(dd^T) = \mathsf{var}(d) + \mu\mu^T.$$

Then, we can write

$$E(IMTR) = Etr(nd^T V_E^{-1} d) \simeq tr(nV^{-1}\frac{V}{n}) + n\mu^T V^{-1}\mu = rank(V) + n\mu^T V^{-1}\mu$$

so, the non-centrality parameter is

$$\lambda = n\mu^T V^{-1}\mu.$$

Recall that to evaluated μ, V for specified values of the parameters under H_1, we use the simulation method from chapter 6.

The theoretical power is found as

$$Power_T = \Pr(\chi^2_\nu(\lambda) > criticalregion_{H_0})$$

which is in our case

$$Power_T = \Pr(\chi^2_1(\lambda) > 2.71)$$

8.6.3 Results and Discussion

As we consider the simulation designed to investigate the power by compare the bootstrap power with empirical and theoretical power. The simulation consider $\alpha = 0.1$, and the results shows in Table 8.3.

If we compare between the values of power, we can say in general the values are comparable for the empirical powers. The theoretical power agrees less well. The reason are almost certainly do with the difficulty of computing λ. This is a matter for further research.

The true model	$Power_E$	$Power_B$	$Power_T$
$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 2x_{i2})$	0.207	0.163	0.103
$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 1.5x_{i2})$	0.218	0.174	0.106
$\pi_i = \text{expit}(0.4 + 0.8x_{i1} + 5x_{i2})$	0.197	0.146	0.102

Table 8.3: Results of the power calculation under $\alpha = 0.1$ for three cases of models.

Chapter 9

Conclusions and Further Work

The work considered in this thesis was centred on the behaviour of maximum likelihood method estimators for a logistic regression under a mis-specified model. We also considered the information matrix test for logistic regression model. The early part of this thesis outlined the maximum likelihood method under missing covariates logistic model. The behaviour of ML method when the assumed model is incorrect is important to find the estimation of the unknown parameters in terms of parameters of the true logistic model, as pointed out by Claeskens and Hjort (2008).

In Chapter 2 we addressed this problem to find a new closed form for the least false values of the parameters obtained by maximising the incorrect likelihood function. In this situation the approximation of the logit by the probit and the properties of the skew-normal distribution were used to compute a good approximation to the least false values under wrong model. Corresponding simulations investigated this form of the least false values, when the covariates are drawn from the multivariate normal distribution: we have found a good agreement in all cases of different variance and correlation. The estimated parameters values and the least false values gave a ratio very close to one between calculated and simulated values. There were slightly different results when $\Pr(Y = 1) \simeq 10\%$, where the formulae appeared slightly sensitive to negative correlation. Notice that if $\beta_a = 0$, i.e. the fitted model is correct, then the least false values are the true parameter values. However, if $\beta_a \neq 0$, then, unlike in this case of a normal linear model where omitting covariates that are uncorrelated with the fitted covariates has no effect on the expectation of $\hat{\beta}_f$, β_f^* will be shrunk towards zero compared with β_f, even if $\Omega_{af} = 0$. Indeed, for given Ω_{aa} the bias will be maximised when $\Omega_{af} = 0$, because the fitted covariates will, in this case, be unable to act as a proxy for the omitted variables.

Chapter 3, considered different assumptions on covariates. First we considered symmetric distributions different from multivariate normal distribution. We considered multivariate t-distribution and multivariate uniform distribution. In fact, with uniform covariates we have found the same behaviour results as found when the covariates draw from normal distribution. If we consider the model with covariates draw from multivariate t-distribution, when the degree of freedom large enough the results almost have the same behaviour in case of normal distribution covariates. However, when the degree of freedom small the ratios appeared slightly affected in case of the sample size $n = 500$. The case of skewed covariates was investigated by using a bivariate distribution including variables from the log-normal distribution. We considered three cases: $(\exp(X_1), X_2)$, $(X_1, \exp(X_2))$ and $(\exp(X_1), \exp(X_2))$, where (X_1, X_2) is bivariate normal. We have found the value of the parameters estimated are close to the least false value which computed by our proposed formulae, where the value of the ratio between them was found close to one in almost cases considered. However, slightly different found by low ratio when the value of correlation was negative and $\Pr(Y = 1) \simeq 10\%$. Some discrepancy was in the ratios noticed when the value of estimated coefficients were very close to zero.

In Chapter 4, a categorical variable is introduced and the least false values computed when one of the covariates is binary and some of the normal covariates are omitted. New formulae were found for the least false values, simulations confirmed our results. Moreover, an application to randomized trials is considered, with an example real data produced by Fleming and Harrington (2005). We computed accurate closed-form approximations for the asymptotic bias, when the important normal distribution covariates are omitted from a model. Our work is a reminder that, even when treatments are allocated at random, the adjusted log odd-ratio is asymptotically biased unless the correct covariates are included in the model. In most circumstances this model will be unknown and in most of the circumstances we describe the misspecification gives rise to a least false value that is shrunk towards zero compared with the true value. The degree of shrinkage depends on the conditional variance of the omitted variables given the fitted variables.

The second part of this thesis considered the information matrix test IMT statistic proposed by White (1982) and investigated by Kuss (2002), who found it had good power for the logistic regression model. The idea and procedure of IMT was discussed in great detail in Chapter 5.

In Chapter 6, we have investigated the distribution of the first two moments of the statistic, and we computed the form of the mean and variance of the statistic and produced a formulae for variance. We were able to use our results as least false values to compute the dispersion matrix, under the wrong model, for IMT and IMT_{DIAG}. We also computed the empirical dispersion matrix. The simulation designed to compare between the diagonal elements of dispersion matrix of empirical V_E, theoretical V_T vs. true dispersion matrix V_{tr}. We found in almost all cases of different parameters and variance, the theoretical variance and empirical variance are close to the true value of variance under true model and also under missing covariate model. The first element was much closer to zero than of the rest. Some slightly strange ratios arose in case of sample size $n = 500$ and with small values of $S.D(\pi)$.

In Chapter 7, we considered investigated the asymptotic distribution of IMT statistic under H_0 specified model and under H_1 mis-specified model. We produced a new form statistic $IMTR$ to avoid the near singularity problem which affected the behaviour of the statistic, by removing overlap with elements of the log-likelihood function. Our proposed form investigated by simulation, was found to have reasonable asymptotic behaviour, with the mean and the variance appearing to be very close to the properties of the χ^2 distribution.

In Chapter 8, we considered Bootstrap test proposed by Efron (1979) as to confirm the results in Chapter 7, which found support results for $IMTR$ statistic. The histogram of the values of P_{Boot} appeared uniformly distribution for $IMTR$, but disagree for IMT under H_0. Moreover, the values of DU were more reasonable in case of P_{Boot} values for $IMTR$ and close to the uniform distribution if compared with the P_{Boot} for IMT. However, the bootstrap test under H_1, the histogram shows there are large number of P_{Boot} have small values. This result investigated by calculation the power of the $IMTR$ statistic, empirical and theoretical power considered. The theoretical power appeared to show strange behavior, the reason related to the difficulty of calculation the non-central parameter of $\chi_\nu(\lambda)$ distribution.

9.1 Further Research Directions

The application of the logistic regression model has been increasingly used in biostatistics; it is important and needed for analysis the medical data. So the development of the model needs more work carefully for some important points under mis-specified model might be needed to investigate. We consider the effect of omitted the covariates on the logistic model, and proposed a new form for the least false values. Our results provide useful insights and extensions to other forms of covariates should be investigated. The assumption on the model and the covariates with small sample size and sparse data might be more sensitive and need more investigate.

The IMT is not widely used in biostatistics and its properties do not seem wholly stable. Our contribution of the reduced version $IMTR$, is helpful in this respect but, further work on its properties under H_1 is needed, before it can be routinely recommended. The statistic form $IMTR$ should be extended with the general multivariate logistic model has p covariates.

Appendix

Appendix I: Additional Table for the Least False Valus

var = 0.1			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta}_1$	β_1^*	R_2	$\hat{\beta}_2$	β_2^*	R_3
0.1	0.2	0.3	0.3931	0.3973	0.99	0.4197	0.4109	1.02	0.5002	0.5102	0.98
0.2	0.1	0.1	0.3995	0.3977	1.00	0.3251	0.3231	1.01	0.4212	0.4226	1.00
0.3	0.1	0.1	0.3954	0.3977	0.99	0.3310	0.3174	1.04	0.4083	0.4168	0.98
0.3	0.4	0.1	0.3992	0.3990	1.00	0.5265	0.5256	1.00	0.6213	0.6253	0.99
0.9	0.9	0.9	0.3992	0.3995	1.00	0.6753	0.6755	1.00	0.7683	0.7754	0.99
0.5	0.7	0.5	0.4009	0.3998	1.00	0.6724	0.6697	1.00	0.7635	0.7697	0.99
0.1	-0.5	0.4	0.3972	0.3991	0.99	-0.1630	-0.1580	1.03	-0.0580	-0.0581	1.00
0.8	-0.6	0.7	0.3977	0.3977	1.00	-0.0630	-0.0490	1.20	0.0692	0.0497	1.30
var = 0.5											
0.1	0.2	0.3	0.3856	0.3872	1.00	0.3968	0.5004	0.99	0.4972	0.3872	1.01
0.2	0.1	0.1	0.3915	0.3891	1.01	0.3137	0.3161	0.99	0.4089	0.4134	0.99
0.3	0.1	0.1	0.3773	0.3890	0.97	0.3066	0.3105	0.99	0.4029	0.4077	0.99
0.3	0.4	0.1	0.3940	0.3951	1.00	0.5192	0.5205	1.00	0.6106	0.6193	0.99
0.9	0.9	0.9	0.3961	0.3976	1.00	0.6798	0.6724	1.01	0.7683	0.7718	0.99
0.5	0.7	0.5	0.3995	0.3992	1.00	0.6689	0.6687	1.00	0.7702	0.7686	1.00
0.1	-0.5	0.4	0.3875	0.3955	0.98	-0.1590	-0.1570	1.01	-0.0570	-0.0580	0.98
0.8	-0.6	0.7	0.3817	0.3890	0.98	-0.0630	-0.0480	1.30	0.0656	0.0486	1.30
var = 1.5											
0.1	0.2	0.3	0.3640	0.3650	1.00	0.3722	0.3774	0.99	0.4460	0.4687	0.95
0.2	0.1	0.1	0.3660	0.3697	0.99	0.3123	0.3004	1.03	0.3820	0.3928	0.97
0.3	0.1	0.1	0.3629	0.3696	0.98	0.2840	0.3874	0.96	0.3760	0.2950	0.97
0.3	0.4	0.1	0.3755	0.3859	0.97	0.5038	0.5084	0.99	0.5986	0.6049	0.99
0.9	0.9	0.9	0.3906	0.3931	0.99	0.6580	0.6647	0.99	0.7604	0.7630	1.00
0.5	0.7	0.5	0.3979	0.3978	1.00	0.6645	0.6664	1.00	0.7638	0.7658	1.00
0.1	-0.5	0.4	0.3855	0.3871	0.99	-0.1470	-0.1530	0.96	-0.0560	-0.0570	0.99
0.8	-0.6	0.7	0.3646	0.3696	0.99	-0.0550	-0.0460	1.10	0.0567	0.0460	1.20

Table 1: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Normal distribution in case $\Pr(Y = 1) \simeq 60\%$

var = 0.1			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta}_1$	β_1^*	R_2	$\hat{\beta}_2$	β_2^*	R_3
0.1	0.2	0.3	-2.187	-2.185	1.00	0.4149	0.4108	1.01	0.4987	0.5102	0.98
0.2	0.1	0.1	-2.185	-2.187	1.00	0.3140	0.3231	0.97	0.4254	0.4226	1.01
0.3	0.1	0.1	-2.187	-2.187	1.00	0.3144	0.3174	0.99	0.4337	0.4168	1.04
0.3	0.4	0.1	-2.196	2.194	1.00	0.5336	0.5256	1.02	0.6188	0.6253	0.99
0.9	0.9	0.9	-2.19	-2.197	1.00	0.6800	0.6755	1.01	0.7709	0.7754	0.99
0.5	0.7	0.5	-2.19	-2.199	1.00	0.6656	0.6697	0.99	0.7702	0.7697	1.00
0.1	-0.5	0.4	-2.190	2.195	1.00	-0.1540	-0.1580	0.97	-0.0630	-0.0580	1.06
0.8	-0.6	0.7	-2.180	-2.187	1.00	-0.0550	-0.0490	1.10	0.0521	0.0497	1.05
var = 0.5											
0.1	0.2	0.3	-2.125	-2.129	1.00	0.4118	0.4004	1.03	0.51167	0.4972	1.03
0.2	0.1	0.1	-2.130	-2.140	1.00	0.3050	0.3161	0.97	0.4216	0.4134	1.02
0.3	0.1	0.1	-2.134	-2.139	1.00	0.3132	0.3105	1.01	0.4041	0.4077	0.99
0.3	0.4	0.1	-2.176	-2.173	1.00	0.5259	0.5205	1.01	0.6251	0.6193	1.01
0.9	0.9	0.9	-2.188	-2.187	1.00	0.6626	0.6724	0.99	0.7831	0.7718	1.01
0.5	0.7	0.5	-2.197	-2.196	1.00	0.6738	0.6687	1.01	0.7662	0.7686	1.00
0.1	-0.5	0.4	-2.177	-2.175	1.00	-0.1520	-0.1570	0.97	-0.0590	-0.0580	1.01
0.8	-0.6	0.7	-2.133	-2.139	1.00	-0.0610	-0.0480	1.20	0.0589	0.0486	1.20
var = 1.5											
0.1	0.2	0.3	-2.001	-2.007	1.00	0.3774	0.3774	1.00	0.4684	0.4687	1.00
0.2	0.1	0.1	-2.037	-2.033	1.00	0.3163	0.3004	1.05	0.3888	0.3928	0.99
0.3	0.1	0.1	-2.021	-2.033	0.99	0.2894	0.2950	0.98	0.3921	0.3874	1.01
0.3	0.4	0.1	-2.120	-2.122	1.00	0.5099	0.5084	1.00	0.6093	0.6049	1.01
0.9	0.9	0.9	-2.164	-2.162	1.00	0.6497	0.6647	0.98	0.7782	0.7630	1.02
0.5	0.7	0.5	-2.192	-2.188	1.00	0.668	0.6664	1.00	0.7672	0.7658	1.00
0.1	-0.5	0.4	-2.115	-2.129	0.99	-0.1510	-0.1530	0.98	-0.0660	-0.0570	1.10
0.8	-0.6	0.7	-2.007	-2.032	0.99	-0.0490	-0.0460	1.08	0.0473	0.0462	1.02

Table 2: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Normal distribution in case $\Pr(Y = 1) \simeq 10\%$

var $= 0.1$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.2	0.3	0.3876	0.3756	1.03	0.4080	0.3884	1.05	0.4970	0.4823	1.03
0.2	0.1	0.1	0.3914	0.3790	1.03	0.3239	0.3079	1.05	0.4167	0.4027	1.03
0.3	0.1	0.1	0.3962	0.3789	1.05	0.3085	0.3024	1.02	0.4172	0.3972	1.05
0.3	0.4	0.1	0.3970	0.3904	1.02	0.5157	0.5143	1.003	0.6092	0.6120	0.99
0.9	0.9	0.9	0.3984	0.3954	1.01	0.6825	0.6685	1.02	0.7633	0.7674	0.99
0.5	0.7	0.5	0.4004	0.3985	1.01	0.6601	0.6676	0.99	0.7670	0.7672	1.00
0.1	-0.5	0.4	0.3958	0.3913	1.01	-0.1390	-0.1550	0.90	-0.0490	-0.0570	0.86
0.8	-0.6	0.7	0.3970	0.3789	1.05	-0.0360	-0.0470	0.77	0.0444	0.0470	0.94
var $= 0.5$											
0.1	0.2	0.3	0.3755	0.3756	1.00	0.3705	0.3884	0.95	0.4621	0.4823	0.96
0.2	0.1	0.1	0.3614	0.3790	0.95	0.2993	0.3079	0.97	0.3800	0.4027	0.94
0.3	0.1	0.1	0.3647	0.3789	0.96	0.2940	0.3024	0.97	0.3841	0.3972	0.97
0.3	0.4	0.1	0.3843	0.3904	0.98	0.5043	0.5143	0.98	0.6006	0.6120	0.98
0.9	0.9	0.9	0.3878	0.3954	0.98	0.6585	0.6685	0.98	0.7655	0.7674	1.00
0.5	0.7	0.5	0.3985	0.3985	1.00	0.6663	0.6676	1.00	0.7576	0.7672	0.99
0.1	-0.5	0.4	0.386	0.391	0.99	-0.147	-0.155	0.94	-0.046	-0.057	0.80
0.8	-0.6	0.7	0.3690	0.3780	0.97	-0.052	-0.0470	1.09	0.0560	0.0470	1.19
var $= 1.5$											
0.1	0.2	0.3	0.3275	0.3756	0.87	0.3369	0.3884	0.87	0.4125	0.4823	0.86
0.2	0.1	0.1	0.3200	0.3790	0.84	0.2727	0.3079	0.88	0.3530	0.4027	0.88
0.3	0.1	0.1	0.3392	0.3789	0.90	0.2664	0.3024	0.88	0.3495	0.3972	0.88
0.3	0.4	0.1	0.3653	0.3904	0.94	0.4699	0.5143	0.91	0.5661	0.6120	0.93
0.9	0.9	0.9	0.3803	0.3954	0.96	0.6534	0.6685	0.98	0.7310	0.7674	0.95
0.5	0.7	0.5	0.3962	0.3985	0.99	0.6536	0.6676	0.98	0.7522	0.7672	0.98
0.1	-0.5	0.4	0.356	0.3913	0.91	-0.1330	-0.1550	0.86	-0.0520	-0.0570	0.91
0.8	-0.6	0.7	0.3244	0.3789	0.86	-0.0340	-0.0470	0.73	0.0424	0.0473	0.90

Table 3: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Uniform distribution in case $\Pr(Y = 1) \simeq 60\%$

var = 0.1			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta}_1$	β_1^*	R_2	$\hat{\beta}_2$	β_2^*	R_3
0.1	0.2	0.3	-2.166	-2.066	1.05	0.4058	0.3884	1.04	0.5038	0.4823	1.04
0.2	0.1	0.1	-2.173	-2.084	1.04	0.3114	0.3079	1.01	0.4154	0.4027	1.03
0.3	0.1	0.1	-2.175	-2.084	1.04	0.3182	0.3024	1.05	0.4136	0.3972	1.04
0.3	0.4	0.1	-2.191	-2.147	1.02	0.5119	0.5143	1.00	0.6208	0.6120	1.01
0.9	0.9	0.9	-2.190	-2.170	1.01	0.6787	0.6685	1.02	0.7711	0.7674	1.01
0.5	0.7	0.5	-2.190	-2.192	1.00	0.6634	0.6676	0.99	0.7621	0.7672	0.99
0.1	-0.5	0.4	-2.190	-2.152	1.01	-0.1440	-0.1550	0.93	-0.0490	-0.0570	0.85
0.8	-0.6	0.7	-2.170	-2.084	1.04	-0.0340	-0.0470	0.73	0.0461	0.0470	0.98
var = 0.5											
0.1	0.2	0.3	-2.037	-2.066	0.99	0.3757	0.3884	0.97	0.4846	0.4823	1.00
0.2	0.1	0.1	-2.069	-2.084	0.99	0.3070	0.3079	1.00	0.3990	0.4027	0.99
0.3	0.1	0.1	-2.071	-2.084	0.99	0.3036	0.3024	1.00	0.4067	0.3972	1.02
0.3	0.4	0.1	-2.138	-2.147	0.99	0.5126	0.5143	1.00	0.6014	0.6120	0.98
0.9	0.9	0.9	-2.170	-2.174	0.99	0.6872	0.6685	1.03	0.7458	0.7674	0.97
0.5	0.7	0.5	-2.180	-2.192	0.99	0.6615	0.6676	0.99	0.7603	0.7672	0.99
0.1	-0.5	0.4	-2.134	-2.152	0.99	-0.1410	-0.1550	0.91	-0.0380	-0.0570	0.66
0.8	-0.6	0.7	-2.070	-2.084	0.99	-0.0400	-0.0470	0.85	0.0634	0.0473	1.30
var = 1.5											
0.1	0.2	0.3	-1.816	-2.066	0.88	0.3444	0.3884	0.89	0.4387	0.4823	0.91
0.2	0.1	0.1	-1.866	-2.084	0.90	0.2869	0.3079	0.93	0.3647	0.4027	0.91
0.3	0.1	0.1	-1.889	-2.084	0.91	0.2741	0.3024	0.91	0.3650	0.3972	0.92
0.3	0.4	0.1	-2.051	-2.147	0.96	0.4955	0.5143	0.96	0.5781	0.6120	0.94
0.9	0.9	0.9	-2.123	-2.174	0.98	0.6441	0.6685	0.96	0.7553	0.7674	0.98
0.5	0.7	0.5	-2.160	-2.192	0.99	0.6554	0.6676	0.98	0.7504	0.7672	0.98
0.1	-0.5	0.4	-2.026	-2.152	0.94	-0.1420	-0.1550	0.91	-0.0370	0.0570	0.65
0.8	-0.6	0.7	-1.850	-2.084	0.89	-0.0330	-0.0470	0.71	0.0565	0.0473	1.20

Table 4: Simulation results of last false values using different values of ρ_{ij} and variance by generated variables from multivariate Uniform distribution in case $\Pr(Y = 1) \simeq 10\%$

$df = 200$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.2	0.3	0.381	0.387	0.99	0.386	0.400	0.97	0.492	0.497	0.99
0.2	0.1	0.1	0.381	0.389	0.98	0.316	0.316	1.00	0.403	0.413	0.98
0.3	0.1	0.1	0.385	0.389	0.99	0.297	0.310	0.96	0.406	0.407	1.00
0.3	0.4	0.1	0.391	0.395	0.99	0.526	0.520	1.01	0.607	0.619	0.98
0.9	0.9	0.9	0.397	0.397	1.00	0.672	0.672	1.00	0.775	0.772	1.00
0.5	0.7	0.5	0.397	0.399	1.00	0.665	0.668	1.00	0.773	0.768	1.01
0.8	-0.6	0.7	0.376	0.389	0.97	-0.041	-0.048	0.85	0.037	0.048	0.76
$df = 20$											
0.1	0.2	0.3	0.389	0.387	1.01	0.393	0.400	0.98	0.494	0.4972	0.99
0.2	0.1	0.1	0.379	0.389	0.97	0.322	0.316	1.02	0.4059	0.413	0.98
0.3	0.1	0.1	0.383	0.389	0.99	0.295	0.310	0.95	0.404	0.407	0.99
0.3	0.4	0.1	0.387	0.395	0.98	0.518	0.521	0.99	0.6174	0.6193	1.00
0.9	0.9	0.9	0.395	0.397	0.99	0.683	0.672	1.01	0.761	0.771	0.99
0.5	0.7	0.5	0.396	0.399	0.99	0.672	0.668	1.01	0.7651	0.768	1.00
0.8	-0.6	0.7	0.378	0.389	0.97	-0.039	-0.048	0.81	0.037	0.048	0.76
$df = 10$											
0.1	0.2	0.3	0.377	0.387	0.97	0.398	0.400	0.99	0.473	0.497	0.95
0.2	0.1	0.1	0.390	0.389	1.00	0.307	0.316	0.97	0.400	0.413	0.97
0.3	0.1	0.1	0.392	0.389	1.01	0.306	0.311	0.99	0.401	0.408	0.98
0.3	0.4	0.1	0.392	0.395	0.99	0.516	0.521	0.99	0.612	0.619	0.99
0.9	0.9	0.9	0.394	0.397	0.99	0.670	0.672	1.00	0.769	0.771	1.00
0.5	0.7	0.5	0.398	0.399	1.00	0.668	0.669	1.00	0.767	0.769	1.00
0.8	-0.6	0.7	0.390	0.389	1.00	-0.035	-0.048	0.72	0.033	0.048	0.69
$df = 5$											
0.1	0.2	0.3	0.375	0.387	0.97	0.360	0.4004	0.90	0.464	0.497	0.93
0.2	0.1	0.1	0.373	0.389	0.96	0.283	0.316	0.89	0.386	0.413	0.93
0.3	0.1	0.1	0.376	0.389	0.97	0.304	0.311	0.98	0.390	0.408	0.96
0.3	0.4	0.1	0.387	0.512	0.98	0.512	0.521	0.98	0.609	0.619	0.98
0.9	0.9	0.9	0.395	0.397	1.00	0.660	0.672	0.98	0.768	0.771	1.00
0.5	0.7	0.5	0.399	0.399	1.00	0.660	0.668	0.99	0.771	0.768	1.00
0.8	-0.6	0.7	0.376	0.389	0.97	-0.030	-0.048	0.63	0.029	0.048	0.60

Table 5: Simulation results of last false values using different values of ρ_{ij} and variance is 0.5 by generated variables from multivariate t-distribution in case $\Pr(Y = 1) \simeq 60\%$

$df = 200$			Parameters estimated, Least false values and Ratio								
Ω_{11}	Ω_{12}	Ω_{22}	$\hat{\alpha}$	α^*	R_1	$\hat{\beta_1}$	β_1^*	R_2	$\hat{\beta_2}$	β_2^*	R_3
0.1	0.2	0.3	-2.127	-2.129	1.00	0.404	0.400	1.01	0.511	0.497	1.03
0.2	0.1	0.1	-2.135	-2.140	1.00	0.325	0.316	1.03	0.415	0.413	1.00
0.3	0.1	0.1	-2.137	-2.139	1.00	0.332	0.310	1.07	0.415	0.408	1.02
0.3	0.4	0.1	-2.173	-2.173	1.00	0.518	0.520	1.00	0.628	0.619	1.01
0.9	0.9	0.9	-2.193	-2.187	1.00	0.662	0.672	0.99	0.788	0.771	1.02
0.5	0.7	0.5	-2.199	-2.196	1.00	0.667	0.668	1.00	0.771	0.768	1.00
0.8	-0.6	0.7	-2.131	-2.139	1.00	-0.059	-0.048	1.20	0.058	0.049	1.20
	$df = 20$										
0.1	0.2	0.3	-2.120	-2.129	1.00	0.418	0.400	1.04	0.500	0.497	1.01
0.2	0.1	0.1	-2.131	-2.140	1.00	0.322	0.316	1.02	0.430	0.413	1.04
0.3	0.1	0.1	-2.132	-2.139	1.00	0.316	0.311	1.02	0.414	0.408	1.02
0.3	0.4	0.1	-2.174	-2.173	1.00	0.521	0.520	1.01	0.621	0.619	1.00
0.9	0.9	0.9	-2.189	-2.187	1.00	0.685	0.672	1.01	0.757	0.772	0.98
0.5	0.7	0.5	-2.198	-2.196	1.00	0.667	0.668	1.00	0.775	0.769	1.01
0.8	-0.6	0.7	-2.131	-2.139	0.99	-0.031	-0.048	0.65	0.049	0.048	1.02
	$df = 10$										
0.1	0.2	0.3	-2.113	-2.129	0.99	0.392	0.400	0.98	0.489	0.497	0.98
0.2	0.1	0.1	-2.124	-2.140	0.99	0.316	0.316	1.00	0.425	0.413	1.03
0.3	0.1	0.1	-2.117	-2.139	0.99	0.316	0.311	1.02	0.411	0.408	1.01
0.3	0.4	0.1	-2.161	-2.173	0.99	0.518	0.521	0.99	0.616	0.619	0.99
0.9	0.9	0.9	-2.187	-2.187	1.00	0.673	0.672	1.00	0.773	0.771	1.00
0.5	0.7	0.5	-2.196	-2.196	1.00	0.667	0.668	1.00	0.769	0.768	1.00
0.8	-0.6	0.7	-2.12	-2.139	0.99	-0.049	-0.048	1.01	0.054	0.049	1.10
	$df = 5$										
0.1	0.2	0.3	-2.100	-2.129	0.99	0.394	0.400	0.98	0.4849	0.4972	0.98
0.2	0.1	0.1	-2.106	-2.140	0.98	0.319	0.316	1.01	0.4158	0.4134	1.01
0.3	0.1	0.1	-2.103	-2.139	0.98	0.298	0.310	0.96	0.401	0.4077	0.98
0.3	0.4	0.1	-2.159	-2.173	0.99	0.508	0.521	0.98	0.6274	0.6193	1.01
0.9	0.9	0.9	-2.181	-2.187	1.00	0.660	0.672	0.98	0.7754	0.7718	1.01
0.5	0.7	0.5	-2.194	-2.196	1.00	0.668	0.669	1.00	0.7726	0.7686	1.01
0.8	-0.6	0.7	-2.097	-2.139	0.98	-0.071	-0.048	1.40	0.075	0.048	1.50

Table 6: Simulation results of last false values using different values of ρ_{ij} and variance is 0.5 by generated variables from multivariate t-distribution in case $\Pr(Y = 1) \simeq 10\%$

Appendix II: additional Plot of p-value Bootstrap

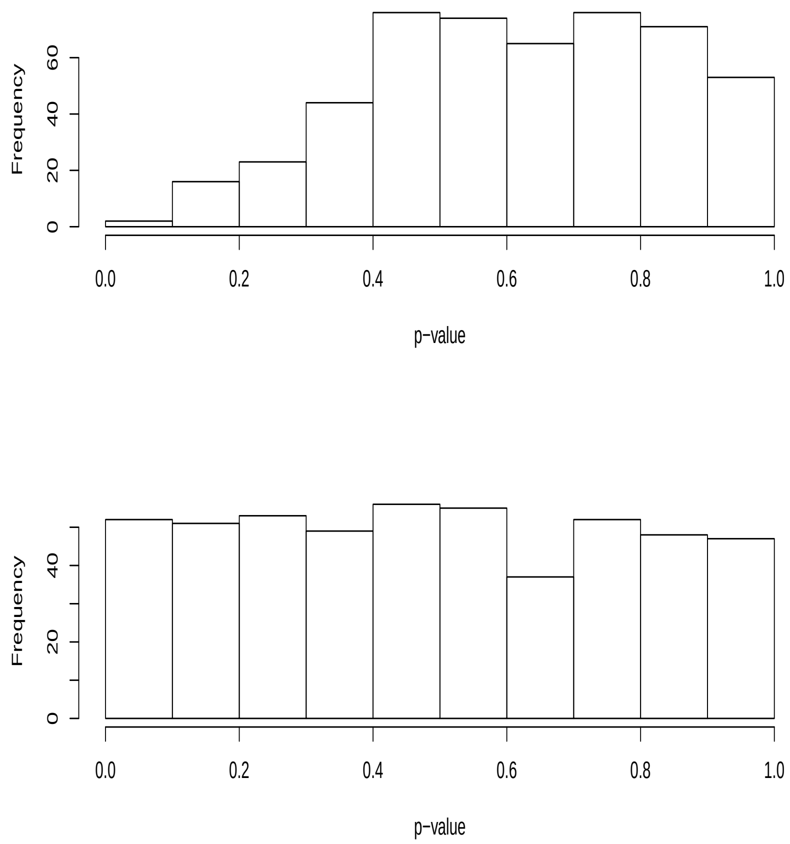

Figure 1: Histogram bootstrap of the p-value of IMT and $IMTR$ respectively, under true model given by $\alpha = 0.2, \beta_1 = 0.3, \beta_2 = 0.5, \beta_3 = 1.4$, sample size and bootstrap sampleis $n = B = 500$ and N=1000 number of simulation.

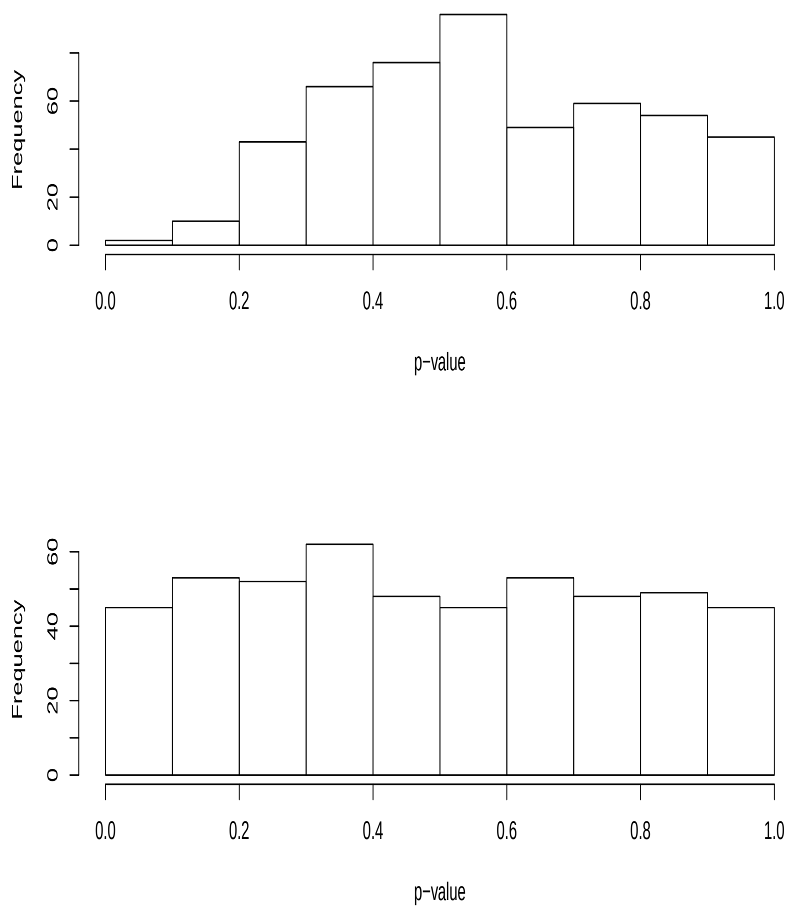

Figure 2: Histogram bootstrap of the p-value of IMT and $IMTR$ respectively, under true model given by $\alpha = 0.9, \beta_1 = 1.3, \beta_2 = 1.1, \beta_3 = 1.5$, sample size and bootstrap sampleis $n = B = 500$ and N=1000 number of simulation.

Bibliography

Aldrich, J. A. and Nelson, F. D. (1984). *Linear Probability, Logit, And Probit Models*. Sage Publications, Inc.,USA.

Arnold, B., Beaver, R. J., Groeneveld, R. A., and Meeker, W. Q. (1993). The non-truncated marginal of a truncated bivariate normal distribution. *Psychometrika*, 58:471–478.

Arnold, B. C. and Beaver, R. J. (2000). Hidden truncation models. *Sankhya* , 62:23–35.

Arnold, B. C. and Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Sociedad de Estadistica e Investigacion Oprativa*, 11:7–54.

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scandinavian Journal of Statistics*, 12:171–178.

Azzalini, A. (1986). Further results on a class of distribution which includes the normal ones. *Statistica*, 46:199–208.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Journal of Statistics*, 32:159–188.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, series B*, 61:579–602.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4):713–726.

Brown, C. C. (1980). On a goodness of fit test for the logistic model based on score statistics. *Communications in Statistics Theory and Methods*, 10:1097–1105.

Cartinhour, J. (1990). One-dimensional marginal density functions of a truncated multivariate normal density function. *Commun. Statist,-Theory Meth.*, 19(1):197–203.

Chesher, A. (1984). Testing for neglected heterogeneity. *Econometrica*, 52:865–872.

Chiogna, M. (1998). Some results on the scalar skew-normal distribution. *Journal of the Italian Statistical Society*, 7:1–13.

Claeskens, G. and Hjort, N. L. (2008). *Model Selection and Model Averaging*. Cambridge University Press.

Copas, J. B. (1989). Unweighted sum of squares test for proportions. *Journal of the Royal Statistical Society, series C*, 38:71–80.

Cox, D. R. (1970). *Analysis of Binary data*. Chapman and Hall,London.

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary data*. Chapman and Hall,New York, 2nd edition.

Davidson, R. and Mackinnon, J. G. (1984). Convenient specification tests for logit and probit models. *Journal of Econometrics*, 25(2):241–262.

Davison, A. and Hinkley, D. (1997). *Bootstrap Methods and their Application*. Cambridge:Cambridge University Press.

Dickson, E. R., Fleming, T. R., Wiesner, R. H., Baldus, W. P., Fleming, C. R., Ludwig, J., and McCall, J. T. (1985). Trial of penicillamine in advanced primary billiary cirrhosis. *New England Journal of Medicine*, 312:1011–1015.

Dobson, A. (1990). *An Introduction to Generalized Linear Models.* Chapman and Hall, London.

Dobson, A. J. and Barnett, A. G. (2008). *An Introduction To Generalized Linear Models.* Chapman and Hall, New York, 3rd edn edition.

Drake, C. and McQuarrie, A. (1995). A note on the bias due to omitted confounders. *Biometrika*, 82:633–638.

Draper, N. R. and Smith, H. (1996). *Applied Regression Analysis*. Wiley-New York, 2rd edition.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstra*. Boca Ration:Chapman and Hall.

Fleming, T. R. and Harrington, D. P. (2005). *Counting Processes and Survival Analysis.* Wiley: Chichester, 2nd edition.

Gail, M. H., Wieand, S., and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regression and omitted covariates. *Biometrika*, 71:431–444.

Gumbel, E. J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association*, 56:335–349.

Hausman, J. A. (1978). Specification tests in econometrics. *The Econometrix Society*, 46(6):1251–1271.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian Journal of Statistics*, 13:271–275.

Hilbe, J. M. (2009). *Logistic Regression Model.* Chapman and Hall, New York.

Hill, M. A. and Dixon, W. J. (1982). Robustness in real life. *Biometrics*, 38(2):377–396.

Hosmer, D. and Lemeshow, S. (1989). *Applied Logistic Regression.* Wily, New York.

Hosmer, D. and Lemeshow, S. (2000). *Applied Logistic Regression.* Wily, Chichester, 2nd edition.

Hosmer, D., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression.* Wily, Chichester, 3rd edition.

Hosmer, D. W., Hosmer, T., Le. Cessie, S., and Lemeshow, S. (1997). A comparision of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, 16:965–980.

Hosmer, D. W., Hosmer, T., and Lemeshow, S. (1980). A goodness-of-fit tests for the multiple logistic regression model. *Communications in Statistics*, 10:1043–1069.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions.* A Wiley-Interscience Publication, U.S.A.

Kleinbaum, D. G. (1994). *Logistic Regression A Self-Learning Text.* Springer-Verlag, New York.

Kuss, O. (2002). Global goodness-of-fit tests in logistic regression with sparse data. *Statistics in Medicine*, 21:3789–3801.

Lancaster, T. (1984). Covariance matrix of the information matrix test. *Econometrica*, 4:1051–1053.

Lee, L. F. (1982). Specification error in multinomial logit model. *Jornal of Econometrics*, 20:197–209.

Lemeshow, S. and Hosmer, D. W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology*, 115:92–106.

Lin, D. Y. and Wel, L. J. (1991). Goodness-of-fit tests for the general cox regression model. *Statist. Sinica*, 1:1–17.

Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statistics and Probability Letters*, 54:381–387.

Ma, Y. and Genton, M. G. (2004). Flexible class of skew-symmetric distribution. *Scandinavian Journal of Statistics*, 31:459–468.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press, New York.

McCullagh, P. (1986). The conditional distribution of goodness-of-fit statistics for discrete data. *Journal of the American Statistical Association*, 81:104–107.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman and Hall, London, 2nd edition.

Nagelkerke, N. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 3:691–692.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, series A*, 135(3):370–384.

Neuhaus, J., Kalbfleisch, J. D., and Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, 59:25–35.

Neuhaus, J. M. and Jewell, N. P. (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika*, 80(4):807–815.

Newey, W. K. (1984). Maximum likelihood specification testing and conditional moment test. *Econometrica*, 53(5):1047–1070.

Orme, C. (1988). The calculation of the information matrix test for binary data models. *EconPapers*, 56(4):370–376.

Robinson, D. and Jewell, N. P. (1991). Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review*, 59(2):227–240.

Weisberg, S. (2005). *Applied Linear Regression*. Wiley-Interscience, 3rd edition.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25.