Imprecise Probability Models for Logistic Regression

A Thesis
Submitted to the College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of
Doctor of Philosophy
in the
Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon

By
Osama Mohammad Bataineh

©Osama Mohammad Bataineh, September/2012. All rights reserved.
PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mathematics and Statistics
University of Saskatchewan,
Saskatoon, Saskatchewan, Canada
S7N 5E6
Abstract

Imprecise probability models are applied to logistic regression to produce interval estimates of regression parameters. The lengths of interval estimates are of main interest. Shorter interval estimates correspond to less imprecision in regression parameters estimates.

This thesis applies imprecise probabilistic methods to the logit model. Imprecise logistic regression, briefly called ImpLogit model, is presented and established for the first time. ImpLogit model is applied based on an inferential paradigm that applies Bayes theorem to a family of prior distributions, yielding interval posterior probabilities. The so-called interval estimates of regression parameters are computed using Metropolis-Hastings algorithm.

Two imprecise prior probability models are applied to 2-parameter ImpLogit model: the imprecise Dirichlet model (IDM) and the imprecise logit-normal model (ILnM). The 2-parameter ImpLogit model is fitted using real life dose-response data. This takes into account the cases of increasing, decreasing and mixed-belief ImpLogit models.

The relation between the lengths of interval estimates of regression parameters and both of covariate values and imprecise prior hyperparameters, in 2-parameter ImpLogit model, is studied by simulation. Different designs are applied in order to investigate a way to shorten the lengths of interval estimates of regression parameters. Having covariate fixed values to surround the prior believed median value of the logistic distribution results in reducing the imprecision in interval estimates. Fixing covariate values around the prior believed median value in a short range increases the lengths of interval estimates.

The number of fixed covariate values (say number of distinct dose levels in a dose-response experiment) affects the produced imprecision. A larger number of fixed covariate values increases the lengths of interval estimates. Therefore, a good design has a small number of fixed covariate values, located and spread out not in a short range.

ImpLogit model designs that are recommended by the simulation study, are compared to optimal designs in the frequentist approach using Fisher information matrix (FIM). Designs in FIM agree with designs that reduce imprecision in 2-parameter ImpLogit model, in the necessity of having covariate values to be fixed around the prior believed
median value of the logistic distribution, not in a short range.

**Keywords**: Imprecise probability model; imprecise Dirichlet model; imprecise logit-normal model; aggregation property; ImpLogit model; interval estimate.
Acknowledgements

First, I would like to thank my supervisor, Dr. Mikelis Bickis for his encouragement, advice, enduring patience throughout the years of my study. His guidance has been vital and he has never been ceasing in his support in every possible way towards making me able, to accomplish my thesis. He has always provided clear explanations when I was stuck, constantly driving the research with new and innovative research methodologies.

I want to sincerely acknowledge the financial support provided by the College of Graduate Studies, Department of Mathematics and Statistics and Dr. Mikelis Bickis.

It is my great pleasure to thank all my friends and well–wishers who encouraged and helped me in the PhD program.

Finally, I am very much grateful to my parents, my sister and my younger brothers, for their encouragement during my stay far away from them.
To my parents whom I love the most ...
# Contents

<table>
<thead>
<tr>
<th>Permission to Use</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Contents</td>
<td>vi</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Logit Model</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Bayesian Approach for the Logit Model</td>
<td>8</td>
</tr>
<tr>
<td>1.4 Bayesian Models Robustness</td>
<td>9</td>
</tr>
<tr>
<td>1.5 Historical Development of Bayesian Methods</td>
<td>11</td>
</tr>
<tr>
<td>1.5.1 Principle of Inverse Probability</td>
<td>12</td>
</tr>
<tr>
<td>1.5.2 Appearance of the Frequentist Approach</td>
<td>13</td>
</tr>
<tr>
<td>1.5.3 Establishment of the Subjective Bayesian Approach</td>
<td>13</td>
</tr>
<tr>
<td>1.6 Imprecise Probabilities</td>
<td>15</td>
</tr>
<tr>
<td>1.7 Imprecise Dirichlet Model</td>
<td>19</td>
</tr>
<tr>
<td>1.8 Applications of Imprecise Probability Models</td>
<td>25</td>
</tr>
<tr>
<td>1.9 Organization of the Thesis</td>
<td>27</td>
</tr>
</tbody>
</table>

| **2 Bayesian 2-Parameter Logit Model**      | 28 |
| 2.1 Increasing 2-Parameter Logit Model      | 28 |
| 2.2 Decreasing 2-Parameters Logit Model     | 38 |
| 2.3 Logit–Normal Model                      | 43 |
| 2.4 Multivariate Logit-Normal Model         | 45 |

| **3 ImpLogit Model**                        | 50 |
| 3.1 Imprecision in 2-Parameter Logit Model  | 51 |
| 3.2 Transmission from $\theta$ to $\beta$ in 2-Parameter ImpLogit Model | 61 |
| 3.2.1 Positive $x_j$ and $x_k$              | 65 |
| 3.2.2 Negative $x_j$ and $x_k$             | 67 |
| 3.2.3 Negative $x_j$ and Positive $x_k$    | 68 |
| 3.3 2-Parameter Increasing ImpLogit Model   | 73 |
| 3.4 2-Parameter Decreasing ImpLogit Model   | 78 |
| 3.5 Mixture of Beliefs in 2-Parameter ImpLogit Model | 82 |
| 3.6 ImpLogit with Imprecise Bivariate Logit-Normal Model | 84 |
| 3.7 ImpLogit with Smaller Sample            | 87 |
## Simulation Study

4.1 Motivation for the Simulation Study ........................................... 89
4.2 Symmetric Two Values ................................................................. 91
4.3 Right and Left-Shifted Two Values ............................................. 94
4.4 Extremely Shifted Two Values ...................................................... 95
4.5 Symmetric Covariate Multiple Values .......................................... 97
4.6 Right and Left-Shifted Multiple Values ....................................... 97
4.7 Extremely Shifted Multiple Values .............................................. 98
4.8 Closer $\theta_j$ and $\theta_k$ ............................................................. 98
4.9 Optimal Frequentist Design .......................................................... 99
4.10 Multiple Parameter ImpLogit Model ............................................ 102

## Conclusion and Future Plan

5.1 Conclusion ............................................................................. 130
5.2 Future Plan .......................................................................... 133

References .............................................................................. 134

A R Code ................................................................................. 143

B Posterior logit-normal density .................................................. 150

C Posterior multivariate logit-normal density ................................. 155

D Calculations and graphs from Chapter 3 .................................... 157

E Graphs from Chapter 4 ............................................................... 163

F Fisher Information Matrix for Logistic Regression ...................... 171
1.1 Motivation

Logistic regression is a well known statistical model that has been applied widely in statistical analysis. It is used in applied statistics to analyze binary and multinomial data. In the logistic regression model (also called logit model), observed binary data are assumed to come from a Bernoulli distribution where the sum of observed binary data follows the binomial distribution. The logit model makes predictions about the probability of occurrence of an event. Finney (1978) used the logit model for applications in statistical bioassay in order to build dose-response relationships. Collet (1991) gives a detailed study of the logit model with applications to real life binary data. Fundamentals of analysis of binary data are given by Cox (1970) and Cox and Snell (1989). Applying the logistic regression model is discussed in Agresti (2007). A recent study for logistic regression model is given in Hilbe (2009).

The logit model assumes a random variable, say \( Y_i, \ i = 1, \ldots, m \), to follow the binomial distribution, \( \text{bin} (n_i, \theta_i) \), where \( \theta_i \) denotes the \( i \)th probability of success, and \( n_i \) is the number of trials in the \( i \)th observation. The main parameters of interest, \( \theta_i \) for all \( i \), are modelled as

\[
\theta_i = \theta (x_i) = \frac{e^{x_i \beta}}{1 + e^{x_i \beta}} = \frac{e^{\beta_0 + x_i \beta_1 + \ldots + x_{ip} \beta_p}}{1 + e^{\beta_0 + x_i \beta_1 + \ldots + x_{ip} \beta_p}};
\]

\[-\infty < \beta_j < \infty, \quad j = 0, \ldots, p,\]

\[-\infty < x_{ij} < \infty, \quad i = 1, \ldots, m, \quad j = 1, \ldots, p,\]

(1.1)

where \( \beta_j, j = 0, \ldots, p, \) are unknown regression parameters and \( x_i = [1, x_{i1}, \ldots, x_{ip}]' \) is the \( i \)th vector of fixed covariates.
The estimation of the logit model parameters results in having a statistical relation between $\theta_i$ and the corresponding $i$th vector of fixed covariates. This comes by modelling the binomial distribution parameters ($\theta_i$'s) as a cumulative density function (cdf) of the logistic distribution. The logistic distribution has the following probability density function (pdf),

$$f(x) = \frac{e^x}{(1 + e^x)^2} = \frac{e^{-x}}{(1 + e^{-x})^2}, \quad -\infty < x < \infty,$$

and a cdf as

$$F(t) = \int_{-\infty}^{t} \frac{e^x}{(1 + e^x)^2} dx = \int_{-\infty}^{t} \frac{e^{-x}}{(1 + e^{-x})^2} dx = \frac{e^t}{1 + e^t} = \frac{1}{1 + e^{-t}},$$

$$-\infty < t < \infty.$$  

(1.2)

(1.3)

The logistic cdf given in (1.3) matches with (1.1) where $t$ is expressed as $t = x'\beta$. Figure 1.1 presents logistic cdf’s for different values of regression parameters $\beta_0$ and $\beta_1$. Figure 1.1 shows four plots where the top ones are for increasing logistic functions with $\beta_1 > 0$, but the lower ones are for decreasing logistic functions with $\beta_1 < 0$. The top-left plot fixes $\beta_0$ and changes $\beta_1$ which impacts the sharpness of logistic curve S-shape. The top-right plot fixes $\beta_1$ and changes $\beta_0$ which results in shifting the logistic curve either to the left or right. A similar thing shows in bottom plots for decreasing logistic functions.

The application of the logit model generally aims to study the effects of the fixed covariates on the predicted probabilities of occurrence. Therefore, there is a focus on estimates and confidence intervals of the regression parameters.

This thesis will investigate applying the imprecise probabilistic methods to the logit model under the Bayesian approach. This requires introducing imprecise probabilities and their models as a main goal of this chapter. Understanding probabilistic imprecision comes after introducing the Bayesian robust and hierarchical methods with a general look at historical developments of Bayesian methods.
Figure 1.1: Plots of different logistic cdf’s.
This introductory chapter is organized as follows. Section 1.2 starts by introducing some technicalities of the logit model. Mathematical aspects are given to describe the relation between the logit model and the exponential families. Also, an example of a real life application of the logit model is given in order to show its importance to the world of applied statistics.

Section 1.3 explains how the Bayesian approach is applied to the logit model. This requires introducing prior and posterior distributions of the main regression parameters in the logit model.

Bayesian robustness in the logit model is discussed in Section 1.4. Understanding robust methods helps to visualize and handle the concept of probabilistic imprecision with less difficulty. Furthermore, more details are given about the history of using and preferring robust methods in the Bayesian approach. A mathematical description of robust methods is given for building interval estimates for parameters of interest.

Section 1.5 takes the reader into a quick historical trip of Bayesian methods developments and applications. Basic and foundational studies in Bayesian methods will be cited. The main goal is to know circumstances that led probabilists, statisticians and researchers to adopt the imprecise probabilistic approach.

Section 1.6 goes deeper and extends sections 1.4 and 1.5 by presenting imprecise probabilities. Section 1.6 will prepare to jump smoothly to Section 1.7 in which the imprecise Dirichlet model is introduced.

In Section 1.7, two imprecise probability models are given: Imprecise Beta Model (IBM) as given in Walley (1991) and its generalized form, the Imprecise Dirichlet Model (IDM), as given in Walley (1996) and Bernard (2005). The study of IBM and IDM is necessary before going into two more imprecise probability models in Chapter 3, the imprecise logit–normal model (ILnM) and its generalized form, the imprecise multivariate logit–normal Model (IMLnM), where both of them are given in Bickis (2009). Aspects of probabilistic imprecision will be shown in IBM and IDM provided that Walley (1991), Walley (1996), Bernard (2005) and Bickis (2009) are considered as driving references and main sources of knowledge.

Section 1.8 gives examples of applications of imprecise probability models to show usefulness of thinking in probabilistic imprecision as an approach in real life problems.
Applications include several fields as classification, clinical trials and regression.

Finally, Section 1.9 looks at the general design of this thesis. A brief description of
next chapters is given to prepare to move from theoretical forms of imprecise probability
models to their new application in the logit model.

1.2 Logit Model

The logit model is a generalized linear model (GLM). In a GLM, a random variable \( Y \)
follows a probability distribution that belongs to a one-dimensional exponential family.
The distribution of a random variable is said to belong to a one-dimensional exponential
family of distributions if its probability density function takes the form of

\[
f(y|\eta, \phi) = \exp \left\{ \frac{g(y)\eta - b(\eta)}{a(\phi)} + c[g(y), \phi] \right\},
\]

(1.4)

where \( \eta \) is called the canonical parameter, \( a \) is an arbitrary function, \( b \) is differentiable, \( c \) is
a function that lets \( f(.) \) to integrate to 1 and \( \phi \), being known, is the dispersion parameter.
The binomial family is a well known exponential family of distributions. This is shown
as follows

\[
f(y|\eta, \phi) \\
= \binom{n}{y} \theta^y (1-\theta)^{n-y} \\
= \exp \left\{ \log \left[ \binom{n}{y} \theta^y (1-\theta)^{n-y} \right] \right\} \\
= \exp \left[ y \log \theta + (n - y) \log(1 - \theta) + \log \binom{n}{y} \right] \\
= \exp \left[ y \log \left( \frac{\theta}{1-\theta} \right) + n \log(1 - \theta) + \log \binom{n}{y} \right],
\]

(1.5)
where $f(.)$ is the density function (with respect to a counting measure) and

$$
\eta = \log \left( \frac{\theta}{1 - \theta} \right)
$$

$$
\implies e^\eta = \frac{\theta}{1 - \theta}
$$

$$
\implies \theta = \frac{e^\eta}{1 + e^\eta}
$$

$$
\implies 1 - \theta = \frac{1}{1 + e^\eta}.
$$

Then

$$
b(\eta) = -n \log (1 - \theta) = n \log [1 + \exp(\eta)].
$$

(1.7)

Also,

$$
c(y, \phi) = \log \left[ \frac{n}{y} \right], \quad a(\phi) = 1 \quad \text{and} \quad g(y) = y.
$$

(1.8)

If $g(y) = y$ in (1.4) which is the case in (1.5), then the random variable $Y$ has the following mean and variance

$$
E(Y) = \mu = b'(\eta),
$$

(1.9)

$$
V(Y) = b''(\eta) a(\phi).
$$

(1.10)


If the logit model has two regression parameters, $\beta_0$ and $\beta_1$, then the $i$th probability of occurrence $\theta_i$ is modelled as

$$
\theta_i = \theta(x_i) = \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} = \frac{1}{1 + e^{-(\beta_0 + x_i \beta_1)}}, \quad 0 < \theta_i < 1, \quad i = 1, ..., m,
$$

(1.11)

which is re-written as

$$
\log \left( \frac{\theta_i}{1 - \theta_i} \right) = \beta_0 + x_i \beta_1.
$$

(1.12)

The left hand side of (1.12) is referred to as the logit function. It can be seen that $\theta_i$ in (1.11) has the following limits,

$$
\lim_{x_i \to -\infty} \theta(x_i) = 0 \quad \text{and} \quad \lim_{x_i \to \infty} \theta(x_i) = 1, \quad \beta_1 > 0,
$$

(1.13)
and

\[ \lim_{x_i \to -\infty} \theta(x_i) = 1 \quad \text{and} \quad \lim_{x_i \to \infty} \theta(x_i) = 0, \quad \beta_1 < 0. \tag{1.14} \]

Limits in (1.13) are identical to the limits of the logistic cdf given in (1.3) when \( \beta_0 = 0 \) and \( \beta_1 = 1 \).

The logit model describes various real life phenomena. For example, Berkson (1944) made an application of logit model in biological assay to find statistical relations between the amount of dose and response (deaths). This relation is called in bioassay as the dose-response relationship. It can be built by fitting a logit model, say with two parameters \( \beta_0 \) and \( \beta_1 \), such that the predicted responses correspond to given doses. Having a dose-response relation gives an opportunity to make comparisons among drugs in terms of potency. In bioassay experiments, different doses are given to groups of experimental individuals or animals. The number of responses in each group to a given dose is considered as an observation on a binomial random variable. In this type of experiment, a parameter \( \theta \) defines the tolerance of the animal on a specific dose \( x \) with no response.

In bioassay, there is often an interest in the dose amount that produces a response in 50% of the experimented animals. This dose amount is called the median effective dose and is referred to as the \( ED_{50} \) value. If the response proportion is 90%, then the effective dose is the \( ED_{90} \) value, and so on. Subject to (1.12), the dose for which \( \theta = 0.5 \) is given as

\[ \log \left( \frac{0.5}{1 - 0.5} \right) = 0 = \beta_0 + \beta_1 (ED_{50}), \tag{1.15} \]

which means that

\[ ED_{50} = -\frac{\beta_0}{\beta_1}. \tag{1.16} \]

Once the logit model is fitted, the parameters in (1.16) are replaced by the estimates \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \).

Estimation of regression parameters requires following a statistical approach. This includes the maximum likelihood and the Bayesian approaches. The Bayesian approach will be presented in Section 1.3.
1.3 Bayesian Approach for the Logit Model

The Bayesian approach states that if \( Y \) is a random variable with a probability density function \( f(y|\theta) \), and \( \theta \) is a distribution parameter that follows a prior distribution \( \pi(\theta) \), then the posterior distribution \( \pi(\theta|y) \) according to Bayes rule is given by

\[
\pi(\theta|y) = \frac{(\text{likelihood})(\text{prior})}{m(y)} = \frac{f(y|\theta) \pi(\theta)}{m(y)},
\]

where \( m(y) \) is the marginal distribution of \( y \) defined as

\[
m(y) = \int_{\theta} f(y|\theta) \pi(\theta) \, d\theta = \int_{\theta} f(y, \theta) \, d\theta.
\]

The Bayesian approach allows representation of pre-experimental beliefs on \( \theta \). This requires to assign a prior distribution in order to produce the posterior distribution \( \pi(\theta|y) \) that is conditioned on observed data. More details about Bayesian methods are in Bernardo and Smith (1994) and Robert (2001) and information about the history of Bayesian methods is given in Stigler (1983, 1986).

In Bayesian analysis of the logit model, a prior distribution is assigned to regression parameters \( \beta = [\beta_0, ..., \beta_p]' \). The regression parameter vector is considered as a random one so that it follows a joint probability distribution called the joint prior distribution, denoted by

\[
\beta \sim \pi(\beta).
\]

The posterior distribution of \( \beta \) given observed values \( y_1, ..., y_m \), assuming independence among \( y_1, ..., y_m \), is determined as

\[
\pi(\beta|y) = \pi(\beta|y_1, ..., y_m) = \frac{f(y, \beta) \pi(\beta)}{m(y)} = \frac{f(y|\beta) \pi(\beta)}{m(y_1, ..., y_m)}
\]

\[
= \frac{\prod_{i=1}^{m} f(y_i|\beta) \cdot \pi(\beta)}{m(y_1, ..., y_m)} = \frac{\prod_{i=1}^{m} f(y_i|\beta) \cdot \pi(\beta)}{\int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{m} f(y_i|\beta) \right\} \pi(\beta) \, d(\beta)},
\]

where \( f(y_i|\beta) \) is the binomial distribution pdf, such that \( \theta_i \) is modelled as in (1.1), and \( m(.) \) is the marginal joint pdf of \( y \). The posterior multivariate pdf is usually intractable,
so computational methods are implemented to infer $\pi (\beta | y)$. To estimate $\beta$, the posterior distribution is simulated, then $\beta$ is averaged in (1.20).

Having a single prior $\pi$ is not always the case, but a set of priors can be assumed. That is, a class of priors is given to the parameters of interest such that $\pi \in \mathcal{A}$, where $\mathcal{A}$ refers to a set of prior distributions. This allows one to study robustness of Bayesian models to changes in the prior knowledge of statistical models parameters. Section 1.4 sheds light on Bayesian models robustness and reasons behind considering it.

### 1.4 Bayesian Models Robustness

Let $\mathcal{A}_{\text{prior}}$ be a set of priors on a single parameter $\theta$, then there is a corresponding set of posteriors $\mathcal{A}_{\text{posterior}}$. If $\theta$, in each single prior, is estimated by finding $E (\theta | y)$, analytically or computationally, then a set of Bayesian estimates is produced for $\theta$. The robust Bayesian approach focuses on the range of the posterior means

$$\left[ \inf_{\pi \in \mathcal{A}_{\text{prior}}} E (\theta | y), \sup_{\pi \in \mathcal{A}_{\text{prior}}} E (\theta | y) \right]. \tag{1.21}$$

The goal is to find the range of posterior estimates and conditions for shortening the interval estimate in (1.21). Gustafson (1996a) mentions that robust analysis of (1.21) is referred to as formal sensitivity analysis.

Regression parameters $\beta$ in the logit model can have a set of prior probability distributions rather than being restricted to follow a precise one. Once the regression parameters are estimated from the produced posterior distributions, then interval estimates for all regression parameters $\beta$ are made. This enables to study the effects of having prior variation on the interval estimate length for each regression parameter $\beta_j$, $j = 0, \ldots, p$.

The interval estimate for each $\beta_j$, $j = 0, \ldots, p$ is denoted by $\hat{\beta}_j \in [\underline{\beta}_j, \overline{\beta}_j]$ where $\underline{\beta}_j$ refers to the lower estimate and $\overline{\beta}_j$ is the upper estimate. In this case, there is

$$\left[ \inf_{\pi \in \mathcal{A}_{\text{prior}}} E (\beta | y_1, \ldots, y_m), \sup_{\pi \in \mathcal{A}_{\text{prior}}} E (\beta | y_1, \ldots, y_m) \right], \tag{1.22}$$

and

$$\hat{\theta} = \inf_{\pi \in \mathcal{A}_{\text{prior}}} E (\theta | y_1, \ldots, y_m), \quad \overline{\theta} = \sup_{\pi \in \mathcal{A}_{\text{prior}}} E (\theta | y_1, \ldots, y_m). \tag{1.23}$$
The idea of prior changes have caught attention in past through various applications. Greenhouse and Wasserman (1995) computed ranges of posterior expectations under priors in $\epsilon$-contaminated class of priors. The $\epsilon$-contaminated class of priors contains all mixture densities of the form

$$
(1 - \epsilon) p(\theta) + \epsilon q(\theta),
$$

where $q$ can be any density. A useful source of knowledge to navigate methods and techniques in robust Bayesian modelling is Berger (1985).

Robustness can be applied in hierarchical models as in Gustafson (1996b). Hierarchical models are given in Lehmann and Casella (1998). The following simple example gives an idea about hierarchical Bayesian modelling. Let $X \sim N(\theta, c)$, where $c$ is a known constant, and let the prior distribution of $\theta$ be $N(\mu, \tau^2)$ given that $\mu$ is known. Then $\tau^2$ is assumed to follow a distribution, say $\tau^2 \sim \text{gamma}(\alpha, \beta)$, given that the hyperparameters $\alpha$ and $\beta$ are known.

Robust methods have applications in testing hypotheses. Having a set of priors results in finding a lower bound for posterior probability of a null hypothesis. Sets of priors are considered in hypothesis testing by various authors. Berger and Delampady (1987) and Berger and Sellke (1987) demonstrated that a set of priors may produce a lower bound of the posterior probability of a true null hypothesis that is larger than the corresponding $p$-value. Carlin and Sargent (1996) tried different prior assumptions in hypothesis testing for a parameter $\theta$ that is assumed to change over a determined interval.

The involvement of robust methods in clinical trials has reserved a place in applied statistics. Robust methods for clinical trials are used in Greenhouse and Wasserman (1995). Spiegelhalter, Freedman and Parmar (1994) suggested three different priors in clinical trials experiments for a certain parameter of interest. If the clinical trials experiments cause a concern of beliefs about priors in the proposed model, then the priors are considered robust if the posterior distributions make similar results.

Bayesian methods literature includes useful review studies of robust methods. Berger (1990, 1994) and Wasserman (1992) are good examples that treat the robustness of Bayesian estimates to prior distributions changes. Pericchi and Walley (1991) is one more advanced source of robustness methods but for Bayesian credible intervals. Walley (1991) is an excellent source of knowledge that discusses in depth the use of classes of priors.
Walley specifies classes of priors called imprecise probability models. Walley (1991) is considered as a very standard and rigorous work. Walley (1991) has to be cited in any work on imprecise probability models and assessments of Bayesian robustness. Walley’s foundational book is considered by Berger (1994) as: “this latter work is particularly to be recommended for its deep and scholarly study of the foundations of imprecision and robustness”.

It is important to look at the historical developments of Bayesian methods that caused statisticians to consider and apply robust methods. Therefore, Section 1.5 goes into a historical trip of developments of Bayesian methods to have a deeper understanding of the applications of Bayesian approach to logit model. Section 1.5 attempts to form a bridge between previous and next sections of this chapter.

1.5 Historical Development of Bayesian Methods

Presenting the history of Bayesian methods development acknowledges reasons and circumstances that led statisticians to think of interval values as parameters estimates. The statistical literature provides several sources of knowledge about the historical establishment of Bayesian methods. Powerful studies are in Stigler (1986), Fienberg (2006) and Aldrich (2008).

Fienberg (2006) made a strongly recommended and well written work. Fienberg’s professional work is the most considered one in this section. He presented rich and rigorous study of historical debates between Bayesian and non-Bayesian statisticians and probabilists.

This section is divided into three subsections. Subsection 1.5.1 starts by defining the principle of inverse probability in order to let the reader to see how mathematicians and probabilists in 19th century thought of probability conditional on observed data. Then Subsection 1.5.2 moves to the circumstances in which the frequentist approach appeared. Subsection 1.5.3 looks at research work in the 20th century regarding the establishment of the subjective Bayesian approach.
1.5.1 Principle of Inverse Probability

Early applications of Bayesian methods took into account the discrete uniform priors with Laplace’s principle of insufficient reason. The principle of insufficient reason implies that if $n$ possibilities are only distinguishable by their names, then each possibility can be given a probability (prior) equal to $\frac{1}{n}$. Examples that apply this principle is in assigning probabilities to coins, dice, and cards. If a coin is thrown, then the results are either a “tail” or “head” and the probabilities of both is equal to $\frac{1}{2}$. Probabilistic equity applies to a symmetric die in which the probability of having any of the die faces after being thrown is $\frac{1}{6}$.

Fienberg (2006) mentions that the principle of insufficient reason was basically established by Laplace (1774) and took its general form in Laplace (1812). Laplace used the principle of insufficient reason to formulate the use of the uniform prior for the binomial distribution parameter. De Morgan (1837) renamed this principle as “principle of inverse probability”. De Morgan included the word “inverse” in the new name to infer backwards from data to probability.

The principle of inverse probability turned out to be a standard method of choice for the 19th century scientists. For example, Edgeworth (1883) made an astonishing derivation of Student’s $t$-distribution by establishing the posterior distribution of the mean $\mu$ of the normal distribution. Some details about the technicalities of Edgeworth contribution are given in Stigler (1978). Pearson (1907) discussed the need and importance of past experience to future expectation.

The principle of inverse probability was renamed by Keynes (1921) as the “principle of indifference”. This is because Keynes thought that the principle of inverse probability only applies to experiments with equal prior probabilities. More details about the history of using and applying the principle of inverse probability is in Dale (1999).

The principle of inverse probability was not always considered. Ronald Fisher, a British statistician, rejected it strongly which led to the birth of the frequentist approach. Subsection 1.5.2 moves to talk generally about the frequentist approach.
1.5.2 Appearance of the Frequentist Approach

Fisher (1922) established his own approach of inference that was called the likelihood approach. In fact, Fisher’s approach was revolutionary to the statistical thinking where he introduced the likelihood approach and its use to find the maximum likelihood estimates. He went farther by presenting the concepts of sufficiency and efficiency. Aldrich (2008) stated what Fisher (1925) mentioned about the principle of inverse probability as: “the theory of inverse probability is founded upon an error, and must be wholly rejected”. Aldrich mentions that Fisher thought the Bayesian approach formulates the problem and produces a solution and then withholds it. Fisher developed his new approach in Fisher (1925) where he made the formal principles of tests of significance. The important contributions made by Fisher in the third decade of the 20th century were followed by a foundational axiomatic work in probability by Kolmogorov (1933, 1950).

The appearance of the frequentist approach did not kill the principle of indifference. The main struggle between supporters of the principle of indifference and the frequentist approach started during the first half of the 20th century. Keynes (1921) attempted to treat the concept of probability in a new way by allowing for the possibility of having a personal degree of belief. Keynes developed a new meaning of probability itself. The introduction of the new probabilistic concept led Keynes to have a great influence on the traditional understanding of probability. Keynes prepared for the rise of what will be called later the “subjective probability”.

The allowance for degree of belief in probability was the seed to establish the subjective Bayesian approach. Subsection 1.5.3 gives more information about this approach and its growth.

1.5.3 Establishment of the Subjective Bayesian Approach

The study of developments and debates of Bayesian methods ends in touching the meaning of the subjective Bayesian approach. Goldstein (2006) is a recent study that pays attention to subjective approach as principle and practice. Machina and Schmeidler (1992) explored past contributions in subjective approach.

The subjective probability approach started to have its own popularity after Keynes
Ramsey (1926) thought that knowledge about probabilities is personal, therefore it is subjective.

De Finetti (1937) justified the concept of subjective probability by introducing the concept of exchangeability and the implicit role of prior distributions. Savage (1954) considered Keynes (1921) new meanings of probability as the earliest account of the modern concept of personal probability. This approves the strong influence of Keynes on Bayesian subjectivists.

The fifties of the 20th century was an important decade to subjective probability approach and subjectivists where more attention and publications arose to adopt, apply and promote Bayesian subjectivism. This was a normal result of the pioneering work of Savage (1954). Savage considered de Finetti (1937) notions to develop a non-frequentist approach. Savage foundational work established the birth of an advanced methodology of subjective methods. Savage (1962) completed his beliefs by introducing the subjective approach as a part of the foundations of statistical inference. Good (1976) thinks that Savage revived the understanding of the whole Bayesian methods.

The fast growth of the subjective Bayesian approach was going in parallel with other attempts of applying prior beliefs in statistical applications. For example, Fisher (1956) defined the Bayesian argument as a fiducial probability by prioritizing a single parameter \( \theta \) in one-dimensional random variables with a uniform distribution. Fisher fiducial argument was considered later in Lindley (1958), but Fisher could not convince all researchers of his time to adopt his method. This is because Fisher looked forward to build a posterior degree of belief without mentioning a prior one. Fisher was critiqued by Savage (1961) as "a bold attempt to make the Bayesian omelet without breaking the Bayesian eggs". Of course, Savage did not mean to underestimate Fisher himself, but he meant that the fiducial approach is used without a complete dependence and specification of a prior distribution.

The strength and richness of research work in subjectivism made the motivation to start developing and establishing the robust Bayesian methods. An early interest in robustness is in Box and Tiao (1961) with a focus on choosing suitable prior distributions. Edwards, Lindman and Savage (1963) applied robustness of statistical modelling in psychological research.
The establishment of subjectivism, being followed by starting studies of robust Bayesian methods, prepared to the birth of hierarchical Bayesian modelling. The allowance of the personal beliefs in the prior assumptions created a thriving interest of applying hierarchical Bayesian methods. Tiao and Tan (1965) presented a hierarchical normal model.

It is interesting to mention that Bayesian subjectivism was not the only approach to follow in Bayesian statistics. Jeffreys (1939) used the principle of indifference to update the degree of personal belief of probabilities and to derive what he called the objective priors.

In a Bayesian model, personal beliefs can be expressed by assigning a group of priors to parameters of interest. As mentioned in Section 1.4, if a class of priors is selected, then an interval estimate for a parameter of interest can be found. Therefore, assignment of priors class is itself a point of interest. How large should the class be so that the interval estimate is kept as short as possible? Imprecise probability theory can answer this important question by making use of the imprecise probability models. Section 1.6 presents basic concepts of imprecise probabilities depending on Walley (1991).

1.6 Imprecise Probabilities

In probability theory, an event $A$ has a precise probability between 0 and 1 to quantify the chance of occurrence. An imprecise probability is thought of in case of having a vagueness or confliction in information for assessing the precise probability. Therefore, generalizing precise probabilities is important in order to have a general assessment of the event $A$. This requires replacing the probability point value by an interval one.

An imprecise probability for an event $A$ is made by assigning lower and upper probabilities. The imprecise probabilities literature denotes the lower probability by $\underline{P}(A)$ and the upper one by $\bar{P}(A)$. A precise probability comes from having equal lower and upper probabilities in an imprecise probability. The idea of an interval probability is not new and goes back long time ago. According to Jane Hutton in the discussion of Walley (1996), Ostrogradsky (1838) was the first to make an explicit use of lower and upper probabilities in the context of judicial decisions.

To illustrate the use of imprecise probabilities, the following example is taken from
Walley (1991). Let $A$ denote an event that a particular thumbtack lands pin-up on its next toss. To construct lower and upper probabilities of $A$, several assessment strategies can be followed to use a relevant evidence about $A$. Evidences include records of previous tosses of this thumbtack or records from other thumbtacks. A simple strategy suggests to make intuitive assessments of $\underline{P}(A)$ and $\overline{P}(A)$. For example, if a thumbtack lands pin-up 3 times in 10 tosses, then an intuitive assessment makes low values for lower and upper probabilities of $A$, say $\underline{P}(A) = 0.1$ and $\overline{P}(A) = 0.4$. If the same thumbtack lands pin-up in 8 times, then lower and upper probabilities can be intuitively assessed by, say $\underline{P}(A) = 0.7$ and $\overline{P}(A) = 0.9$. Intuitive assessment is not a unique strategy. The lower and upper probabilities can be calculated based on observed values of a random variable which will be shown in Section 1.7.

Imprecise probabilities have their own applications in real world problems. Imprecise probabilities found their applications to the world of artificial intelligence early. Dempster (1967) used probabilistic imprecision to extend the belief functions. Belief functions were used in Shafer (1976). Theories in Shafer (1976) and Shafer and Vovk (2001) are used in De Cooman and Hermans (2008) in order to apply imprecise probabilities to trees with applications to artificial intelligence.

Other interesting applications are in applying imprecise probabilities to linguistics as in Zadeh (2002, 2006). This came after Zadeh (1975) where the concept of a linguistic variable with values as words or sentences was used in the artificial language to provide a basis for approximate reasoning. Imprecise probability theory can offer its services to the fields of finance and economics as in Vicig (2008). Imprecise probabilities are found to apply in engineering design by Aughenbaugh and Paredis (2006).

Imprecise probabilities stayed in shadows and depended on developments made by personal interests of mathematicians and statisticians until the nineties of the 20th century. The theory of imprecise probabilities gathered and received strong attention after Walley (1991). Walley presented a very rigid, foundational and strong theoretical description of imprecise probabilities. Walley went far in establishing the properties of imprecise probabilities and their uses in building the imprecise probability models. Walley refers to the interval probability by the term “imprecise probability”. However, the term “interval probability” was recently used in literature as in Weichselberger (2000).
Imprecise probabilities are meant to reflect the coherence of personal imprecise probabilistic assessments. Coherence is a self consistency requirement where an imprecise probability for an event is thought of as a result of thinking in how to handle a collection of probabilistic assessments.

To define lower and upper probabilities mathematically, let \( \mathcal{X} \) be a linear space of random variables on a sample space \( \Omega \), then a lower prevision, denoted by bolded \( \underline{P} \), is a real-valued function that maps \( \mathcal{X} \) to real numbers. That is

\[
\underline{P} : \mathcal{X} \rightarrow \mathbb{R}. \tag{1.25}
\]

A lower prevision \( \underline{P} \) is coherent when it is characterized by three axioms. The axioms are, for all \( X \in \mathcal{X}, Y \in \mathcal{X} \) and positive \( \lambda \),

\[
\begin{align*}
\underline{P}(X) & \geq \inf X, \\
\underline{P}(\lambda X) & = \lambda \underline{P}(X), \\
\underline{P}(X + Y) & \geq \underline{P}(X) + \underline{P}(Y),
\end{align*} \tag{1.26}
\]

where \( X \) is a real-valued function on \( \Omega \). Having \( \underline{P} \) and \( \mathcal{X} \) forms the lower prevision triple \( (\Omega, \mathcal{X}, \underline{P}) \).

The coherent upper prevision is a conjugate to the lower one. That is, with an upper prevision triple \( (\Omega, \mathcal{X}, \overline{P}) \), a coherent upper prevision, denoted by bolded \( \overline{P} \) and defined as

\[
\overline{P} : \mathcal{X} \rightarrow \mathbb{R}, \tag{1.27}
\]

is characterized by

\[
\begin{align*}
\overline{P}(X) & \leq \sup X, \\
\overline{P}(\lambda X) & = \lambda \overline{P}(X), \\
\overline{P}(X + Y) & \leq \overline{P}(X) + \overline{P}(Y).
\end{align*} \tag{1.28}
\]

Both of \( (\Omega, \mathcal{X}, \underline{P}) \) and \( (\Omega, \mathcal{X}, \overline{P}) \) assume \( \mathcal{X} \) to be a linear space. The space \( \mathcal{X} \) is linear if for all \( X \in \mathcal{X}, Y \in \mathcal{X} \) and \( \lambda \in \mathbb{R} \), there is

\[
\begin{align*}
\lambda X & \in \mathcal{X}, \\
Z = X + Y & \rightarrow Z \in \mathcal{X}. \tag{1.29}
\end{align*}
\]
Based on (1.26) and (1.28), there is
\[
\inf X \leq P(X) \leq \overline{P}(X) \leq \sup X, \ \forall X \in \mathcal{X}.
\] (1.30)

Walley goes deeper by defining lower and upper probabilities for events. Let \( \mathcal{A} \) be an arbitrary class of events, and consider it as a class of 0 - 1 (binary) random variables. Define the lower prevision \( P \) on \( \mathcal{A} \), then \( P \) is called the lower probability on \( \mathcal{A} \) and \( P(A) \) is called the lower probability on an event \( A \in \mathcal{A} \). Now, the upper probability comes by defining the upper prevision \( \overline{P} \) on \( \mathcal{A}^c \), where \( \mathcal{A}^c = \{ A^c : A \in \mathcal{A} \} = \{ 1 - A : A \in \mathcal{A} \} \). This means that
\[
\overline{P}(A) = 1 - P(A^c).
\] (1.31)

Imprecise probabilities have certain properties. A basic property is that an imprecise probability for an event \( A \) can have a minimum and maximum bounds as
\[
0 \leq P(A) \leq \overline{P}(A) \leq 1,
\] (1.32)
where the degree of imprecision is defined as
\[
\overline{P}(A) - P(A).
\] (1.33)
Also
\[
P(\emptyset) = \overline{P}(\emptyset) = 0, \ \text{and} \ P(\Omega) = \overline{P}(\Omega) = 1.
\] (1.34)
Another property to know is that
\[
\overline{P}\left( \bigcup_{j=1}^{n} A_j \right) \leq \sum_{j=1}^{n} \overline{P}(A_j).
\] (1.35)
If \( A_1, A_2, \ldots \) are disjoint events then
\[
P\left( \bigcup_{j=1}^{n} A_j \right) \geq \sum_{j=1}^{n} P(A_j),
\] (1.36)
and
\[
P\left( \bigcup_{j=1}^{\infty} A_j \right) \geq \sum_{j=1}^{\infty} P(A_j).
\] (1.37)

Imprecise probabilities have their own statistical models. Walley (1991, 1996), Bernard (2005) and Bickis (2009) presented imprecise probability models with specific names. Section 1.7 looks at widely used imprecise probability model called the Imprecise Dirichlet Model.
An advanced look is given in this section on designing imprecise probabilities in clearly defined and marked models. Attention is paid to imprecise probability models used for the binomial distribution. This comes first by introducing the imprecise beta model as a special case of the imprecise Dirichlet model. The beta distribution is used frequently as a prior for the binomial distribution parameter, the probability of occurrence $\theta$. If a class of beta priors is assigned to the binomial distribution parameter $\theta$, then the posterior distributions will be expressed in closed forms. This is attractively usual since the beta distribution is a conjugate prior to the binomial distribution, that is the posterior distribution of $\theta$ is also a beta distribution with hyperparameters different from those in the beta prior. Then the concept of probabilistic imprecision becomes less difficult to handle and understand.

According to Casella and Berger (2002), if $Y$ is a random variable such that $Y \sim \text{bin}(n, \theta)$ and $\theta \sim \text{beta}(\alpha, \beta)$, then the joint pdf of $Y$ and $\theta$ is

$$f(y, \theta) = f(y|\theta)\pi(\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (1-\theta)^{n-y+\beta-1},$$

(1.38)

where $y = 0, ..., n$, $0 < \theta < 1$, $\alpha > 0$ and $\beta > 0$. The marginal pdf of $Y$ is found by integrating (1.38) with respect to $\theta$,

$$m(y) = \int_0^1 f(y, \theta) \, d\theta = \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y + \alpha)\Gamma(n - y + \beta)}{\Gamma(n + \alpha + \beta)}.$$  

(1.39)

The probability density function in (1.39) is for a probability distribution known as the beta-binomial. Beta-binomial distribution is then used in Bayes rule to find the posterior distribution

$$\pi(\theta|y) = \frac{f(y, \theta)}{m(y)} = \frac{\Gamma(n + \alpha + \beta)}{\Gamma(y + \alpha)\Gamma(n - y + \beta)} \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1},$$

(1.40)

where the posterior pdf in (1.40) is for $\theta|y \sim \text{beta}(y + \alpha, n - y + \beta)$. Therefore, prior and posterior distributions of $\theta$ belong to the same family of beta distributions but with different parameters. Figure 1.2 shows a plot for both of beta prior and posterior pdf’s.
Figure 1.2: Plots of conjugate beta prior (solid) and posterior (dashed) pdf’s of
the binomial random variable parameter $\theta$ when $n = 10$, $y = 3$, $\alpha = 2$ and $\beta = 5$.

The Bayesian estimate (under the squared quadratic loss function) of the parameter
$\theta$ is the mean of (1.40), that is

$$\hat{\theta} = E(\theta|y) = \int_0^1 \theta f(\theta|y) d\theta = \frac{y + \alpha}{n + \alpha + \beta}. \quad (1.41)$$

This simple Bayesian review is necessary to prepare to build the probabilistic imprecision
skeleton in beta and Dirichlet distributions.

Now, a set of beta priors can be considered for the parameter $\theta$. A set of beta
distributions is particularly called the imprecise Beta model (IBM) by Walley (1991). Let
$\theta \sim \text{beta} (\alpha_1, \alpha_2)$ (instead of $\text{beta} (\alpha, \beta)$) with the following alternative parametrization

$$\theta \sim \text{beta} (\alpha_1, \alpha_2) = \text{beta} (\nu \varphi_1, \nu \varphi_2),$$

where $\nu = \alpha_1 + \alpha_2$, $\varphi_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2}$, $\varphi_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2}. \quad (1.42)$

Based on (1.40) and (1.42), the posterior distribution of $\theta$ is rewritten as

$$\theta|y \sim \text{beta} (y + \nu \varphi_1, n - y + \nu \varphi_2). \quad (1.43)$$

Note that for fixed $\nu > 0$, the set of priors is defined as

$$A_{\text{prior}} = \{\text{beta} (\nu \varphi_1, \nu \varphi_2) : (\varphi_1, \varphi_2) \in \varphi \}, \quad (1.44)$$
where $\varphi$ is the parameter space of all possible values of $\varphi_1$ and $\varphi_2$.

If IBM is considered, then there is a set of corresponding posteriors. The set of posteriors, given an observation $y$, is then defined as

$$A_{\text{posterior}} = \{ \text{beta} (y + \nu \varphi_1, n - y + \nu \varphi_2) : (\varphi_1, \varphi_2) \in \varphi \}.$$  \hfill (1.45)

Minimizing and maximizing $E(\theta)$ and $E(\theta|y)$ over the sets of prior and posterior distributions given in (1.44) and (1.45), with respect to $(\varphi_1, \varphi_2) \in \varphi$, gives lower and upper estimates in both sets of distributions. Walley (1991) refers to the lower and upper expectations by $\underline{E}(\theta)$ and $\overline{E}(\theta)$, $\underline{E}(\theta|y)$ and $\overline{E}(\theta|y)$.

Posterior lower and upper expectations come from estimating posterior distributions for which $\varphi_1 \to 0$ and $\varphi_1 \to 1$, respectively. If $\varphi_1 \to 0$ in (1.42) then $\varphi_2 \to 1$ and the posterior distribution in (1.43) becomes $\theta|y \sim \text{beta} (y, n - y + \nu)$, but if $\varphi_1 \to 1$ then $\varphi_2 \to 0$ and $\theta|y \sim \text{beta} (y + \nu, n - y)$. So, finding posterior expectations as in (1.41) for both of $\text{beta} (y, n - y + \nu)$ and $\text{beta} (y + \nu, n - y)$ gives the posterior lower and upper expectations of $\theta$. For any $\theta$ in IBM, the prior lower and upper expectations are

$$\begin{align*}
\text{as } \varphi_1 \to 0 & \text{ then } \underline{E}(\theta) = 0, \quad \text{(1.46)} \\
\text{and as } \varphi_1 \to 1 & \text{ then } \overline{E}(\theta) = 1, \quad \text{(1.47)}
\end{align*}$$

while the posterior lower and upper expectations are

$$\begin{align*}
\text{as } \varphi_1 \to 0 & \text{ then } \underline{\theta} = \underline{E}(\theta|y) = \frac{y}{n + \nu}, \quad \text{(1.48)} \\
\text{and as } \varphi_1 \to 1 & \text{ then } \overline{\theta} = \overline{E}(\theta|y) = \frac{y + \nu}{n + \nu}. \quad \text{(1.49)}
\end{align*}$$

The lower and upper estimates in (1.48) and (1.49) are combined to form an imprecise estimate of $\theta$ as $\hat{\theta} \in \left[ \underline{\theta}, \overline{\theta} \right]$. Figure 1.3 shows plots of lower and upper beta posterior cdf’s when $\varphi_1 \to 0$ and $\varphi_1 \to 1$ for different binomial random variable values.

The lower and upper estimates in (1.48) and (1.49) can be applied to the thumbtack example given in Section 1.6. Fix $\nu = 1$, then for a thumbtack that lands pin-up 3 times in 10 tosses, the lower and upper estimates in (1.48) and (1.49) are

$$\left[ \frac{3}{10 + 1} = \frac{3}{11}, \quad \frac{3 + 1}{10 + 1} = \frac{4}{11} \right]. \quad \text{(1.50)}$$
For a thumbtack that lands pin-up 8 times in 10 tosses, the lower and upper estimates are

\[
\begin{align*}
\hat{\theta} &= \frac{8}{11}, \\
\bar{\theta} &= \frac{9}{11}.
\end{align*}
\] (1.51)

It is important to distinguish the previous interval estimates from credible intervals in Bayesian methods. Recall that for a posterior distribution \( P(\theta|y) \), then a 95\% credible interval comes by finding \( c_1 \) and \( c_2 \) in

\[
\int_{c_1}^{c_2} P(\theta|y) \, d\theta = 0.95,
\] (1.52)

where the credible interval \([c_1, c_2]\) is assigned with a probability of 0.95. On the contrary, the interval estimates in (1.50) and (1.51) are not assigned with probabilities but they come from finding the point estimates in an infinite family of posterior distributions as in (1.45). For each point estimate in an interval estimate, a corresponding credible interval can be constructed.
Figure 1.3: Plots of lower and upper beta posterior cdf’s of $\theta$ when $\varphi_1 \to 0$ (solid) and $\varphi_1 \to 1$ (dashed) with $\nu = 2$, $n = 10$, $y = 1$ (top-left), $y = 3$ (top-right), $y = 7$ (bottom-left) and $y = 9$ (bottom-right).
The IBM is a special case of the Imprecise Dirichlet Model (IDM). The imprecise Dirichlet family of distributions forms a conjugate imprecise prior to the family of multinomial distributions. IDM was presented in Walley (1996) and discussed in Bernard (2005).

If \( Y \) follows the multinomial distribution as

\[
Y = (Y_1, ..., Y_m) \sim \text{mult} \left( N, \theta_1, ..., \theta_m \right),
\]

then \( Y \) has the following pdf

\[
f(Y|\theta) = \frac{N!}{\prod_{i=1}^{m} y_i!} \prod_{i=1}^{m} \theta_i^{y_i}, \tag{1.53}
\]

where \( N = \sum_{i=1}^{m} y_i, \sum_{i=1}^{m} \theta_i = 1, \ 0 < \theta_i < 1, \ \forall i. \)

If \( \theta \) follows the Dirichlet distribution as

\[
\theta = (\theta_1, ..., \theta_m) \sim \text{Dir} \left( \alpha_1, ..., \alpha_m \right), \tag{1.54}
\]

then \( \theta \) has the following pdf

\[
f(\theta|\alpha) = \frac{\Gamma \left( \sum_{i=1}^{m} \alpha_i \right)}{\prod_{i=1}^{m} \Gamma (\alpha_i)} \left( \prod_{i=1}^{m-1} \theta_i^{\alpha_i-1} \right) \left( 1 - \sum_{i=1}^{m} \theta_i \right)^{\alpha_{m-1}}, \tag{1.55}
\]

where \( \theta_m \) is written as \( \theta_m = 1 - \sum_{i=1}^{m-1} \theta_i, \ \alpha_i > 0, \ \forall i. \)

The posterior distribution of \( \theta|Y \) is \( \text{Dir} (y_1 + \alpha_1, ..., y_m + \alpha_m) \) with pdf

\[
f(\theta|\alpha) = \frac{\Gamma \left( N + \sum_{i=1}^{m} \alpha_i \right)}{\prod_{i=1}^{m} \Gamma (y_i + \alpha_i)} \left( \prod_{i=1}^{m-1} \theta_i^{y_i+\alpha_i-1} \right) \left( 1 - \sum_{i=1}^{m-1} \theta_i \right)^{y_m+\alpha_{m-1}}, \tag{1.56}
\]

where \( \theta_m \) is written as \( \theta_m = 1 - \sum_{i=1}^{m-1} \theta_i, \ \alpha_i > 0, \ \forall i. \)
Probabilistic imprecision in IDM is expressed in a similar way to IBM. Let
\[ \nu = \sum_{i=1}^{m} \alpha_i = \alpha_1 + \ldots + \alpha_m, \quad \varphi_i = \frac{\alpha_i}{\nu}, \quad \varphi_i \in [0, 1], \quad \forall i. \] (1.57)

For any \( \theta_i \) in IDM, the prior lower and upper expectations are
\[
\text{as } \varphi_i \to 0 \text{ then } E(\theta_i) = 0, \quad \text{and as } \varphi_i \to 1 \text{ then } E(\theta_i) = 1, \] (1.58) (1.59)

while the posterior lower and upper expectations are
\[
\text{as } \varphi_i \to 0 \text{ then } \hat{\theta}_i = E(\theta_i | y_1, \ldots, y_m) = \frac{y_i}{n + \sum_{i=1}^{m} \alpha_i} = \frac{y_i}{n + \nu}, \quad \text{and as } \varphi_i \to 1 \text{ then } \bar{\theta}_i = E(\theta_i | y_1, \ldots, y_m) = \frac{y_i + \sum_{i=1}^{m} \alpha_i}{n + \sum_{i=1}^{m} \alpha_i} = \frac{y_i + \nu}{n + \nu}. \] (1.60) (1.61)

Imprecise probability models have their own valuable applications. Section 1.8 gives examples of research fields in which imprecise probability models are found useful.

### 1.8 Applications of Imprecise Probability Models

Imprecise probability models have applications in several fields as clinical trials, classification and artificial intelligence. For example, Walley, Gurrin and Burton (1996) made a powerful study and application of the IBM to clinical trials analysis. IBM was applied for prior ignorance about binomial distribution parameter (probability of occurrence) to analyze clinical trials data. The imprecise probabilistic method was applied to data from clinical trials of extracorporeal membrane oxygenation (ECMO). The study mainly focused on the chance (probability) of survival for new-born babies who were treated by ECMO due to having acute respiratory failure or certain anomalies in cardiovascular circulation. An interesting application is in Coolen (1997) in which failure data are modelled imprecisely under IDM.
Imprecise probability models found their track to applications in statistical classification. Zaffalon (2002) made use of imprecise probabilities to create a new classifier called naive credal classifier (NCC). The imprecise classifier NCC extended a well known classifier called naive Bayes classifier (NBC). In NBC, if $C$ is a class variable that takes values $c_1, ..., c_n$, and $A_1, ..., A_n$ is a set of attributes variables where each attribute takes values $a_{i1}, ..., a_{ik}$, then an instance of attributes, say $a_1, ..., a_n$, is classified to a class $c_i \in C$ by maximizing

$$P(C|a_1, ..., a_n) = \frac{P(a_1, ..., a_n|C)P(C)}{P(a_1, ..., a_n)} \propto P(a_1, ..., a_n|C)P(C), \quad (1.62)$$

where independence is assumed among attributes, as in Duda and Hart (1973), conditional on the class variable

$$P(A_1, ..., A_n|C) = \prod_{i=1}^n P(A_i|C). \quad (1.63)$$

Then (1.62) becomes

$$P(C|a_1, ..., a_n) \propto P(C) \prod_{i=1}^n P(a_i|C). \quad (1.64)$$

The classifier NCC maps instances of attributes to a subset of classes in $C$ rather than only a single class. NCC is used when $P(a_i|C)$ is interval-valued, $\forall i$ in (1.64).

Imprecise probability models were used in regression analysis by Walter, Augustin and Peters (2007), in which the imprecise normal model was considered as an imprecise prior to the parameters of linear regression model. This application followed a remarkable work done by Quaeghebeur and De Cooman (2005). Quaeghebeur and De Cooman built a general theory for using imprecise probability models in exponential families. Quaeghebeur and De Cooman (2005) came after a study of imprecise conjugate priors for one parameter exponential family made by Coolen (1993).

An interesting application of imprecise probabilities is in Cozman (2000) where Bayesian networks are extended to credal networks. Credal networks are directed acyclic graphs that are associated with sets of probability measures. Cozman (2005) established the theory of graphical models for imprecise probabilities, where “graphical models” include credal networks and other types of graphs as the undirected graphs.
The last few years had excellent developments in the theoretical aspects of imprecise probabilities. De Cooman, Hermans and Quaeghebeur (2009) applied the imprecise transition probabilities to Markov chains. This may pace the track for a revolutionary field of applications of Markov chains and to establish what could be later called as the Imprecise Markov Chains. Imprecise probabilities have passed expectations to have applications in decision theory in Troffaes (2007) and in minimum distance estimation in Hable (2010).

According to Augustin and Hable (2010), imprecise probability theory can generalize the results of robust statistics. The theory of imprecise probabilities is promising enough to present imprecise hierarchical Bayesian models as in De Cooman (2002).

1.9 **Organization of the Thesis**

This thesis is organized as follows. Chapter 2 gives a new method of inducing a prior distribution for regression parameters in logistic regression model.

Chapter 3 dives into the main body of this thesis where a new statistical regression model called ImpLogit is established, developed and fitted. ImpLogit is a new coined model that makes use of IDM and ILnM as imprecise priors for logistic regression parameters. ImpLogit model is presented as a promising and novel model. ImpLogit model is fitted in purpose of finding conditions in which imprecision (interval estimates lengths) is reduced in regression parameters estimates.

A detailed simulation study is given in Chapter 4. The simulation study aims to explore statistical behaviour aspects in 2-parameter ImpLogit model. The relation between the 2-parameter ImpLogit model design and the amount of imprecision in regression parameters is a point of main interest. It is important to see how imprecision of estimates of $\beta_0$ and $\beta_1$, relates to covariate ranges and values of imprecise priors parameters.

The conclusion and future research plan on ImpLogit model are given in Chapter 5.
CHAPTER 2

BAYESIAN 2-PARAMETER LOGIT MODEL

This chapter goes into Bayesian modelling of the logit model with only two regression parameters, \( \beta_0 \) and \( \beta_1 \). Cases of increasing and decreasing logistic curves are considered in sections 2.1 and 2.2, respectively. The Dirichlet distribution is considered as a prior distribution in logit model. Sections 2.1 and 2.2 show how to induce a prior joint pdf for \( \beta = [\beta_0, \beta_1]' \).

Section 2.3 introduces the logit–normal distribution. The multivariate form of logit–normal distribution is presented in Section 2.4. The prior distribution is shown to play as a conjugate one for the posterior distribution of the binomial distribution parameter \( \theta \).

2.1 Increasing 2-Parameter Logit Model

In the logit model, it is reasonable to have correlation among the binomial distribution parameters \( \theta_i, \forall i \). Correlation comes by noticing that an increasing logistic curve means mathematically that \( \theta_1 < \theta_2 < ... < \theta_m \), for distinct \( x_i \) values. This implies having correlation between \( \theta_{i-1} \) and \( \theta_i \) since \( \theta_{i-1} \) restricts the value of \( \theta_i \). The same thing is thought of in the decreasing logistic curve where \( \theta_1 > \theta_2 > ... > \theta_m \).

The prior distribution of the successive differences, \( \theta_1, \theta_2 - \theta_1, ..., \theta_m - \theta_{m-1}, 1 - \theta_m \), in the logistic curve is assumed to follow the Dirichlet distribution as

\[
(\theta_1, \theta_2 - \theta_1, ..., \theta_m - \theta_{m-1}, 1 - \theta_m)
\sim \text{Dir}(\alpha_1, \alpha_2 - \alpha_1, ..., \alpha_m - \alpha_{m-1}, \nu - \alpha_m), \tag{2.1}
\]

such that \( \theta_1 < \theta_2 < ... < \theta_m \) and \( \alpha_1 + \alpha_2 - \alpha_1 + ... + \alpha_m - \alpha_{m-1} + \nu - \alpha_m = \nu \). Then the
hyperparameters in (2.1) are reparametrized as in (1.42)

\[(\theta_1, \theta_2 - \theta_1, ..., \theta_m - \theta_{m-1}, 1 - \theta_m)\]
\[\sim \text{Dir}\left(\nu H_1, \nu H_2 - \nu H_1, ..., \nu H_m - \nu H_{m-1}, \nu - \nu H_m\right),\]  

where \(0 \leq H_1 < H_2 < ... < H_m \leq 1\).

A prior joint distribution for \(\beta = [\beta_0, \beta_1]\) can be built by assigning a joint distribution for two \(\theta\)'s, say \(\theta_j, \theta_k\). Such a joint prior with a pdf \(f(\theta_j, \theta_k)\) is then used to induce a joint pdf \(\pi(\beta_0, \beta_1)\) by applying the method of change of variables. To find \(f(\theta_j, \theta_k)\), the aggregation property of the Dirichlet distribution is recalled. Briefly, the aggregation property is described as follows: if

\[\theta = (\theta_1, ..., \theta_m) \sim \text{Dir}(\alpha_1, ..., \alpha_m),\]  

then

\[\theta^* = (\theta_1, ..., \theta_j + \theta_k, ..., \theta_m) \sim \text{Dir}(\alpha_1, ..., \alpha_j + \alpha_k, ..., \alpha_m), \quad \forall j, k,\]  

and by applying this property to (2.2), the joint distribution of \(\theta_j, \theta_k\) is

\[(\theta_j, \theta_k - \theta_j, 1 - \theta_k) \sim \text{Dir}(\nu H_j, \nu H_k - \nu H_j, \nu - \nu H_k).\]  

This implies

\[\theta_j \sim \text{beta}(\nu H_j, \nu - \nu H_j),\]  

and

\[\theta_k \sim \text{beta}(\nu H_k, \nu - \nu H_k).\]

Based on (2.5) and (1.55), \(\theta_j\) and \(\theta_k\) have the following joint pdf (because the Jacobian determinant \(|J| = 1\))

\[f(\theta_j, \theta_k) = \frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \times (\theta_j)^{\alpha H_j - 1} (\theta_k - \theta_j)^{\nu H_k - \nu H_j - 1} (1 - \theta_k)^{\nu - \nu H_k - 1},\]  

\[0 < \theta_j < \theta_k < 1.\]
Then (2.8) is written in terms of $\beta_0$ and $\beta_1$ by applying the change of variables method

$$
\pi(\beta) = f(\theta(\beta)) \left| \frac{\partial \theta}{\partial \beta} \right|, 
$$

(2.9)

where the Jacobian matrix determinant is defined as

$$
\left| \frac{\partial \theta}{\partial \beta} \right| = \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} ,
$$

which becomes

$$
(x_k - x_j) \left\{ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right\} = (x_k - x_j) \left\{ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right\} .
$$

(2.10)

The prior joint pdf is induced by substituting (2.10) in (2.9) as

$$
\pi(\beta) = \pi(\beta_0, \beta_1) = \frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \times \left\{ \theta(\beta_0 + x_j \beta_1) \right\}^{\nu H_j - 1} \times \left\{ \theta(\beta_0 + x_k \beta_1) - \theta(\beta_0 + x_j \beta_1) \right\}^{\nu H_k - \nu H_j - 1} \times (x_k - x_j) \left\{ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right\} 
$$

$$
\times \left( \frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \times \left( \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu H_j - 1} \left( \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu H_k - \nu H_j - 1} \right) 
$$

$$
\times (x_k - x_j) \left\{ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right\} .
$$

(2.11)

The posterior joint pdf $\pi(\beta_0, \beta_1|y_1, ..., y_m)$ is then found by plugging (2.11) in (1.20). Since

$$
y_i \sim \text{bin}(\theta_i, n_i) \quad \text{and} \quad \theta_i = \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}}, \ \forall i,
$$

30
then

\[
\pi(\boldsymbol{\beta}|\mathbf{y}) = \pi(\beta_0, \beta_1|y_1, \ldots, y_m) \propto f(\mathbf{y}, \boldsymbol{\beta})
\]

\[
= \prod_{i=1}^{m} \left[ \frac{n_i}{y_i} \left( \frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}} \right)^{y_i} \left( \frac{1}{1+e^{\beta_0+x_i\beta_1}} \right)^{n_i-y_i} \right] \times \left[ \frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \times \left[ \frac{1}{1+e^{\beta_0+x_j\beta_1}} \right]^{\nu H_j-1} \left( \frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} - \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu H_k-\nu H_j-1} \\
\times \left( 1 - \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu-\nu H_k-1} \\
\times (x_k - x_j) \left[ \left( \frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_j} \left( \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right) \left( 1 - \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu H_k-\nu H_j} \\
\times (1 + \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}}) \cdot \left( 1 + \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right) \cdot (1 + \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}}) \right] \\
\times (x_k - x_j) \left[ \left( \frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_j} \left( \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right) \right] \right]
\]

\[
= \prod_{i=1}^{m} \left[ \frac{n_i}{y_i} \left( \frac{e^{\beta_0+x_i\beta_1}}{1+e^{\beta_0+x_i\beta_1}} \right)^{y_i} \left( \frac{1}{1+e^{\beta_0+x_i\beta_1}} \right)^{n_i-y_i} \right] \times \left[ \frac{\Gamma(\nu)}{\Gamma(\nu H_j) \Gamma(\nu H_k - \nu H_j) \Gamma(\nu - \nu H_k)} \right] \times \left[ \frac{1}{1+e^{\beta_0+x_j\beta_1}} \right]^{\nu H_j-1} \left( \frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} - \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu H_k-\nu H_j-1} \\
\times \left( 1 - \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right)^{\nu-\nu H_k-1} \\
\times \left( 1 + \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right) \cdot \left( 1 + \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right) \cdot (1 + \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}}) \right] \\
\times (x_k - x_j) \left[ \left( \frac{e^{\beta_0+x_j\beta_1}}{1+e^{\beta_0+x_j\beta_1}} \right)^{\nu H_j} \left( \frac{e^{\beta_0+x_k\beta_1}}{1+e^{\beta_0+x_k\beta_1}} \right) \right] \right].
\]
The assignment of joint prior distribution in (2.5) can be made for any \( j, k \), where \( \theta_j < \theta_k \). Therefore, it becomes important to find the prior correlation between any selected \( \theta_j \) and \( \theta_k \). For Dirichlet distribution, if

\[
\theta = (\theta_1, ..., \theta_m) \sim \text{Dir} (\alpha_1, ..., \alpha_m),
\]

then according to Kotz, Balakrishnan and Johnson (2000),

\[
\begin{align*}
E (\theta_j) &= \frac{\alpha_j}{\sum_{i=1}^{m} \alpha_i}, \\
\text{Var} (\theta_j) &= \frac{\alpha_j \left( \sum_{i=1}^{m} \alpha_i - \alpha_j \right)}{\left( \sum_{i=1}^{m} \alpha_i \right) \left( \sum_{i=1}^{m} \alpha_i + 1 \right)} , \\
\text{Cov} (\theta_j, \theta_k) &= \frac{-\alpha_j \alpha_k}{\left( \sum_{i=1}^{m} \alpha_i \right) \left( \sum_{i=1}^{m} \alpha_i + 1 \right)}. 
\end{align*}
\]

To find the correlation between any \( \theta_j \) and \( \theta_k \) in (2.5), using (2.16) and (2.17),

\[
\text{Cov} (\theta_j, \theta_k - \theta_j) = \text{Cov} (\theta_j, \theta_k) - \text{Var} (\theta_j)
\]

\[\Rightarrow\]

\[
\frac{-\nu H_j (\nu H_k - \nu H_j)}{\nu^2 (\nu + 1)} = \text{Cov} (\theta_j, \theta_k) - \frac{\nu H_j (\nu - \nu H_j)}{\nu^2 (\nu + 1)}
\]

\[\Rightarrow\]

\[
\text{Cov} (\theta_j, \theta_k) = \frac{\nu^2 H_j - (\nu H_j)^2 - \nu^2 H_j H_k + (\nu H_j)^2}{\nu^2 (\nu + 1)} = \frac{\nu^2 H_j (1 - H_k)}{\nu^2 (\nu + 1)} = \frac{H_j (1 - H_k)}{\nu + 1}.
\]

The covariance between any prior parameters \( \theta_j \) and \( \theta_k \) in (2.20) is always positive since having \( \theta_j < \theta_k \), \( \forall j, k \), requires \( \theta_k \) to be positively correlated to \( \theta_j \). This is different from the covariance in (2.17) which is always negative. To find \text{corr} (\theta_j, \theta_k), both of \text{Var} (\theta_j) and \text{Var} (\theta_k) have to be used, where
\[
\text{Var}(\theta_j) = \frac{\nu H_j (\nu - \nu H_j)}{\nu^2 (\nu + 1)} = \frac{H_j (1 - H_j)}{\nu + 1},
\]
(2.21)

and

\[
\text{Var}(\theta_k) = \frac{\nu H_k (\nu - \nu H_k)}{\nu^2 (\nu + 1)} = \frac{H_k (1 - H_k)}{\nu + 1}.
\]
(2.22)

Then the correlation \(\text{corr}(\theta_j, \theta_k)\) is

\[
\text{corr}(\theta_j, \theta_k) = \frac{\text{Cov}(\theta_j, \theta_k)}{\sqrt{\text{Var}(\theta_j) \text{Var}(\theta_k)}} = \frac{H_j (1 - H_k)}{\nu + 1} \sqrt{\frac{H_j (1 - H_j) H_k (1 - H_k)}{\nu + 1}}.
\]
(2.23)

Bayesian estimates of \(\beta_0\) and \(\beta_1\) are computed by finding the expectation of the posterior pdf in (2.12). Unlike the situation in the beta and Dirichlet probability models where expectations can be found by hand, the expected value of (2.12) has to be found computationally. This can be done by estimating \(\beta\) using Markov Chain Monte Carlo (MCMC) methods through applying Metropolis-Hastings (MH) sampling algorithm. This algorithm was established by Metropolis et al (1953) and being generalized by Hastings (1970). More information about MCMC methods and algorithms can be found in Robert and Casella (1999). In MH algorithm, a proposal pdf, say \(q(.)\), is designed to be the stationary distribution that plays the role of the posterior density in (2.12). The proposal function will be used to generate random values for \(\beta_0\) and \(\beta_1\), where every random drawn value will be either accepted or rejected. The means of the accepted values are then computed and considered as estimates of the parameters of interest. Here is a brief description of the MH algorithm:

- Given \(\beta^{(t)} = \left[\beta_0^{(t)}, \beta_1^{(t)}\right]^\top\).
- Generate

\[
Z \sim q\left(\beta^{(t+1)}|\beta^{(t)}\right).
\]
(2.24)
Then let
\[ p = \min [\alpha_{MH}, 1], \quad (2.25) \]
where
\[ \alpha_{MH} = \frac{\pi (Z|y) q(\beta^{(t)}|Z)}{\pi (\beta^{(t)}|y) q(Z|\beta^{(t)})}, \quad (2.26) \]
and \( \pi(.) \) is the posterior density in (2.12).

- Draw \( u \sim U[0,1] \), so that
\[ \beta^{(t+1)} = \begin{cases} Z, & u \leq p, \\ \beta^{(t)}, & u > p. \end{cases} \quad (2.27) \]

Random values of the regression parameters where drawn independently from \( t(5) \) distribution as a proposal pdf. The previous algorithm was applied and iterated over 2000 times after finding the likelihood estimates of \( \beta_0 \) and \( \beta_1 \). In each iteration, the regression parameters are generated as
\[ \beta_j^{(t)} = \hat{\sigma}_{j,LH} \cdot \text{[generated } t(5)-\text{dist value]} + \hat{\beta}_{j,LH}, \quad j = 0, 1, \quad (2.28) \]
where \( \hat{\sigma}_{j,LH} \) is the standard error of \( \hat{\beta}_j \) (determined by the information matrix in maximum likelihood estimates) and it is fixed over all iterations. Both of \( \hat{\sigma}_{j,LH} \) and \( \hat{\beta}_{j,LH} \) are required to make a transformation for \( t(5) \). Out of 2000 iterations, the first 1000 generations were burned and the other 1000 iterations were considered for estimation purposes. The generation of random draws of regression parameters for the purpose of Bayesian estimation will follow (2.28) in all of the following sections and chapters. Computations are coded using R software package and the code is given in Appendix A.

Table 2.1 presents observed data given in Govindarajulu (1988). The data describes the application of a toxin called retenone to *Macrosiphoniella sanborni*. Estimates (posterior means) of \( \beta_0 \) and \( \beta_1 \) in the logit model are given in tables 2.2 and 2.3 where the term “selected pair” indicates that the prior distribution was defined on the given \( \theta_j \) and \( \theta_k \). Such estimates are close to the maximum likelihood estimates \( \hat{\beta}_0,LH = -3.223 \) and \( \hat{\beta}_1,LH = 0.606 \). Plots of the predicted logistic curves are given in Figure 2.1. Figures 2.2
and 2.3 give the sequences of the generated random values of $\beta_0$ and $\beta_1$ for the top-left plot of Figure 2.1.

In Figure 2.2, the computational algorithm starts with $\beta_0 = -3.2$, then it keeps accepting or rejecting the next random draws. It can be seen that random draws tend to stabilize by moving around the value of -3.21. A similar thing happens to $\beta_1$ in Figure 2.3 where the sequence starts with $\beta_1 = 0.64$ then the adopted random draws tend to stabilize around 0.6.

**Table 2.1:** Binomial data where the dose is given in mg/l, $n$ is the number of experimented insects and $y_{obs}$ is the number of affected insects.

<table>
<thead>
<tr>
<th>Dose</th>
<th>$n$</th>
<th>$y_{obs}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>2.6</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>3.8</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>7.7</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>10.2</td>
<td>50</td>
</tr>
</tbody>
</table>

**Table 2.2:** $\beta_0$ estimates in increasing logit model with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\nu = 0.5$</th>
<th>$\nu = 1$</th>
<th>$\nu = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6$</td>
<td>-3.210</td>
<td>-3.221</td>
<td>-3.204</td>
</tr>
<tr>
<td>$\theta_2, \theta_5$</td>
<td>-3.216</td>
<td>-3.206</td>
<td>-3.229</td>
</tr>
<tr>
<td>$\theta_3, \theta_4$</td>
<td>-3.227</td>
<td>-3.235</td>
<td>-3.235</td>
</tr>
</tbody>
</table>

**Table 2.3:** $\beta_1$ estimates in increasing logit model with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\nu = 0.5$</th>
<th>$\nu = 1$</th>
<th>$\nu = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6$</td>
<td>0.606</td>
<td>0.606</td>
<td>0.603</td>
</tr>
<tr>
<td>$\theta_2, \theta_5$</td>
<td>0.605</td>
<td>0.602</td>
<td>0.602</td>
</tr>
<tr>
<td>$\theta_3, \theta_4$</td>
<td>0.604</td>
<td>0.607</td>
<td>0.605</td>
</tr>
</tbody>
</table>
Figure 2.1: Plots of increasing logistic functions with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$. Top: The selected pairs are $\theta_1$, $\theta_6$ (top-left), $\theta_2$, $\theta_5$ (top-middle), $\theta_3$, $\theta_4$ (top-right) and $\nu = 0.5$. Middle: $\theta_1$, $\theta_6$ (middle-left), $\theta_2$, $\theta_5$ (middle-middle), $\theta_3$, $\theta_4$ (middle-right) and $\nu = 1$. Bottom: $\theta_1$, $\theta_6$ (bottom-left), $\theta_2$, $\theta_5$ (bottom-middle), $\theta_3$, $\theta_4$ (bottom-right) and $\nu = 2$. 

$\nu = 0.5$, Pair:1,6 $\hat{\beta}_0 = -3.21$ $\hat{\beta}_1 = 0.606$ $\nu = 1$, Pair:2,5 $\hat{\beta}_0 = -3.206$ $\hat{\beta}_1 = 0.602$ $\nu = 2$, Pair:3,4 $\hat{\beta}_0 = -3.235$ $\hat{\beta}_1 = 0.607$
Figure 2.2: Sequence of 2000 random draws of $\beta_0$ in MH algorithm of top-left plot in Figure 2.1.

Figure 2.3: Sequence of 2000 random draws of $\beta_1$ in MH algorithm of top-left plot in Figure 2.1.
2.2 Decreasing 2-Parameters Logit Model

The decreasing logit model is considered if there is a belief of having lower probability of occurrence with larger value of $x_i$. The prior distribution assumption is somehow similar in its form to that in (2.2). The prior distribution is

$$(1 - \theta_1, \theta_1 - \theta_2, ..., \theta_{m-1} - \theta_m, \theta_m) \sim \text{Dir}(\nu - \nu H^*_1, \nu H^*_1 - \nu H^*_2, ..., \nu H^*_m - \nu H^*_m),$$ (2.29)

where after considering the aggregation property and hyperparameters reparametrization, the joint prior distribution for any $\theta_j, \theta_k$ becomes

$$(1 - \theta_j, \theta_j - \theta_k, \theta_k) \sim \text{Dir}(\nu - \nu H^*_j, \nu H^*_j - \nu H^*_k, \nu H^*_k),$$ (2.30)

with joint pdf as

$$f(\theta_j, \theta_k) = \frac{\Gamma(\nu)}{\Gamma(\nu - \nu H^*_j) \Gamma(\nu H^*_j - \nu H^*_k) \Gamma(\nu H^*_k)} \times (1 - \theta_j)^{\nu - \nu H^*_j - 1} (\theta_j - \theta_k)^{\nu H^*_j - \nu H^*_k - 1} (\theta_k)^{\nu H^*_k - 1},$$

$$1 > \theta_j > \theta_k > 0.$$ (2.31)

To find $\text{corr}(\theta_j, \theta_k)$,

$$\text{Cov}(\theta_j - \theta_k, \theta_k) = -\frac{(\nu H^*_j - \nu H^*_k) \nu H^*_k}{\nu^2(\nu + 1)} = \text{Cov}(\theta_j, \theta_k) - \text{Var}(\theta_k)$$

$$= \text{Cov}(\theta_j, \theta_k) - \text{Var}(\theta_k).$$ (2.32)

Since

$$(1 - \theta_k) \sim \text{beta}(\nu - \nu H^*_k, \nu H^*_k),$$ (2.33)

then

$$\text{Var}(\theta_k) = \frac{\nu H^*_k (\nu - \nu H^*_k)}{\nu^2(\nu + 1)} = \frac{H^*_k (1 - H^*_k)}{(\nu + 1)},$$ (2.34)

and the same can be found for $\theta_j$

$$\text{Var}(\theta_j) = \frac{\nu H^*_j (\nu - \nu H^*_j)}{\nu^2(\nu + 1)} = \frac{H^*_j (1 - H^*_j)}{(\nu + 1)}.$$ (2.35)
This leads to have

\[
\text{Cov}(\theta_j, \theta_k) = \frac{-\nu^2 H_j^* H_k^* + (\nu H_k^*)^2 + \nu^2 H_k^* - (\nu H_j^*)^2}{\nu^2 (\nu + 1)} = \frac{H_k^* (1 - H_j^*)}{\nu + 1},
\]

(2.36)

which results in

\[
\text{corr}(\theta_j, \theta_k) = \frac{\text{Cov}(\theta_j, \theta_k)}{\sqrt{\text{Var}(\theta_j)} \sqrt{\text{Var}(\theta_k)}} = \sqrt{\frac{H_k^* (1 - H_j^*)}{H_j^* (1 - H_k^*)}}.
\]

(2.37)

The joint posterior distribution is induced and written in terms of \(\beta_0\) and \(\beta_1\) by following the same method in deriving (2.12) so that

\[
\pi(\beta|y_1, \ldots, y_m) \propto f(y, \beta)
= \left\{ \prod_{i=1}^{m} \left[ \frac{n_i}{y_i} \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \times \frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_k^*)} \left[ 1 - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right]^{\nu - \nu H_j^* - 1} \times \left[ 1 - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right]^{\nu - \nu H_k^* - 1} \times |x_k - x_j| \left\{ \left[ \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right]^2 \left[ \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right]^2 \right\}
\]

(2.38)
the univariate and multivariate forms of imprecise logit–normal model in Chapter 3. The multivariate logit–normal model is given in Section 2.4. This prepares to introduce distribution. Section 2.3 presents a new probability model called the logit–normal model. Section 2.4 uses the data given in Table 2.1 but for number of insects that are not affected (n - y_{obs}). Tables 2.5 and 2.6 present the numerical estimates of \( \beta_0 \) and \( \beta_1 \).

Sections 2.1 and 2.2 discussed 2-parameters logit model with a prior of Dirichlet distribution. Section 2.3 presents a new probability model called the logit–normal model. The multivariate logit–normal model is given in Section 2.4. This prepares to introduce the univariate and multivariate forms of imprecise logit–normal model in Chapter 3.

\[
\begin{align*}
&= \left\{ \prod_{i=1}^{m} \left( \frac{n_i}{y_i} \right) \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right\} \\
&\times \left[ \frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_j^* - \nu H_k^*) \Gamma(\nu H_k^*)} \right] \\
&\times \left[ \frac{1}{1 + e^{\beta_0 + x_j \beta_1}} \right]^{\nu - \nu H_j^*} \left( \frac{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})(1 + e^{\beta_0 + x_k \beta_1})}\right)^{\nu H_k^* - \nu H_j^*} \\
&\times (1 + e^{\beta_0 + x_j \beta_1}) \cdot \left( \frac{1 + e^{\beta_0 + x_j \beta_1}}{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}} \right) \cdot \left( \frac{1 + e^{\beta_0 + x_j \beta_1}}{e^{\beta_0 + x_k \beta_1}} \right) \\
&\times |x_k - x_j| \left\{ \left[ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[ \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\} \\
&= \left\{ \prod_{i=1}^{m} \left( \frac{n_i}{y_i} \right) \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right\} \\
&\times \left[ \frac{\Gamma(\nu)}{\Gamma(\nu - \nu H_j^*) \Gamma(\nu H_j^* - \nu H_k^*) \Gamma(\nu H_k^*)} \right] \\
&\times \left[ \frac{1}{1 + e^{\beta_0 + x_j \beta_1}} \right]^{\nu - \nu H_j^*} \left( \frac{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})(1 + e^{\beta_0 + x_k \beta_1})}\right)^{\nu H_k^* - \nu H_j^*} \\
&\times (1 + e^{\beta_0 + x_j \beta_1}) \cdot \left( \frac{1 + e^{\beta_0 + x_j \beta_1}}{e^{\beta_0 + x_j \beta_1} - e^{\beta_0 + x_k \beta_1}} \right) \cdot \left( \frac{1 + e^{\beta_0 + x_j \beta_1}}{e^{\beta_0 + x_k \beta_1}} \right) .
\end{align*}
\]

(2.39)

Figure 2.4 gives 9 different logistic decreasing fits for data given in Table 2.4. Table 2.4 uses the data given in Table 2.1 but for number of insects that are not affected (n - y_{obs}). Tables 2.5 and 2.6 present the numerical estimates of \( \beta_0 \) and \( \beta_1 \).
Table 2.4: Binomial data where the dose is given in mg/l, \( n \) is the number of experimented insects and \( n - y_{obs} \) is the number of insects that are not affected.

<table>
<thead>
<tr>
<th>Dose</th>
<th>( n )</th>
<th>( n - y_{obs} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>2.6</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>3.8</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>7.7</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>10.2</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 2.5: \( \beta_0 \) estimates in decreasing logit model with \( H_1^* = 0.85, H_2^* = 0.7, H_3^* = 0.55, H_4^* = 0.45, H_5^* = 0.3 \) and \( H_6^* = 0.15 \).

| selected pair | \( \nu = 0.5 \) | \( \nu = 1 \) | \( \nu = 2 \) |
|---------------|-----------------|-----------------|
| \( \theta_1, \theta_6 \) | 3.216 | 3.214 | 3.177 |
| \( \theta_2, \theta_5 \) | 3.218 | 3.232 | 3.174 |
| \( \theta_3, \theta_4 \) | 3.220 | 3.214 | 3.223 |

Table 2.6: \( \beta_1 \) estimates in decreasing logit model with \( H_1^* = 0.85, H_2^* = 0.7, H_3^* = 0.55, H_4^* = 0.45, H_5^* = 0.3 \) and \( H_6^* = 0.15 \).

| selected pair | \( \nu = 0.5 \) | \( \nu = 1 \) | \( \nu = 2 \) |
|---------------|-----------------|-----------------|
| \( \theta_1, \theta_6 \) | -0.599 | -0.596 | -0.584 |
| \( \theta_2, \theta_5 \) | -0.605 | -0.606 | -0.591 |
| \( \theta_3, \theta_4 \) | -0.605 | -0.599 | -0.607 |
Figure 2.4: Plots of decreasing logistic functions with $H_1^* = 0.85$, $H_2^* = 0.7$, $H_3^* = 0.55$, $H_4^* = 0.45$, $H_5^* = 0.3$ and $H_6^* = 0.15$. Top: The selected pairs are $\theta_1$, $\theta_6$ (top-left), $\theta_2$, $\theta_5$ (top-middle), $\theta_3$, $\theta_4$ (top-right) and $\nu = 0.5$. Middle: $\theta_1$, $\theta_6$ (middle-left), $\theta_2$, $\theta_5$ (middle-middle), $\theta_3$, $\theta_4$ (middle-right) and $\nu = 1$. Bottom: $\theta_1$, $\theta_6$ (bottom-left), $\theta_2$, $\theta_5$ (bottom-middle), $\theta_3$, $\theta_4$ (bottom-right) and $\nu = 2$. 

\[
\begin{align*}
\beta_0^* &= 3.216 \\
\beta_1^* &= -0.599 \\
\nu &= 0.5, \text{Pair: 1,6} \\
\beta_0^* &= 3.218 \\
\beta_1^* &= -0.605 \\
\nu &= 0.5, \text{Pair: 2,5} \\
\beta_0^* &= 3.214 \\
\beta_1^* &= -0.596 \\
\nu &= 1, \text{Pair: 1,6} \\
\beta_0^* &= 3.223 \\
\beta_1^* &= -0.607 \\
\nu &= 2, \text{Pair: 3,4} \\
\beta_0^* &= 3.218 \\
\beta_1^* &= -0.605 \\
\nu &= 1, \text{Pair: 2,5} \\
\beta_0^* &= 3.22 \\
\beta_1^* &= -0.614 \\
\nu &= 2, \text{Pair: 3,4} \\
\beta_0^* &= 3.177 \\
\beta_1^* &= -0.584 \\
\nu &= 2, \text{Pair: 1,6} \\
\beta_0^* &= 3.174 \\
\beta_1^* &= -0.591 \\
\nu &= 2, \text{Pair: 2,5} \\
\beta_0^* &= 3.218 \\
\beta_1^* &= -0.606 \\
\nu &= 2, \text{Pair: 3,4} \\
\beta_0^* &= 3.177 \\
\beta_1^* &= -0.592 \\
\nu &= 2, \text{Pair: 1,6} \\
\beta_0^* &= 3.174 \\
\beta_1^* &= -0.591 \\
\nu &= 2, \text{Pair: 2,5} \\
\beta_0^* &= 3.22 \\
\beta_1^* &= -0.607 \\
\nu &= 2, \text{Pair: 3,4}
\end{align*}
\]
2.3 Logit–Normal Model

The logit-normal pdf is

\[ \pi (\theta | \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{\left[ \log \left( \frac{\theta}{1-\theta} \right) - \mu \right]^2}{2\sigma^2} \right\} \cdot \frac{1}{\theta (1-\theta)}, \]

where \( \mu \) is a location parameter and \( \sigma \) is a scale parameter. The pdf in (2.40) does not have explicit solutions for its mean, mode and variance as mentioned in Mead (1965). The logit-normal model was used as a prior distribution in previous studies and applications. This includes Bloch and Watson (1967) and Leonard (1973). Properties of the logit-normal model can be seen in Aitchison and Shen (1980). Some recent attempts to present methods for computing the moments for the logit-normal distribution are in Frederic and Lad (2008). Figure 2.5 gives plots for logit-normal density.

The relation between the logit-normal distribution and the normal distribution is given in Appendix B. Appendix B shows that the prior and posterior logit-normal distributions (for \( \theta \) in binomial distribution) belong to an exponential family of distributions.

\[ \]
Figure 2.5: Top-left: Plots of logit-normal pdf’s with $\mu = 0$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed). Top-right: $\mu = 0.5$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed). Bottom-left: $\mu = 1$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed). Bottom-right: $\mu = 1.5$, $\sigma = 0.5$ (solid), $\sigma = 1$ (dashed), $\sigma = 1.5$ (dotted), $\sigma = 2$ (dot-dashed), $\sigma = 2.5$ (long-dashed).
2.4 Multivariate Logit-Normal Model

The multivariate logit-normal pdf for \( \theta_{\text{m} \times 1} = [\theta_1, ..., \theta_m]' \) is

\[
\pi(\theta | \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \left[ \log \left( \frac{\theta}{1-\theta} \right) - \mu \right]' \Sigma^{-1} \left[ \log \left( \frac{\theta}{1-\theta} \right) - \mu \right] \right\} \times \prod_{i=1}^{m} \left[ \frac{1}{\theta_i (1-\theta_i)} \right],
\]

(2.41)

where

\[
\log \left( \frac{\theta}{1-\theta} \right) = \left[ \log \left( \frac{\theta_1}{1-\theta_1} \right), ..., \log \left( \frac{\theta_m}{1-\theta_m} \right) \right]',
\]

and

\[
\mu = [\mu_1, ..., \mu_m]',
\]

such that \( 0 < \theta_i < 1, -\infty < \mu_i < \infty, \Sigma_{m \times m} = [\sigma_{ij}]_{m \times m}. \)

Also, \( \Sigma_{m \times m} = [\sigma_{ij}]_{m \times m} = \sigma^2 R_{m \times m} \), where the correlation equation given in (2.23) is considered, for the increasing logistic curve, as

\[
R = \begin{bmatrix}
1 & \sqrt{\frac{H_1 (1-H_2)}{H_2 (1-H_1)}} & \cdots & \sqrt{\frac{H_1 (1-H_m)}{H_m (1-H_1)}} \\
& \ddots & \ddots & \ddots \\
& & \ddots & \ddots \\
& & & \ddots & \ddots \\
\sqrt{\frac{H_1 (1-H_m)}{H_m (1-H_1)}} & \cdots & \sqrt{\frac{H_{m-1} (1-H_m)}{H_m (1-H_{m-1})}} & 1
\end{bmatrix},
\]

(2.42)
and (2.37) for the decreasing logistic curve as

\[
R = \begin{bmatrix}
1 & \sqrt{\frac{H_2^* (1 - H_1^*)}{H_1^* (1 - H_2^*)}} & \cdots & \sqrt{\frac{H_m^* (1 - H_1^*)}{H_1^* (1 - H_m^*)}} \\
\vdots & \ddots & \ddots & \vdots \\
\sqrt{\frac{H_m^* (1 - H_1^*)}{H_1^* (1 - H_m^*)}} & \cdots & \sqrt{\frac{H_m^* (1 - H_{m-1}^*)}{H_{m-1}^* (1 - H_m^*)}} & 1
\end{bmatrix}
\]  \tag{2.43}

The posterior multivariate logit-normal pdf is

\[
\pi (\theta | y) = \pi (\theta | y_1, \ldots, y_m) = C \cdot \exp \left\{- \frac{1}{2} \left[ \log \left( \frac{\theta}{1 - \theta} \right) - (\mu + \Sigma y) \right] \right\}^{'} \Sigma^{-1} \left[ \log \left( \frac{\theta}{1 - \theta} \right) - (\mu + \Sigma y) \right] \right\}
\times \exp \left\{ 1^{'} \log \left[ \left( \frac{1 - \theta}{\theta} \right)^{n-1} \right] \right\}, \tag{2.44}
\]

where \( Y \) is a vector of independent binomial random variables.

The prior and posterior multivariate distributions in (2.41) and (2.44) belong to an exponential family of distributions. This is shown in Appendix C. If 2-parameters logit model to be fitted, then a bivariate logit–normal model is assumed as a priori after selecting \( \theta_j \) and \( \theta_k \).

To have a prior bivariate logit-normal model of \( \theta_j \) and \( \theta_k \)

\[
f (\theta_j, \theta_k | \mu_{2x1}, \Sigma_{2x2}) = \frac{1}{(2\pi)^{\frac{1}{2}} \Sigma_{2x2}^{\frac{1}{2}}} \cdot \exp \left\{- \frac{1}{2} \left[ \log \left( \frac{\theta}{1 - \theta} \right) - \mu \right]^{'} \Sigma^{-1} \left[ \log \left( \frac{\theta}{1 - \theta} \right) - \mu \right] \right\}
\times \prod_{i=j,k} \frac{1}{\theta_i (1 - \theta_i)}, \tag{2.45}
\]

where

\[
\left[ \log \left( \frac{\theta}{1 - \theta} \right) \right]_{2x1} = \begin{bmatrix}
\log \left( \frac{\theta_j}{1 - \theta_j} \right) \\
\log \left( \frac{\theta_k}{1 - \theta_k} \right)
\end{bmatrix}, \quad [\mu]_{2x1} = \begin{bmatrix}
\mu_j \\
\mu_k
\end{bmatrix}_{2x1}, \tag{2.46}
\]

46
and

$$\Sigma_{2x2} = \begin{bmatrix} \sigma_j^2 & \sigma_{jk} \\ \sigma_{kj} & \sigma_k^2 \end{bmatrix}, \quad R_{2x2} = \sigma^2 \begin{bmatrix} 1 & \rho_{\theta_j, \theta_k} \\ \rho_{\theta_j, \theta_k} & 1 \end{bmatrix}, \quad \sigma_j = \sigma_k, \quad (2.47)$$

where $\rho_{\theta_j, \theta_k}$ can take the form of (2.23) or (2.37), and fixing the value of $\sigma_{jk}$ results in determining the value of $\sigma$ in (2.47).

The prior bivariate logit-normal for regression parameters in 2-parameter ImpLogit model is

$$\pi(\beta) = f(\theta(\beta)) \frac{\partial \theta}{\partial \beta} = f(\theta_j, \theta_k) \frac{\partial \theta}{\partial \beta} = \frac{1}{(2\pi) |\Sigma_{2x2}|^2} \times \exp \left\{ -\frac{1}{2} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right] \Sigma_{2x2}^{-1} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right] \right\} \times \prod_{i=1}^2 \left\{ \frac{1}{\theta_i(\beta_0, \beta_1) [1 - \theta_i(\beta_0, \beta_1)]} \right\} \times |x_k - x_j| \left\{ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right\} \left\{ \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right\}. \quad (2.48)$$

Then the posterior density of regression parameters is

$$\pi(\beta_0, \beta_1 | y_1, \ldots, y_m) \propto f(\gamma, \beta) = \prod_{i=1}^m \left\{ \frac{n_i}{y_i} \left( 1 + e^{\beta_0 + x_i \beta_1} \right)^{y_i} \left( 1 + e^{\beta_0 + x_i \beta_1} \right)^{n_i - y_i} \right\} \times \frac{1}{(2\pi) |\Sigma_{2x2}|^2} \exp \left\{ -\frac{1}{2} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right] \Sigma_{2x2}^{-1} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right] \right\} \times \prod_{i=j,k} \left\{ \frac{1}{\theta_i(\beta_0, \beta_1) [1 - \theta_i(\beta_0, \beta_1)]} \right\} \times |x_k - x_j| \left\{ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right\} \left\{ \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right\}. \quad (2.48)$$
\[
\begin{align*}
&= \prod_{i=1}^{m} \left[ \frac{n_i}{y_i} \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \\
&\times \frac{1}{(2\pi)^{1/2} |\Sigma_{2x2}|^{1/2}} \exp \left\{ -\frac{1}{2} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right]^{\prime} \Sigma_{2x2}^{-1} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right] \right\} \\
&\times \left[ \frac{(1 + e^{\beta_0 + x_j \beta_1})^2}{e^{\beta_0 + x_j \beta_1}} \right] \left[ \frac{1 + e^{\beta_0 + x_k \beta_1}}{e^{\beta_0 + x_k \beta_1}} \right] \\
&= \prod_{i=1}^{m} \left[ \frac{n_i}{y_i} \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \\
&\times \frac{1}{(2\pi)^{1/2} |\Sigma_{2x2}|^{1/2}} \exp \left\{ -\frac{1}{2} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right]^{\prime} \Sigma_{2x2}^{-1} \left[ \log \left( \frac{\theta(\beta)}{1 - \theta(\beta)} \right) - \mu \right] \right\} \\
&\times |x_k - x_j|, \quad (2.49)
\end{align*}
\]

where
\[
\left[ \log \left( \frac{\theta}{1 - \theta} \right) \right]_{2x1} = \left[ \begin{array}{c}
\beta_0 + x_j \beta_1 \\
\beta_0 + x_k \beta_1
\end{array} \right]_{2x1}, \quad [\mu]_{2x1} = \left[ \begin{array}{c}
\mu_j \\
\mu_k
\end{array} \right]_{2x1}. \quad (2.50)
\]

Figure 2.6 shows plots of increasing logistic regression model under the bivariate log–normal prior distribution for data in Table 2.1. Computed estimates for \( \beta_0 \) and \( \beta_1 \) are given in tables 2.7 and 2.8.

**Table 2.7:** \( \beta_0 \) estimates in increasing logit model with \( H_1 = 0.15, H_2 = 0.3, H_3 = 0.45, H_4 = 0.55, H_5 = 0.7 \) and \( H_6 = 0.85 \), \( \mu_i = 0 \) (\( \mu_j = \mu_k \)), and \( \sigma_i = 0.5, 1 \) and 1.5 \( (\sigma_j = \sigma_k) \), \( i = j, k \).

<table>
<thead>
<tr>
<th>selected pair</th>
<th>( \sigma_i = 0.5 )</th>
<th>( \sigma_i = 1 )</th>
<th>( \sigma_i = 1.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_1, \theta_6 )</td>
<td>-3.233</td>
<td>-3.227</td>
<td>-3.234</td>
</tr>
<tr>
<td>( \theta_2, \theta_5 )</td>
<td>-3.241</td>
<td>-3.236</td>
<td>-3.209</td>
</tr>
<tr>
<td>( \theta_3, \theta_4 )</td>
<td>-3.236</td>
<td>-3.232</td>
<td>-3.213</td>
</tr>
</tbody>
</table>

**Table 2.8:** \( \beta_1 \) estimates in increasing logit model with \( H_1 = 0.15, H_2 = 0.3, H_3 = 0.45, H_4 = 0.55, H_5 = 0.7 \) and \( H_6 = 0.85 \), \( \mu_i = 0 \) (\( \mu_j = \mu_k \)), and \( \sigma_i = 0.5, 1 \) and 1.5 \( (\sigma_j = \sigma_k) \), \( i = j, k \).

<table>
<thead>
<tr>
<th>selected pair</th>
<th>( \sigma_i = 0.5 )</th>
<th>( \sigma_i = 1 )</th>
<th>( \sigma_i = 1.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_1, \theta_6 )</td>
<td>0.602</td>
<td>0.607</td>
<td>0.605</td>
</tr>
<tr>
<td>( \theta_2, \theta_5 )</td>
<td>0.613</td>
<td>0.605</td>
<td>0.602</td>
</tr>
<tr>
<td>( \theta_3, \theta_4 )</td>
<td>0.607</td>
<td>0.603</td>
<td>0.609</td>
</tr>
</tbody>
</table>
Figure 2.6: Plots of increasing logistic functions under logit–normal prior distribution with $H_1 = 0.15$, $H_2 = 0.3$, $H_3 = 0.45$, $H_4 = 0.55$, $H_5 = 0.7$ and $H_6 = 0.85$. Top: A plot of increasing logistic functions with selected pair $\theta_1$, $\theta_6$ (top-left), $\theta_2$, $\theta_5$ (top-middle), $\theta_3$, $\theta_4$ (top-right) and $\sigma_i = 0.5$. Middle: $\theta_1$, $\theta_6$ (middle-left), $\theta_2$, $\theta_5$ (middle-middle), $\theta_3$, $\theta_4$ (middle-right) and $\sigma_i = 1$. Bottom: $\theta_1$, $\theta_6$ (bottom-left), $\theta_2$, $\theta_5$ (bottom-middle), $\theta_3$, $\theta_4$ (bottom-right) and $\sigma_i = 2$. 

$\mu_i = 0$, $\sigma_i = 0.5$

$\hat{\beta}_0 = -3.233$, $\hat{\beta}_1 = 0.602$

$\mu_i = 0$, $\sigma_i = 0.5$

$\hat{\beta}_0 = -3.241$, $\hat{\beta}_1 = 0.613$

$\mu_i = 0$, $\sigma_i = 0.5$

$\hat{\beta}_0 = -3.236$, $\hat{\beta}_1 = 0.607$

$\mu_i = 0$, $\sigma_i = 1$

$\hat{\beta}_0 = -3.227$, $\hat{\beta}_1 = 0.607$

$\mu_i = 0$, $\sigma_i = 1$

$\hat{\beta}_0 = -3.236$, $\hat{\beta}_1 = 0.605$

$\mu_i = 0$, $\sigma_i = 1.5$

$\hat{\beta}_0 = -3.234$, $\hat{\beta}_1 = 0.605$

$\mu_i = 0$, $\sigma_i = 1.5$

$\hat{\beta}_0 = -3.209$, $\hat{\beta}_1 = 0.602$

$\mu_i = 0$, $\sigma_i = 1.5$

$\hat{\beta}_0 = -3.235$, $\hat{\beta}_1 = 0.609$
Chapter 3

ImpLogit Model

This chapter extends Chapter 2 to involve probabilistic imprecision in logit model regression parameters. Using imprecise priors in logit model creates a novel imprecise logistic regression model. The imprecise logistic regression model is suggested to be called as the “ImpLogit model”. ImpLogit model performs a set of Bayesian logit fits with interval estimates for regression parameters.

This chapter establishes a general framework for ImpLogit model. The prior distribution in the logit model will be expressed imprecisely by re-parameterizing the hyper-parameters of interest in a way similar to that made for the beta and Dirichlet models in Section 1.7. ImpLogit is the first step in establishing a certain criterion for the analysis of imprecise logistic regression. ImpLogit model will focus on computing interval estimates of the regression parameters.

The 2-parameter ImpLogit model is presented simply in Section 3.1. Then Section 3.2 explores the relation between any selected pair of binomial parameters, $\theta_j, \theta_k$ in ImpLogit model and the regression parameters, $\beta_0$ and $\beta_1$. Details of the structure of increasing and decreasing cases of ImpLogit model are given in sections 3.3 and 3.4. Section 3.5 builds the ImpLogit model with a mixture of beliefs of increasing and decreasing logistic functions where there are no assumptions on the values of $\beta_1$ which gives a general manipulation of the ImpLogit model. Section 3.6 considers fitting the ImpLogit model using the imprecise bivariate logit normal model as an imprecise prior. Finally, Section 3.7 fits ImpLogit under IDM with a small sample.
3.1 Imprecision in 2-Parameter Logit Model

ImpLogit model requires a set of logistic functions to be fitted so that a set of estimates for each probability of occurrence, $\hat{\theta}_i$ for all $i$, are found. The logistic functions in ImpLogit model that correspond to the interval endpoints of $\beta_0$ and $\beta_1$ are given as

$$\theta_i = \theta(x_i) = \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \quad \Rightarrow \quad \log \left( \frac{\theta_i}{1 - \theta_i} \right) = \beta_0 + x_i \beta_1, \quad (3.1)$$

and

$$\theta_i = \theta(x_i) = \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \quad \Rightarrow \quad \log \left( \frac{\theta_i}{1 - \theta_i} \right) = \beta_0 + x_i \beta_1, \quad (3.2)$$

where $\beta_j \in \left[ \beta_j, \bar{\beta}_j \right]$, $j = 0, 1$, and $i = 1, ..., m.$

Figure 3.1 describes ImpLogit model shape where logistic curves that correspond to endpoints of $\beta_0$ and $\beta_1$ in (3.1) and (3.2) are plotted. The two plotted logistic curves intersect at some fixed covariate point. This means that lower and upper values of regression parameters do not correspond to lower and upper logistic curves. That is, having

$$\theta_i = \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \quad \text{and} \quad \theta_i^* = \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}},$$

does not necessarily imply that $\theta_i < \theta_i^*$, $\forall i$. This can be figured out in another way where Figure 3.1 shows that regression parameters lower and upper points do not necessarily match with lower and upper points of $\theta_i$ at a fixed $x_i$, $\forall i$. That is, $\underline{\beta}_0$ and $\underline{\beta}_1$ may correspond to points that belong to the interval value of $\theta_i$, but not the lower value $\underline{\theta}_i$ at $x_i$. This is an interesting point that deserves to be investigated. Therefore, the relation between $\beta_0$, $\beta_1$ and $\theta_j$, $\theta_k$ can provide an understanding of the behaviour of the ImpLogit model. Such relation will be studied in Section 3.2.

Figure 3.2 gives sets of logistic function curves. The curves are plotted under different combinations of values $\beta_0$ and $\beta_1$ selected from determined interval values. Generally, the intervals of $\theta_i$ become shorter at extreme fixed values of $x_i$, but longer around $x_i = 0$. In Figure 3.2, the top left and right plots seem more spread and imprecise compared to the bottom ones. This is due to the length of the interval value of $\beta_1$ which is larger in both top plots. This does not mean that $\beta_0$ interval value has a negligible effect on the
ImpLogit fit imprecision, but $\beta_1$ seems to play a stronger role in determining the spread of the whole set of logistic function plots. This comes from comparing both of the top and bottom plots in Figure 3.2.

Figure 3.2 can provide information about the role of $\beta_0$ on ImpLogit imprecision. In both top plots, where $\beta_1$ interval endpoints and length are similar, $\beta_0$ made a significant change of imprecision in the whole fit. In the top-left plot there is $\beta_0 \in [0.5, 0.7]$ with interval length of $0.7 - 0.5 = 0.2$, and in the top-right plot there is $\beta_0 \in [-0.5, 0.7]$ with interval length $0.7 - (-0.5) = 1.2$. It can be seen that the whole spread in the top-right plot is more than that in the top-left one. However, this situation is not the same in the bottom plots. Despite that $\beta_0$ interval endpoints and lengths are similar to the top plots, but $\beta_1$ has a similar length with larger interval endpoints. Bottom plots show that larger $\beta_1$ endpoints may result in less effect than $\beta_0$ parameter on the imprecision in ImpLogit model fit.

Figure 3.2 shows that both logistic curves with lower and upper regression parameters values tend to coincide for extreme values of the covariate $x$. This helps to know where can imprecision be reduced in ImpLogit model. This can be useful in putting a general design to make logistic curves as close as possible.

In Figure 3.2, the green color does not pass the blue and red curves over some intervals of the covariate $x$. It is interesting to find where can this happen, for an increasing or decreasing 2-parameter logistic curves, by solving

$$
\frac{e^{\beta_0 + x \beta_1}}{1 + e^{\beta_0 + x \beta_1}} < \theta < \frac{e^{\beta_0 + x \beta_1}}{1 + e^{\beta_0 + x \beta_1}},
$$

and

$$
\frac{e^{\beta_0 + x \beta_1}}{1 + e^{\beta_0 + x \beta_1}} > \theta > \frac{e^{\beta_0 + x \beta_1}}{1 + e^{\beta_0 + x \beta_1}}.
$$

Inequality (3.3) is re-written as

$$
1 + e^{-(\beta_0 + x \beta_1)} < 1 + e^{-(\beta_0 + x \beta_1)} < 1 + e^{-(\beta_0 + x \beta_1)}
$$

$$
\Rightarrow
$$

$$
1 + e^{-(\beta_0 + x \beta_1)} < 1 + e^{-(\beta_0 + x \beta_1)}.
$$
\[ 1 + e^{-(\beta_0 + x\beta_1)} < 1 + e^{-(\beta_0 + x\beta_1)}. \] (3.6)

Then solving (3.5) gives

\[ - (\beta_0 + x\beta_1) < - \left( \frac{\beta_0 - \beta_0}{\beta_1 - \beta_1} \right), \]

\[ \Rightarrow \]

\[ \beta_0 - \beta_0 < x \left( \beta_1 - \beta_1 \right) \Rightarrow \frac{\beta_0 - \beta_0}{\beta_1 - \beta_1} < x, \] (3.7)

and solving (3.6) gives

\[ - (\beta_0 + x\beta_1) < - (\beta_0 + x\beta_1), \]

\[ \Rightarrow \]

\[ \beta_0 + x\beta_1 < \beta_0 + x\beta_1 \Rightarrow x \left( \beta_1 - \beta_1 \right) < \beta_0 - \beta_0, \]

\[ \Rightarrow \]

\[ x > \frac{\beta_0 - \beta_0}{\beta_1 - \beta_1}. \] (3.8)

Finally, merging (3.7) and (3.8) gives

\[ \min \left( \frac{\beta_0 - \beta_0}{\beta_1 - \beta_1}, \frac{\beta_0 - \beta_0}{\beta_1 - \beta_1} \right) < x. \] (3.9)

Now, inequality (3.4) is solved in a similar way to (3.3). Then (3.5) and (3.6) become

\[ 1 + e^{-(\beta_0 + x\beta_1)} > 1 + e^{-(\beta_0 + x\beta_1)}, \] (3.10)

and

\[ 1 + e^{-(\beta_0 + x\beta_1)} > 1 + e^{-(\beta_0 + x\beta_1)}. \] (3.11)

Both of (3.10) and (3.11) give

\[ x < \frac{\beta_0 - \beta_0}{\beta_1 - \beta_1}, \] (3.12)
and
\[ x < \beta_0 - \beta_1, \tag{3.13} \]
then
\[ x \leq \max \left( \beta_0 - \frac{\beta_0}{\beta_1}, \frac{\beta_0}{\beta_1} - \beta_1 \right). \tag{3.14} \]

The point of intersection of logistic curves under lower and upper values of regression parameters is
\[ \frac{1}{1 + e^{-(\beta_0 + x\beta_1)}} = \frac{1}{1 + e^{-(\beta_0 + x\beta_1)}} \]
\[ \implies e^{-(\beta_0 + x\beta_1)} = e^{-(\beta_0 + x\beta_1)} \]
\[ \beta_0 + x\beta_1 = \beta_0 + x\beta_1 \]
\[ \frac{\beta_0}{\beta_1} - \beta_0 = x \left( \frac{\beta_1}{\beta_1} - \beta_1 \right), \]
which becomes
\[ x = \frac{\beta_0}{\beta_1} - \beta_0, \tag{3.15} \]
where \( x \) in (3.15) is always having a negative value. This can be figured out in figures 3.1 and 3.2. The same is found for inequality (3.14) where \( x \) is always negative valued.

If \( x \geq 0 \) then it can be seen that
\[ \frac{e^{\beta_0 + x\beta_1}}{1 + e^{\beta_0 + x\beta_1}} < \frac{e^{\beta_0 + x\beta_1}}{1 + e^{\beta_0 + x\beta_1}}, \forall x \geq 0. \tag{3.16} \]
This means that with \( \beta_0 \in \left( \beta_0, \overline{\beta}_0 \right) \) and \( \beta_1 \in \left( \beta_1, \overline{\beta}_1 \right) \), then \( \underline{\theta} < \theta < \overline{\theta} \) as in (3.3) if \( x \) belongs to (3.9), and \( \underline{\theta} > \theta > \overline{\theta} \) as in (3.4) if \( x \) belongs to (3.14).

Figures 3.3, 3.4 and 3.5 give a better idea to visualize the interesting relation. Figure 3.3 shows clearly that all combinations of \( \beta_0 \) and \( \beta_1 \), which form a square in the top-left plot, do not necessarily transmit to a square in the rest of plots. A similar thing can be seen in figures 3.4 and 3.5.

Note that the covariate \( x_i \) in (3.9), (3.14) and (3.16) can have either negative or nonnegative values. Now, define
\[ h = \min_{i=1,\ldots,n} x_i, \tag{3.17} \]
and set

\[ \tilde{x}_i = x_i + |h|, \]  

(3.18)

so that all \( \tilde{x}_i \) values are nonnegative. This will affect the parameters in logit model as

\[ \tilde{\eta}_i = \tilde{\beta}_0 + \tilde{x}_i\tilde{\beta}_1, \]  

(3.19)

such that

\[ \eta_i = \beta_0 + x_i\beta_1, \]  

(3.20)

which means that (3.19) and (3.20) become equal if

\[ \tilde{\beta}_1 = \beta_1, \]  

(3.21)

and

\[ \tilde{\beta}_0 = \beta_0 - \beta_1|h|. \]  

(3.22)

Having the regression parameters as in (3.21) and (3.22) comes by having nonnegative values of the covariate \( \tilde{x}_i \) which falls under the case of (3.16). This means that lower and upper logistic curves will correspond to lower and upper values of the transformed regression parameters in (3.21) and (3.22) over the \( \tilde{x}_i \) values in (3.18).
Figure 3.1: Top-Left: Plots of logistic functions with $\beta_0 = 0.5$, $\beta_1 = 1.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 3.7$ (dashed). Top-Right: $\beta_0 = -0.5$, $\beta_1 = 1.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 3.7$ (dashed). Bottom-Left: $\beta_0 = 0.5$, $\beta_1 = 3.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 5.7$ (dashed). Bottom-Right: $\beta_0 = -0.5$, $\beta_1 = 3.5$ (solid), $\beta_0 = 0.7$, $\beta_1 = 5.7$ (dashed).
Figure 3.2: The red curve refers to the lower logistic curve with lower values of regression parameters. The blue curve refers to the upper logistic curve with upper values of regression parameters, while the green area is for logistic curves with in between values of regression parameters. **Top-Left:** Sets of logistic functions with $\beta_0 \in [0.5, 0.7]$ and $\beta_1 \in [1.5, 3.7]$. **Top-Right:** $\beta_0 \in [-0.5, 0.7]$ and $\beta_1 \in [1.5, 3.7]$. **Bottom-Left:** $\beta_0 \in [0.5, 0.7]$ and $\beta_1 \in [3.5, 5.7]$. **Bottom-Right:** $\beta_0 \in [-0.5, 0.7]$ and $\beta_1 \in [3.5, 5.7]$. 

57
Figure 3.3: Transformation plots from $\beta_0, \beta_1$ to $\theta_j, \theta_k$ under symmetric $x_j$ and $x_k$.

Top-left: A plot of $\beta_0$ vs $\beta_1$. Top-right: A plot of $\theta_j$ vs $\theta_k$ that corresponds to combinations of $\beta_0$ and $\beta_1$ with $x_j = -3.2$ and $x_k = 3.2$. Bottom-left: $x_j = -1.8$ and $x_k = 1.8$. Bottom-right: $x_j = -1.1$ and $x_k = 1.1$. 
Figure 3.4: Transformation plots from $\beta_0, \beta_1$ to $\theta_j, \theta_k$ under left-shifted $x_j$ and $x_k$.

Top-left: A plot of $\beta_0$ vs $\beta_1$. Top-right: A plot of $\theta_j$ vs $\theta_k$ that corresponds to combinations of $\beta_0$ and $\beta_1$ with $x_j = -4.7$ and $x_k = 1.7$. Bottom-left: $x_j = -3.3$ and $x_k = 0.3$. Bottom-right: $x_j = -2.6$ and $x_k = -0.4$. 
Figure 3.5: Transformation plots from $\beta_0, \beta_1$ to $\theta_j, \theta_k$ under right-shifted $x_j$ and $x_k$. Top-left: A plot of $\beta_0$ vs $\beta_1$. Top-right: A plot of $\theta_j$ vs $\theta_k$ that corresponds to combinations of $\beta_0$ and $\beta_1$ with $x_j = -1.7$ and $x_k = 4.7$. Bottom-left: $x_j = -0.3$ and $x_k = 3.3$. Bottom-right: $x_j = 0.4$ and $x_k = 2.6$. 
3.2 Transmission from $\theta$ to $\beta$ in 2-Parameter ImpLogit Model

The mathematical relation that maps sets of $\beta = [\beta_0, \beta_1]'$ to sets of $\theta = [\theta_j, \theta_k]'$ and vice versa is important to know. Figure 3.6 divides the plotted area of $\theta_j$ vs $\theta_k$ into 4 regions.

Figure 3.6: A plot of $\theta_j$ vs $\theta_k$. 
For region 1, points of \((\theta_j, \theta_k)\) satisfy

\[
\begin{align*}
\theta_j & \leq \theta_k, \\
1 - \theta_j & \leq \theta_k,
\end{align*}
\]

which results in

\[
1 \leq 2\theta_k \implies \theta_k \geq \frac{1}{2},
\]

and this becomes

\[
\frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \geq \frac{1}{2},
\]

\[
\implies
\]

\[
\beta_0 + x_k \beta_1 \geq \log \left( \frac{\frac{1}{2}}{1 - \frac{1}{2}} \right)
\]

\[
\implies
\]

\[
0 \leq \beta_0 + x_k \beta_1.
\]

(3.23)

For region 2

\[
\begin{align*}
\theta_j & \leq \theta_k, \\
1 - \theta_j & \geq \theta_k \implies \theta_j - 1 \leq -\theta_k,
\end{align*}
\]

which results in

\[
2\theta_j - 1 \leq 0 \implies \theta_j \leq \frac{1}{2},
\]

and this becomes as

\[
\frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \leq \frac{1}{2},
\]

\[
\implies
\]

\[
\beta_0 + x_j \beta_1 \leq \log \left( \frac{\frac{1}{2}}{1 - \frac{1}{2}} \right)
\]
\[ \beta_0 + x_j \beta_1 \leq 0. \] (3.24)

For region 3

\[ \theta_j \geq \theta_k \]
\[ 1 - \theta_j \leq \theta_k \implies \theta_j - 1 \geq -\theta_k \]

which results in

\[ 2\theta_j \geq 1 \implies \theta_j \geq \frac{1}{2}, \]

and this becomes as

\[ e^{\beta_0 + x_j \beta_1} \left( 1 + e^{\beta_0 + x_j \beta_1} \right) \geq \frac{1}{2}, \]

\[ \implies \]

\[ \beta_0 + x_j \beta_1 \geq \log \left( \frac{1}{2} \right) \]

\[ \implies \]

\[ 0 \leq \beta_0 + x_j \beta_1. \] (3.25)

For region 4

\[ \theta_j \geq \theta_k, \]
\[ 1 - \theta_j \geq \theta_k, \]

which results in

\[ 1 \geq 2\theta_k \implies \theta_k \leq \frac{1}{2}, \]

and this becomes as

\[ e^{\beta_0 + x_k \beta_1} \left( 1 + e^{\beta_0 + x_k \beta_1} \right) \leq \frac{1}{2}, \]
\[ \beta_0 + x_k \beta_1 \leq \log \left( \frac{1}{1 - \frac{1}{2}} \right) \]

\[ \Rightarrow \]

\[ \beta_0 + x_k \beta_1 \leq 0. \]  \hspace{1cm} (3.26)

Now, merging (3.23), (3.24), (3.25) and (3.26). Note that from (3.23)

\[ 0 \leq \beta_0 + x_k \beta_1, \]  \hspace{1cm} (3.27)

\[ \Rightarrow \]

\[ \frac{-\beta_0}{\beta_1} \leq x_k. \]  \hspace{1cm} (3.28)

The same can be done for (3.24), (3.25) and (3.26), to have

\[ x_j \leq \frac{-\beta_0}{\beta_1}, \quad \frac{-\beta_0}{\beta_1} \leq x_j \text{ and } x_k \leq \frac{-\beta_0}{\beta_1}, \]  \hspace{1cm} (3.29)

respectively.

The relation from \( \theta_j \) and \( \theta_k \) to \( \beta_0 \) and \( \beta_1 \) can also be seen from another corner. If

\[ a_1 < \theta_j < a_2 \quad \Rightarrow \quad a_1 < \theta_j = \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} < a_2, \]

and

\[ b_1 < \theta_k < b_2 \quad \Rightarrow \quad b_1 < \theta_k = \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} < b_2. \]

\[ \Rightarrow \quad \log \left( \frac{a_1}{1 - a_1} \right) < \beta_0 + x_j \beta_1 < \log \left( \frac{a_2}{1 - a_2} \right), \]  \hspace{1cm} (3.30)

and

\[ \log \left( \frac{b_1}{1 - b_1} \right) < \beta_0 + x_k \beta_1 < \log \left( \frac{b_2}{1 - b_2} \right). \]  \hspace{1cm} (3.31)
Let
\[
\lambda_1 = \log \left( \frac{a_1}{1 - a_1} \right), \quad \lambda_2 = \log \left( \frac{a_2}{1 - a_2} \right), \\
\lambda_3 = \log \left( \frac{b_1}{1 - b_1} \right), \quad \lambda_4 = \log \left( \frac{b_2}{1 - b_2} \right),
\]
then, from (3.30) and (3.31), solve
\[
\lambda_1 < \beta_0 + x_j \beta_1 \quad (3.32)
\]
and
\[
\beta_0 + x_k \beta_1 < \lambda_4 \quad (3.33)
\]
The inequalities in (3.32) and (3.33) can be rewritten as
\[
-\beta_0 - x_j \beta_1 < -\lambda_1 \quad (3.34)
\]
\[
\beta_0 + x_k \beta_1 < \lambda_4 \quad (3.35)
\]
Then
\[
\beta_1(x_k - x_j) < \lambda_4 - \lambda_1,
\]
⇒
\[
\beta_1 < \frac{\lambda_4 - \lambda_1}{x_k - x_j} \quad (3.36)
\]
given that \(x_j < x_k\), where without loss of generality, the following cases can be considered
\[
\begin{align*}
&x_j > 0, \ x_k > 0, \\
&x_j < 0, \ x_k < 0, \\
&x_j < 0, \ x_k > 0.
\end{align*}
\]
(3.37)

3.2.1 Positive \(x_j\) and \(x_k\)

Multiply (3.34) by \(\frac{x_k}{x_j}\) to get
\[
-\frac{x_k}{x_j} \beta_0 - \frac{x_k}{x_j} x_j \beta_1 < -\frac{x_k}{x_j} \lambda_1,
\]
and add to (3.35),
\[
\beta_0 + x_k \beta_1 < \lambda_4,
\]
to get
\[
\beta_0 \left( 1 - \frac{x_k}{x_j} \right) < \lambda_4 - \frac{x_k \lambda_1}{x_j}
\]

\[\Rightarrow\]
\[
\beta_0 < \frac{\lambda_4 - \frac{x_k \lambda_1}{x_j}}{1 - \frac{x_k}{x_j}}.
\]  \hspace{1cm} (3.38)

Furthermore, from (3.30) and (3.31),
\[
\beta_0 + x_j \beta_1 < \lambda_2 \quad \Rightarrow \quad -\lambda_2 < -\beta_0 - x_j \beta_1, \quad (3.39)
\]
and
\[
\lambda_3 < \beta_0 + x_k \beta_1. \quad (3.40)
\]

Then
\[
\lambda_3 - \lambda_2 < \beta_1 (x_k - x_j)
\]

\[\Rightarrow\]
\[
\frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_1, \quad (3.41)
\]

and after multiplying (3.39) by \(\frac{x_k}{x_j}\), (3.39) and (3.40) are written as
\[
- \frac{x_k}{x_j} \lambda_2 < - \frac{x_k}{x_j} \beta_0 - \frac{x_k}{x_j} x_j \beta_1,
\]
\[
\lambda_3 < \beta_0 + x_k \beta_1,
\]
which implies
\[
\lambda_3 - \frac{x_k}{x_j} \lambda_2 < \beta_0 \left( 1 - \frac{x_k}{x_j} \right)
\]

\[\Rightarrow\]
\[
\frac{\lambda_3 - \frac{x_k \lambda_2}{x_j}}{1 - \frac{x_k}{x_j}} < \beta_0. \quad (3.42)
\]
Finally, merging (3.36) with (3.41) yields
\[
\frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_1 < \frac{\lambda_4 - \lambda_1}{x_k - x_j},
\]
and merging (3.38) and (3.42) yields
\[
\frac{\lambda_3 - x_j \lambda_2}{1 - \frac{x_k}{x_j}} < \beta_0 < \frac{\lambda_4 - \frac{x_k}{x_j} \lambda_1}{1 - \frac{x_k}{x_j}}.
\]

If (3.43) is multiplied by \(x_j > 0\) and added to (3.44), then
\[
\frac{\lambda_3 - \frac{x_k \lambda_2}{x_j}}{1 - \frac{x_k}{x_j}} + x_j \frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_0 + x_j \beta_1 < \frac{\lambda_4 - \frac{x_k}{x_j} \lambda_1}{1 - \frac{x_k}{x_j}} + x_j \frac{\lambda_4 - \lambda_1}{x_k - x_j},
\]
and if (3.43) is multiplied by \(x_k > 0\) and added to (3.44), then
\[
\frac{\lambda_3 - \frac{x_k \lambda_2}{x_j}}{1 - \frac{x_k}{x_j}} + x_k \frac{\lambda_3 - \lambda_2}{x_k - x_j} < \beta_0 + x_k \beta_1 < \frac{\lambda_4 - \frac{x_k}{x_j} \lambda_1}{1 - \frac{x_k}{x_j}} + x_k \frac{\lambda_4 - \lambda_1}{x_k - x_j},
\]
where both of (3.45) and (3.46) can be added and divided by 2 to get
\[
\frac{\lambda_3 - \frac{x_k \lambda_2}{x_j}}{1 - \frac{x_k}{x_j}} + \frac{\lambda_3 - \lambda_2}{2 \frac{x_k - x_j}} < \beta_0 + \frac{\lambda_4 + \frac{x_k}{x_j} \lambda_1 + x_k \lambda_4 - \lambda_1}{2 \frac{x_k - x_j}} < \frac{\lambda_4 - \frac{x_k}{x_j} \lambda_1}{1 - \frac{x_k}{x_j}} + \frac{\lambda_4 - \lambda_1}{2 \frac{x_k - x_j}}.
\]

### 3.2.2 Negative \(x_j\) and \(x_k\)

Inequality (3.47) changes if \(x_j < 0\) and \(x_k < 0\). If (3.43) is multiplied by \(x_j < 0\) and (3.44) by -1, then adding them gives
\[
- \left[ \frac{\lambda_4 - \frac{x_k \lambda_1}{x_j}}{1 - \frac{x_k}{x_j}} \right] + x_j \frac{\lambda_4 - \lambda_1}{x_k - x_j} < x_j \beta_1 - \beta_0 < - \left[ \frac{\lambda_3 - \frac{x_k \lambda_2}{x_j}}{1 - \frac{x_k}{x_j}} \right] + x_j \frac{\lambda_3 - \lambda_2}{x_k - x_j},
\]
and if (3.43) is multiplied by \( x_k < 0 \) and (3.44) by -1, then adding them gives

\[
- \left[ \frac{\lambda_4 - \frac{x_k}{x_j} \lambda_1}{1 - \frac{x_k}{x_j}} \right] + x_k \frac{\lambda_4 - \lambda_1}{x_k - x_j} < x_k \beta_1 - \beta_0 < - \left[ \frac{\lambda_3 - \frac{x_k}{x_j} \lambda_2}{1 - \frac{x_k}{x_j}} \right] + x_k \frac{\lambda_3 - \lambda_2}{x_k - x_j}, \tag{3.49}
\]

where both of (3.48) and (3.49) can be added and divided by 2 to get

\[
- \left[ \frac{\lambda_4 - \frac{x_k}{x_j} \lambda_1}{1 - \frac{x_k}{x_j}} \right] + \frac{(x_j + x_k)}{2} \frac{\lambda_4 - \lambda_1}{x_k - x_j} < - \beta_0 + \frac{(x_j + x_k)}{2} \beta_1 < \\
- \left[ \frac{\lambda_3 - \frac{x_k}{x_j} \lambda_2}{1 - \frac{x_k}{x_j}} \right] + \frac{(x_j + x_k)}{2} \frac{\lambda_3 - \lambda_2}{x_k - x_j}. \tag{3.50}
\]

Inequalities (3.43), (3.44), (3.47) and (3.50) determine the set of \( \beta \) points that correspond to \( \theta \) (in case of \( x_j > 0, x_k > 0 \) or \( x_j < 0, x_k < 0 \)). This can be better seen by looking at figures 3.7, 3.8 and 3.9.

### 3.2.3 Negative \( x_j \) and Positive \( x_k \)

If \( x_j < 0 \) and \( x_k > 0 \), then multiplying (3.32) by \( \frac{x_k}{x_j} \) gives

\[
\frac{x_k}{x_j} \beta_0 + \frac{x_k}{x_j} x_j \beta_1 < \frac{x_k}{x_j} \lambda_1
\]

and add to (3.35), \( \beta_0 + x_k \beta_1 < \lambda_4 \), to get

\[
\beta_0 \left( 1 + \frac{x_k}{x_j} \right) + \beta_1 (2x_k) < \lambda_4 + \frac{x_k}{x_j} \lambda_1. \tag{3.51}
\]

Also, after multiplying (3.39) by \( \frac{x_k}{x_j} \),

\[
\frac{x_k}{x_j} \lambda_2 < \frac{x_k}{x_j} \beta_0 + \frac{x_k}{x_j} x_j \beta_1,
\]

and add to (3.40), \( \lambda_3 < \beta_0 + x_k \beta_1 \),
to get
\[ \lambda_3 + \frac{x_k}{x_j} \lambda_2 < \beta_0 \left( 1 + \frac{x_k}{x_j} \right) + \beta_1 (2x_k). \] (3.52)

Inequalities (3.51) and (3.52) are merged as
\[ \lambda_3 + \frac{x_k}{x_j} \lambda_2 < (1 + \frac{x_k}{x_j}) \beta_0 + 2x_k \beta_1 < \lambda_4 + \frac{x_k}{x_j} \lambda_1. \] (3.53)

Plots of maps from intervals of \( \theta = [\theta_j, \theta_k]^\prime \) to intervals of \( \beta = [\beta_0, \beta_1]^\prime \) are shown in figures 3.7, 3.8 and 3.9 where rectangles in \( \theta \) transmit to convex parallelograms in \( \beta \). Every \( \beta \) point has to fit inequality (3.53) except in the bottom-right plot of Figure 3.8 which fits (3.50) because \( x_j \) and \( x_k \) are fixed at negative values. The same thing is for bottom-right plot of Figure 3.9 which fits (3.47) because \( x_j \) and \( x_k \) are fixed at positive values.
Figure 3.7: Transformation plots from $\theta_j, \theta_k$ to $\beta_0, \beta_1$ under symmetric fixed $x_j$ and $x_k$. top-right: A plot of $\beta_0$ vs $\beta_1$ that corresponds to combinations of $\theta_j \in [0.05, 0.95]$ and $\theta_k \in [0.05, 0.95]$ with $x_j = -3.2$ and $x_k = 3.2$. Bottom-left: $x_j = -1.8$ and $x_k = 1.8$. Bottom-right: $x_j = -1.1$ and $x_k = 1.1$. 
Figure 3.8: Transformation plots from $\theta_j, \theta_k$ to $\beta_0, \beta_1$ under left-shifted fixed $x_j$ and $x_k$. Top-right: A plot of $\beta_0$ vs $\beta_1$ that corresponds to combinations of $\theta_j \in [0.05, 0.95]$ and $\theta_k \in [0.05, 0.95]$ with $x_j = -4.7$ and $x_k = 1.7$. Bottom-left: $x_j = -3.3$ and $x_k = 0.3$. Bottom-right: $x_j = -2.6$ and $x_k = -0.4$. 
Figure 3.9: Transformation plots from $\theta_j, \theta_k$ to $\beta_0, \beta_1$ under right-shifted fixed $x_j$ and $x_k$. top-right: A plot of $\beta_0$ vs $\beta_1$ that corresponds to combinations of $\theta_j \in [0.05, 0.95]$ and $\theta_k \in [0.05, 0.95]$ with $x_j = -1.7$ and $x_k = 4.7$. Bottom-left: $x_j = -0.3$ and $x_k = 3.3$. Bottom-right: $x_j = 0.4$ and $x_k = 2.6$. 

72
3.3 2-Parameter Increasing ImpLogit Model

This section builds on Section 2.1 to involve imprecision in the increasing ImpLogit model. In this case, $\beta_1$ interval estimate has to have positive limits but no restrictions on interval limits of the regression parameter $\beta_0$. Let $H_j = \varphi_{jk1}$, $H_k - H_j = \varphi_{jk2}$ and $1 - H_k = \varphi_{jk3}$ where $\varphi_{jk1} + \varphi_{jk2} + \varphi_{jk3} = 1$. Then the posterior distribution given in (2.12) is rewritten, under a selected $\theta_j, \theta_k$, as

$$
\pi(\beta | y_1, \ldots, y_m) \propto f(y, \beta) = \prod_{i=1}^{m} \left[ \binom{n_i}{y_i} \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right]
$$

$$
\times \left[ \frac{\Gamma(\nu)}{\Gamma(\nu \varphi_{jk1}) \Gamma(\nu \varphi_{jk2}) \Gamma(\nu \varphi_{jk3})} \right] \times \left[ \left( \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu \varphi_{jk1} - 1} \left( \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu \varphi_{jk2} - 1} \right] \times \left( 1 - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu \varphi_{jk3} - 1} \right]
\times |x_k - x_j| \left\{ \left[ \frac{e^{\beta_0 + x_j \beta_1}}{(1 + e^{\beta_0 + x_j \beta_1})^2} \right] \left[ \frac{e^{\beta_0 + x_k \beta_1}}{(1 + e^{\beta_0 + x_k \beta_1})^2} \right] \right\},
$$

(3.54)

where the prior parameters sum to

$$
\nu H_j + (\nu H_k - \nu H_j) + (\nu - \nu H_k) = \nu H_j + \nu H_k - \nu H_j + \nu (1 - H_k) = \nu \varphi_{jk1} + \nu \varphi_{jk2} + \nu \varphi_{jk3} = \nu (\varphi_{jk1} + \varphi_{jk2} + \varphi_{jk3}) = \nu.
$$

Also, since $H_j \in [0, 1]$ then

$$
\min (\varphi_{jk1}) = \min (H_j) = 0 \quad , \quad \max (\varphi_{jk1}) = \max (H_j) = 1,
$$

and

$$
\min (\varphi_{jk3}) = \min (1 - H_k) = 0 \quad , \quad \max (\varphi_{jk3}) = \max (1 - H_k) = 1.
$$

If

$$
\varphi_{jk1} \rightarrow 0 \quad \text{and} \quad \varphi_{jk3} \rightarrow 1, \quad \text{then} \quad \varphi_{jk2} \rightarrow 0,
$$

(3.55)

and

$$
\varphi_{jk1} \rightarrow 1 \quad \text{and} \quad \varphi_{jk3} \rightarrow 0, \quad \text{then} \quad \varphi_{jk2} \rightarrow 0,
$$

(3.56)
The prior distribution in (2.11) is not considered as a pdf when \(\nu = 0\), but it is called an improper prior. This is because if \(\nu = 0\) then \(\Gamma(\nu)\) in (2.11) is undefined. The posterior distribution in (3.54) becomes an improper one if \(\nu = 0\) and \(y_i = 0\) or \(n_i, \forall i\). The same thing happens if \(\varphi_{jkl} = 0\) or 1, \(\forall l = 1, 2, 3\), and \(y_i = 0\) or \(n_i, \forall i\). In these cases, the improper prior does not define a posterior distribution.

The joint posterior densities in (3.54) can now be fitted, subject to (3.55) and (3.56), to find the interval estimates of \(\beta_0\) and \(\beta_1\). Since all \(\varphi_{jk1}, \varphi_{jk2}\) and \(\varphi_{jk3}\) belong to \([0, 1]\), then a set of logistic function fits can be performed for (3.54) with a set of values of \(\varphi_{jk1}, \varphi_{jk2}\) and \(\varphi_{jk3}\). Estimating regression parameters in (3.54) at cases of (3.55) and (3.56) does not necessarily produce lower and upper estimates. This situation is different from that in Section 1.7 where a re-look at (1.60) and (1.61) shows that lower and upper expectations correspond to limiting values of \(\varphi_i, \forall i\). The induced prior pdf in (2.11) is not a conjugate one to the posterior density function in (3.54). Consequently, estimating the regression parameters under (3.55) and (3.56) does not suffice to find their interval estimates, but should be done under a set of values of \(\varphi_{jk1}, \varphi_{jk2}\) and \(\varphi_{jk3}\).

Tables 3.1, 3.2 and 3.3 provide the computed lower, upper and interval estimates for \(\beta_0\) and \(\beta_1\) when \(\nu = 0.5, 1\) and 2, respectively, for the real data given in Chapter 2 in Table 2.1. Computations are performed by following Metropolis Hastings algorithm. The class of logistic fits is gained by changing the values of \(\varphi_{jk1}, \varphi_{jk2}\) and \(\varphi_{jk3}\), provided that they sum to 1 and \(H_j < H_k, \forall j, k\). Then \(H_j = \varphi_{jk1}\) was assigned 19 values and other 19 values are given to \(H_k = 1 - \varphi_{jk3}\). The assigned values for \(H_j\) are

\[
0.001, 0.14, 0.19, \ldots, 0.998, \tag{3.57}
\]

while the values for \(H_k\) are

\[
0.002, 0.15, 0.20, \ldots, 0.999, \tag{3.58}
\]

provided that \(H_j < H_k, \forall j, k\). This means that there is \(19+18+\ldots+1 = 190\) logistic functions to be fitted in each ImpLogit fit.
Table 3.1: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to top plots in Figure 3.10 for 2-parameter increasing ImpLogit model when $\nu = 0.5$. $\text{imp}(\hat{\beta}_i)$ refers to the interval estimate length.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>$\text{imp}(\hat{\beta}_0)$</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>$\text{imp}(\hat{\beta}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>-3.226</td>
<td>-3.221</td>
<td>0.005</td>
<td>0.589</td>
<td>0.613</td>
<td>0.024</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>-3.226</td>
<td>-3.221</td>
<td>0.005</td>
<td>0.586</td>
<td>0.612</td>
<td>0.026</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>-3.228</td>
<td>-3.223</td>
<td>0.005</td>
<td>0.585</td>
<td>0.613</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Table 3.2: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to middle plots in Figure 3.10 for 2-parameter increasing ImpLogit model when $\nu = 1$. $\text{imp}(\hat{\beta}_i)$ refers to the interval estimate length.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>$\text{imp}(\hat{\beta}_0)$</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>$\text{imp}(\hat{\beta}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>-3.226</td>
<td>-3.221</td>
<td>0.005</td>
<td>0.595</td>
<td>0.621</td>
<td>0.026</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>-3.227</td>
<td>-3.218</td>
<td>0.009</td>
<td>0.593</td>
<td>0.622</td>
<td>0.029</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>-3.228</td>
<td>-3.216</td>
<td>0.012</td>
<td>0.591</td>
<td>0.624</td>
<td>0.033</td>
</tr>
</tbody>
</table>

Table 3.3: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to bottom plots in Figure 3.10 for 2-parameter increasing ImpLogit model when $\nu = 2$. $\text{imp}(\hat{\beta}_i)$ refers to the interval estimate length.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>$\text{imp}(\hat{\beta}_0)$</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>$\text{imp}(\hat{\beta}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>-3.227</td>
<td>-3.219</td>
<td>0.008</td>
<td>0.597</td>
<td>0.624</td>
<td>0.027</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>-3.229</td>
<td>-3.216</td>
<td>0.013</td>
<td>0.594</td>
<td>0.624</td>
<td>0.030</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>-3.229</td>
<td>-3.215</td>
<td>0.014</td>
<td>0.592</td>
<td>0.626</td>
<td>0.034</td>
</tr>
</tbody>
</table>
Numerical results show that increasing $\nu$ value results in having higher imprecision (longer interval estimates) in regression parameters. Having a closer sequential order of selected $\theta_j$ and $\theta_k$ increases the lengths of interval estimates of $\beta_0$ and $\beta_1$.

Tables 3.4, 3.5 and 3.6 present interval estimates of $\theta$ when $x = 5$ that correspond to regression parameters interval estimates in tables 3.1, 3.2 and 3.3, respectively. Since $x_i \geq 0, \forall i = 1, ..., 6$, for data given in Table 2.1, then based on inequality (3.16), each $\hat{\theta}$ corresponds to $\overline{\beta} = \left[ \hat{\beta}_0, \hat{\beta}_1 \right]'$ and each $\bar{\theta}$ corresponds to $\overline{\beta} = \left[ \overline{\beta}_0, \overline{\beta}_1 \right]'$.

ImpLogit plots are given in Figure 3.10. Each ImpLogit plot in Figure 3.10 is a collection of 190 logistic function fits.

### Table 3.4: Interval estimates of $\theta$ for Table 3.1 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\bar{\theta}$</th>
<th>$\text{imp}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>0.430</td>
<td>0.461</td>
<td>0.031</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>0.426</td>
<td>0.460</td>
<td>0.034</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>0.425</td>
<td>0.461</td>
<td>0.036</td>
</tr>
</tbody>
</table>

### Table 3.5: Interval estimates of $\theta$ for Table 3.2 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\bar{\theta}$</th>
<th>$\text{imp}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>0.438</td>
<td>0.471</td>
<td>0.033</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>0.435</td>
<td>0.473</td>
<td>0.038</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>0.432</td>
<td>0.476</td>
<td>0.044</td>
</tr>
</tbody>
</table>

### Table 3.6: Interval estimates of $\theta$ for Table 3.3 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\bar{\theta}$</th>
<th>$\text{imp}(\hat{\theta})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>0.440</td>
<td>0.475</td>
<td>0.035</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>0.436</td>
<td>0.476</td>
<td>0.040</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>0.433</td>
<td>0.479</td>
<td>0.046</td>
</tr>
</tbody>
</table>
Figure 3.10: Top: Plots of increasing ImpLogit model for the real data given in Table 2.1 with \( \nu = 0.5 \) and selected pairs \( \theta_1, \theta_6 \) (top-left), \( \theta_2, \theta_5 \) (top-middle) and \( \theta_3, \theta_4 \) (top-right). Middle: \( \nu = 1 \) and selected pairs \( \theta_1, \theta_6 \) (middle-left), \( \theta_2, \theta_5 \) (middle-middle) and \( \theta_3, \theta_4 \) (middle-right). Bottom: \( \nu = 2 \) and selected pairs \( \theta_1, \theta_6 \) (bottom-left), \( \theta_2, \theta_5 \) (bottom-middle) and \( \theta_3, \theta_4 \) (bottom-right).
3.4 2-Parameter Decreasing ImpLogit Model

This section presents ImpLogit model with negative interval limits for $\beta_1$. Let $\varphi_{jk1} = 1 - H_j^*$, $\varphi_{jk2} = H_j^* - H_k^*$ and $\varphi_{jk3} = H_k^*$, then the posterior distribution in (2.38) is rewritten as

$$
\pi(\beta|y_1, \ldots, y_m) \propto f(y, \beta)
$$

$$
= \left\{ \prod_{i=1}^{m} \left[ \frac{n_i}{y_i} \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^y \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \times \left[ \Gamma(\nu_{\varphi_{jk1}}) \Gamma(\nu_{\varphi_{jk2}}) \Gamma(\nu_{\varphi_{jk3}}) \right] \times \left[ \left( 1 - \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu_{\varphi_{jk1}} - 1} \left( \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu_{\varphi_{jk2}} - 1} \left( \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu_{\varphi_{jk3}}} \right] \times |x_k - x_j| \left\{ \left[ \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right] \left[ \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right] \right\} \right\} \right\} \right\} \right\}

(3.59)

Now, the regression parameters in (3.59) can be estimated. Tables 3.7, 3.8 and 3.9 confirm that $\beta_1$ interval estimates include only negative values which fits the decreasing ImpLogit model for data given in Table 2.4. The assigned values for $H_j^*$ are

$$0.999, \ldots, 0.20, 0.15, 0.002,$$

(3.60)

while the values for $H_k^*$ are

$$0.998, \ldots, 0.19, 0.14, 0.001,$$

(3.61)

provided that $H_j^* > H_k^*$, $\forall j, k$. Increasing $\nu$ value results in having longer interval estimates of regression parameters.
Table 3.7: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to top plots in Figure 3.11 for 2-parameter decreasing ImpLogit model when $\nu = 0.5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_0$ imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_1$ imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>3.221</td>
<td>3.227</td>
<td>0.006</td>
<td>-0.612</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>3.220</td>
<td>3.228</td>
<td>0.008</td>
<td>-0.614</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>3.218</td>
<td>3.229</td>
<td>0.011</td>
<td>-0.616</td>
</tr>
</tbody>
</table>

Table 3.8: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to middle plots in Figure 3.11 for 2-parameter decreasing ImpLogit model when $\nu = 1$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_0$ imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_1$ imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>3.220</td>
<td>3.228</td>
<td>0.008</td>
<td>-0.614</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>3.219</td>
<td>3.229</td>
<td>0.010</td>
<td>-0.616</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>3.217</td>
<td>3.231</td>
<td>0.014</td>
<td>-0.618</td>
</tr>
</tbody>
</table>

Table 3.9: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to bottom plots in Figure 3.11 for 2-parameter decreasing ImpLogit model when $\nu = 2$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_0$ imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_1$ imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>3.218</td>
<td>3.232</td>
<td>0.014</td>
<td>-0.615</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>3.216</td>
<td>3.233</td>
<td>0.017</td>
<td>-0.617</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>3.215</td>
<td>3.231</td>
<td>0.016</td>
<td>-0.618</td>
</tr>
</tbody>
</table>
Interval estimates of $\theta$ are in tables 3.10, 3.11 and 3.12. Figure 3.11 gives the decreasing ImpLogit fits.

Table 3.10: Interval estimates of $\theta$ for Table 3.7 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\tilde{\theta}$</th>
<th>imp($\tilde{\theta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>0.540</td>
<td>0.565</td>
<td>0.025</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>0.537</td>
<td>0.568</td>
<td>0.031</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>0.534</td>
<td>0.572</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Table 3.11: Interval estimates of $\theta$ for Table 3.8 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\tilde{\theta}$</th>
<th>imp($\tilde{\theta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>0.537</td>
<td>0.568</td>
<td>0.031</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>0.535</td>
<td>0.571</td>
<td>0.036</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>0.532</td>
<td>0.575</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Table 3.12: Interval estimates of $\theta$ for Table 3.9 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\tilde{\theta}$</th>
<th>imp($\tilde{\theta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>0.536</td>
<td>0.570</td>
<td>0.034</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>0.533</td>
<td>0.576</td>
<td>0.043</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>0.531</td>
<td>0.577</td>
<td>0.046</td>
</tr>
</tbody>
</table>
Figure 3.11: Top: Plots of decreasing ImpLogit model for the real data given in Table 2.4 with \( \nu = 0.5 \) and selected pairs \( \theta_1, \theta_6 \) (top-left), \( \theta_2, \theta_5 \) (top-middle) and \( \theta_3, \theta_4 \) (top-right). Middle: \( \nu = 1 \) and selected pairs \( \theta_1, \theta_6 \) (middle-left), \( \theta_2, \theta_5 \) (middle-middle) and \( \theta_3, \theta_4 \) (middle-right). Bottom: \( \nu = 2 \) and selected pairs \( \theta_1, \theta_6 \) (bottom-left), \( \theta_2, \theta_5 \) (bottom-middle) and \( \theta_3, \theta_4 \) (bottom-right).
3.5 Mixture of Beliefs in 2-Parameter ImpLogit Model

Cases of increasing and decreasing ImpLogit models can be generalized and merged in one general form. This happens if there is a mixture of beliefs about the behaviour of the logistic models. To think of the logit model as a mixture of two models, assume that $-\infty < \beta_1 < \infty$, then the joint prior pdf is written as a mixture of two joint pdf’s of the increasing and decreasing logistic regression models, $f_1 (\theta_j, \theta_k)$ and $f_2 (\theta_j, \theta_k)$, as follows

$$f (\theta_j, \theta_k) = w f_1 (\theta_j, \theta_k) + (1 - w) f_2 (\theta_j, \theta_k) ,$$

where $w \in [0, 1]$ refers to the weight that reflects the degree of belief in $f_1 (\theta_j, \theta_k)$ and $f_2 (\theta_j, \theta_k)$. If $w = 1$, then $f (\theta_j, \theta_k)$ is the same as in Section 3.3, and if $w = 0$, then the mixture joint density function is similar to that in Section 3.4.

The joint prior pdf is

$$\pi (\beta | y_1, ..., y_m) \propto f (y, \beta) = \left\{ \prod_{i=1}^{m} \left[ \binom{n_i}{y_i} \left( \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{y_i} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)^{n_i - y_i} \right] \right\} \times w \left[ \frac{\Gamma (\nu)}{\Gamma (\nu H_j) \Gamma (\nu H_k - \nu H_j) \Gamma (\nu - \nu H_k)} \right] \times \left[ \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right]^{\nu H_j - 1} \left( \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_k - \nu H_j - 1} \times \left( 1 - \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu - \nu H_k - 1} \cdot |x_k - x_j| \left\{ \left[ \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right] \left[ \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right] \right\} \times (1 - w) \left[ \frac{\Gamma (\nu)}{\Gamma (\nu - \nu H_j^*) \Gamma (\nu H_j^* - \nu H_k^*) \Gamma (\nu H_k^*)} \right] \times \left[ \frac{1}{1 + e^{\beta_0 + x_j \beta_1}} \right]^{\nu - \nu H_j^* - 1} \left( \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right)^{\nu H_j^* - \nu H_k^* - 1} \times \left( \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right)^{\nu H_k^* - 1} \cdot |x_k - x_j| \left\{ \left[ \frac{e^{\beta_0 + x_j \beta_1}}{1 + e^{\beta_0 + x_j \beta_1}} \right] \left[ \frac{e^{\beta_0 + x_k \beta_1}}{1 + e^{\beta_0 + x_k \beta_1}} \right] \right\} .$$

The correlation between $\theta_j$ and $\theta_k$ is found by deriving the prior covariance of $\theta_j$ and
\( \theta_k \) from the joint pdf given in (3.62) as

\[
\text{Cov}(\theta_j, \theta_k) = E\{[\theta_j - E(\theta_j)][\theta_k - E(\theta_k)]\} \\
= wE\{[\theta_j - E_1(\theta)] [\theta_k - E_1(\theta_k)]\} + (1-w)E\{[\theta_j - E_2(\theta_j)] [\theta_k - E_2(\theta_k)]\} \\
= wE\{[\theta_j - E_1(\theta)] [\theta_k - E_1(\theta_k)]\} \\
+ (1-w)E\{[\theta_j - E_2(\theta_j)] [\theta_k - E_2(\theta_k)]\} \\
= w\text{Cov}_1(\theta_j, \theta_k) + (1-w)\text{Cov}_2(\theta_j, \theta_k) \\
= w\frac{H_j (1-H_k)}{\nu + 1} + (1-w)\frac{H_k^* (1-H_j^*)}{\nu + 1}. \tag{3.64}
\]

where \( \text{Cov}_1(\theta_j, \theta_k) \) and \( \text{Cov}_2(\theta_j, \theta_k) \) are given in (2.20) and (2.36), respectively. If \( w = 1 \), then (3.64) becomes as in (2.20), and if \( w = 0 \), the covariance is similar to that in (2.36). Now, \( \text{Var}(\theta_j) \) and \( \text{Var}(\theta_k) \) are required. Using the aggregation property gives the probability density function of \( \theta_i \) as

\[
f(\theta_i) = wf_1(\theta_i) + (1-w)f_2(\theta_i). \tag{3.65}
\]

Using variance equations in (2.22) and (2.34) gives, for any \( i = 1, ..., m \),

\[
\text{Var}(\theta_i) = w\frac{H_i (1-H_i)}{\nu + 1} + (1-w)\frac{H_i^* (1-H_i^*)}{\nu + 1}. \tag{3.66}
\]

Now based on (3.64) and (3.66), the correlation is

\[
\text{corr}(\theta_j, \theta_k) = \\
\frac{w\frac{H_j (1-H_k)}{\nu + 1} + (1-w)\frac{H_k^* (1-H_j^*)}{\nu + 1}}{\sqrt{\left[w\frac{H_j (1-H_j)}{\nu + 1} + (1-w)\frac{H_j^* (1-H_j^*)}{\nu + 1}\right] \sqrt{\left[w\frac{H_k (1-H_k)}{\nu + 1} + (1-w)\frac{H_k^* (1-H_k^*)}{\nu + 1}\right]}}}. \tag{3.67}
\]

If \( w = 1 \), then (3.67) becomes as in (2.23), and if \( w = 0 \), the correlation is similar to (2.37). Calculations and graphs of this section and sections 3.6 and 3.7 are given in Appendix D. Tables D.1, D.2 and D.3 provide interval estimates of regression parameters by fitting data in Table 2.1 for different values of \( w \). Interval estimates of \( \theta \) are in tables D.4, D.5 and D.6. Values of \( H_j, H_k, H_j^* \) and \( H_k^* \) are the same as in (3.57), (3.58), (3.60) and (3.61).
Increasing $w$ value makes the mixed-belief 2-parameter ImpLogit model closer to being an increasing one, where interval estimates lengths become shorter as can be seen in tables D.1, D.2 and D.3. The same thing can be found in interval estimates of $\theta$ at $x = 5$, in tables D.4, D.5 and D.6. ImpLogit fits are in Figure D.1.

### 3.6 ImpLogit with Imprecise Bivariate Logit-Normal Model

Sections 2.3 and 2.4 are now extended to involve imprecision in logit–normal distribution with application to ImpLogit model. This section introduces imprecise logit–normal model according to Bickis (2009) who was the first to present it. The new imprecise model will be briefly referred to by ILnM. There will be a focus on the imprecise bivariate logit–normal model (IBLnM).

Bickis (2009) applied the imprecise logit–normal model to the estimation of hazard functions. The application of imprecise models to estimate hazard functions was also considered by Bickis and Bickis (2007), where lower and upper probabilities are derived for the imminent recurrence of pandemic influenza. Bickis and Bickis (2007) made the application using data given in Patterson (1987).

In ILnM, a family of logit–normal pdf’s will be built. In Bickis (2009), the mode is considered as an estimator of the parameter $\theta$. If the derivative of the log of the logit-normal density function is taken, then a solution for the mode can be found. The derivative of (B.6) is

$$
\frac{\partial}{\partial \theta} \left\{ \log \left[ f(\theta|\mu, \sigma) \right] \right\} =
$$

$$
\frac{\partial}{\partial \theta} \left[ \log \left( \frac{1}{\sigma \sqrt{2\pi}} \right) - \frac{\mu^2}{2\sigma^2} \right] + \frac{\partial}{\partial \theta} \left\{ -\frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]^2 \right\}
$$

$$
+ \frac{\partial}{\partial \theta} \left[ \left( \frac{\mu}{\sigma^2} - 1 \right) \log(\theta) \right] + \frac{\partial}{\partial \theta} \left[ -\left( \frac{\mu}{\sigma^2} + 1 \right) \log(1-\theta) \right]
$$

$$
= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \theta} \left\{ \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]^2 \right\} + \left( \frac{\mu}{\sigma^2} - 1 \right) \frac{\partial}{\partial \theta} [\log(\theta)]
$$

$$
- \left( \frac{\mu}{\sigma^2} + 1 \right) \frac{\partial}{\partial \theta} [\log(1-\theta)]
$$
\[
\begin{align*}
&= -\frac{1}{\sigma^2} \left[ \log\left(\frac{\theta}{1-\theta}\right) \right] 
&\quad + \frac{1}{\sigma^2} \left[ \frac{(\mu - \theta)^2}{\theta (1 - \theta)} \right] 
&\quad + \left( \frac{\mu}{\sigma^2} - 1 \right) \frac{1}{\theta} 
&\quad + \left( \frac{\mu}{\sigma^2} + 1 \right) \frac{1}{1 - \theta} \\
&= -\frac{1}{\sigma^2} \left[ \log\left(\frac{\theta}{1-\theta}\right) \right] 
&\quad + \frac{\mu}{\sigma^2} \left( \frac{1}{\theta} + \frac{1}{1 - \theta} \right) 
&\quad - \frac{1}{\theta} + \frac{1}{1 - \theta} \\
&= -\frac{1}{\sigma^2} \left[ \log\left(\frac{\theta}{1-\theta}\right) \right] 
&\quad + \frac{\mu}{\sigma^2} \left[ \frac{1}{\theta (1 - \theta)} \right] 
&\quad + 2\theta - 1 \\
&= -\frac{1}{\sigma^2} \left[ \log\left(\frac{\theta}{1-\theta}\right) \right] 
&\quad + \frac{\mu}{\sigma^2} + 2\theta - 1 = 0 \\
\Rightarrow \quad \frac{\mu}{\sigma^2} + 2\theta - 1 &= \frac{1}{\sigma^2} \left[ \log\left(\frac{\theta}{1-\theta}\right) \right], \quad (3.68)
\end{align*}
\]
and this results in having the mode for the logit–normal pdf in (2.40) to satisfy
\[
\log\left(\frac{\theta}{1-\theta}\right) = \sigma^2 (2\theta - 1) + \mu, \quad (3.69)
\]
where the line in (3.69) intersects the \(\theta\)-axis when
\[
0 = \sigma^2 (2\theta - 1) + \mu, \quad (3.70)
\]
which means that
\[
\theta_0 = \frac{1}{2} \left( \frac{-\mu}{\sigma^2} + 1 \right) = \frac{(-1 + \mu)}{-2}. \quad (3.71)
\]
Based on (B.7), (B.9) and (3.69), there is
\[
\log\left(\frac{\theta}{1-\theta}\right) = \frac{(\eta_2^* + \eta_3^*) \theta - \eta_2^*}{2\eta_1^*} \\
= \frac{(\eta_2 + y + \eta_3 + n - y) \theta - \eta_2 - y}{2\eta_1} \\
= \frac{\left( \frac{\mu}{\sigma^2} - 1 + y - \frac{\mu}{\sigma^2} - 1 + n - y \right) \theta - \frac{\mu}{\sigma^2} + 1 - y}{2\sigma^2} \\
= -\sigma^2 \left[ \left( \frac{\mu}{\sigma^2} - 1 + y - \frac{\mu}{\sigma^2} - 1 + n - y \right) \theta - \frac{\mu}{\sigma^2} + 1 - y \right] \\
= -\sigma^2 (n - 2) \theta + \sigma^2 \left( \frac{\mu}{\sigma^2} + y - 1 \right) \\
= -(n - 2) \sigma^2 \theta + \mu + (y - 1) \sigma^2. \quad (3.72)
\]
The line in (3.72) intersects the \( \theta \)-axis at

\[
\theta_0 = \frac{y - 1}{n - 2} + \frac{\mu}{(n - 2)\sigma^2} = \frac{y - 1 + \frac{\mu}{\sigma^2}}{n - 2}, \quad n > 2, \tag{3.73}
\]

According to Bickis (2009), the posterior modes must lie in

\[
\left\{ \inf_{\pi(\theta|\mu, \sigma) \in \mathcal{A}_{\text{prior}}} \text{mode}[\pi(\theta|y)], \sup_{\pi(\theta|\mu, \sigma) \in \mathcal{A}_{\text{prior}}} \text{mode}[\pi(\theta|y)] \right\}
\]

\[
= \left\{ \inf_{\pi(\theta|\mu, \sigma) \in \mathcal{A}_{\text{prior}}} \left[ \frac{y - 1 + \frac{\mu}{\sigma^2}}{n - 2} \right], \sup_{\pi(\theta|\mu, \sigma) \in \mathcal{A}_{\text{prior}}} \left[ \frac{y - 1 + \frac{\mu}{\sigma^2}}{n - 2} \right] \right\}, \tag{3.74}
\]

where the class of logit–normal distributions is

\[
\mathcal{A}_{\text{prior}} = \left\{ \text{logit-normal}(\mu, \sigma) : \sigma = \sigma_0 + \tau |\mu|^\frac{1}{2}, -\infty < \mu < \infty \right\}, \tag{3.75}
\]

such that \( \sigma_0 > 0 \) and \( \tau < 1 \). A simple comparison of (3.71) to (3.73) shows that (3.73) updates (3.71).

Bickis (2009) uses (3.75) to treat the logit-normal model imprecisely. Bickis broke the habit by looking at an imprecise prior and posterior distributions other than the usual used IDM model as in Walley (1996) and Bernard (2005).

In this section, to suggest a class of priors in imprecise logit–normal prior for 2-parameter ImpLogit model, the bivariate case denoted by IBLnM is used. Under any selected pair of \( \theta_j, \theta_k \), let the class of logit–normal priors be

\[
\mathcal{A}_{\text{prior}} = \left\{ \text{logit-normal}(\mu_j, \sigma_j, \sigma_k, \sigma_{jk}, \sigma_{kj}) : (\mu_j, \mu_k) \in (\mu), (\sigma_j, \sigma_k, \sigma_{jk}, \sigma_{kj}) \in (\sigma) \right\}. \tag{3.76}
\]

Imprecision in bivariate logit–normal model is taken into account by varying the values of elements in \( \mu_{2 \times 1} \) and \( \Sigma_{2 \times 2} \). This requires allowing the location parameters \( \mu_j \) and \( \mu_k \) to take values over intervals that belong to the line of real numbers, \( \mathbb{R} \). The same can be done for the scale parameters \( \sigma_j \) and \( \sigma_k \) but over intervals of positive values.

The correlation form in (3.67) is used to model the covariance matrix with \( w = 1 \), \( H_j \) and \( H_k \) values as in (3.57) and (3.58). Regression parameters estimates are found using the posterior of bivariate logit-normal distribution given in (2.49), under a set of prior believed values of \( \mu \) and \( \sigma \).
Tables D.7, D.8 and D.9 provide the computed interval estimates for data in Table 2.1. The location parameters are fixed as $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$. Recall that

$$E \left[ \log \left( \frac{\theta_i}{1 - \theta_i} \right) \right] = \mu_i, \quad \forall i = 1, ..., m. \quad (3.77)$$

Therefore, having $\mu_{j1} = -0.5$ and $\mu_{j2} = -1$ results in

$$E \left[ \log \left( \frac{\theta_{j1}}{1 - \theta_{j1}} \right) \right] \approx \log \left[ \frac{E(\theta_{j1})}{1 - E(\theta_{j1})} \right] = -0.5,$$

$$\implies E(\theta_{j1}) \approx \frac{e^{-0.5}}{1 + e^{-0.5}} \approx 0.38,$$

and $E(\theta_{j2}) \approx \frac{e^{-1}}{1 + e^{-1}} \approx 0.27. \quad (3.78)$

The same thing is for $\mu_{k1} = 0.5$ and $\mu_{k2} = 1$, where

$$E(\theta_{k1}) \approx \frac{e^{0.5}}{1 + e^{0.5}} \approx 0.62,$$

and $E(\theta_{k2}) \approx \frac{e^{1}}{1 + e^{1}} \approx 0.73. \quad (3.79)$

Values in (3.78) and (3.79) are located on both sides of $\theta = 0.5$ (median). Tables D.7, D.8 and D.9 show that increasing the scale parameter values in IBLnM results in producing shorter interval estimates for $\beta_0$ and $\beta_1$.

Figure D.2 shows ImpLogit plots under IBLnM. The impact of location and scale parameters on the produced imprecision under IBLnM will be paid attention in Chapter 4. Section 3.7 finalizes this chapter by fitting ImpLogit model under IDM with a small sample size.

### 3.7 ImpLogit with Smaller Sample

A re-look at tables 3.1, 3.2 and 3.3 for data in Table 2.1 shows that interval estimates lengths are near to each other under different values of $\nu$ and different selections of $\theta_j$ and $\theta_k$. This can be due to having a total number of binary observations over 6 binomial dose levels as $49+50+48+46+49+50 = 292$. Therefore, this sample size will be decreased by taking a smaller sample from data given in Table 2.1. To take such a small sample, a random variable say $u \sim$ bernoulli $(0.1, 1)$, will be used to decide whether to select each
binary observation out of 292 observations. The main goal behind searching 2-parameter 
ImpLogit behaviour with smaller sample is to see if imprecision of regression parameters 
estimates can be impacted. Table 3.13 gives the smaller sample.

Fitting 2-parameter increasing ImpLogit model under IDM with this small sample 
results in a longer interval estimates as being seen in tables D.10, D.11 and D.12 in 
comparison to tables 3.1, 3.2 and 3.3. The total sum of of binary trials in Table 3.13 is 
37.

Table 3.13: Smaller size binomial data sample taken from the population data in 
Table 2.1. Dose is given in mg/l, \( n \) is the number of experimented insects and \( y_{obs} \) 
is the number of affected insects.

<table>
<thead>
<tr>
<th>Dose</th>
<th>( n )</th>
<th>( y_{obs} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2.6</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>3.8</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>5.1</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>7.7</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>10.2</td>
<td>7</td>
</tr>
</tbody>
</table>
Chapter 4
Simulation Study

4.1 Motivation for the Simulation Study

A simulation study is necessary to have a deeper knowledge of 2-parameter ImpLogit model. Several points of interest can arise and require more investigation. For example, the sample size is an important factor that relates to the imprecision of $\beta_0$ and $\beta_1$ estimates. The number of binomial observations ($m$) is a point that deserves to be paid attention.

Furthermore, the locations of the covariate $x_i$ distinct values can play a strong role in increasing or decreasing the imprecision of parameters of interest. This comes by searching how to shorten the interval estimates of regression parameters. A simulation study can search the conditions under which the true values of regression parameters do belong to the produced interval estimates. The location and distance among the selected values of the covariate $x_i$ have to considered. It is important to know how far the selected values of $x_i$ should be from each other in order to control the lengths of interval estimates.

The proposed simulation study in this chapter studies differences in ImpLogit fit and performance under both of IDM and IBLnM imprecise priors. It is interesting to study the effect of changing the imprecise priors parameters on the produced imprecision of regression parameters. Also, the relation between the values of parameters in imprecise priors and lengths of the produced interval estimates is important to be figured out.

To start the simulation study, let $\beta_0 = 0$ and $\beta_1 = 0.4$. Having true $\beta_0 = 0$ makes the true median of the logistic distribution to be

$$\log \left( \frac{0.5}{1 - 0.5} \right) = \beta_0 + x_i (\beta_1) = (0.4)x_i \implies x_{\text{median}} = 0.$$
Also, having $\beta_1 = 0.4$ makes the true $\theta$ at $x_i = -1$ and $x_i = 1$ as

$$
\begin{align*}
\theta_i = \theta (x_i = -1) &= \frac{e^{0+0.4(-1)}}{1 + e^{0+0.4(-1)}} = \frac{e^{-0.4}}{1 + e^{-0.4}} \approx 0.4,
\quad \text{and} \\
\theta_i = \theta (x_i = 1) &= \frac{e^{0+0.4(1)}}{1 + e^{0+0.4(1)}} = \frac{e^{0.4}}{1 + e^{0.4}} \approx 0.6.
\end{align*}
$$

So, the true $\theta$ changes within 0.4 to 0.6 over covariate range of -1 to 1, where the true median value is at $x_i = 0$. In other words, the $x_i$’s can be scaled such that $x_i = 0$ corresponds to a response of 0.5 and $x_i = -1$ and 1 correspond to responses of 0.4 and 0.6, respectively.

For the simulated data, each binomial observation $y_i$ comes from

$$
y_i \sim \text{bin} \left( n_i, \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \right), \quad \beta_0 = 0, \quad \beta_1 = 0.4, \quad i = 1, \ldots, m,
$$

$$
n_1 = n_2 = \ldots = n_m,
$$

where the design is balanced and the total number of trials over all binomial observations is fixed as $\sum_{i=1}^{m} n_i = 40$. Three combinations of $n_i$ and $m$ are chosen, $n_i = 20$ with $m = 2$, $n_i = 4$ with $m = 10$ and $n_i = 2$ with $m = 20$, such that $n_i \times m = 40$, $\forall i$. .

Since having true $\beta_1 = 0.4$ refers to an increasing logistic function, then the induced prior distribution under IDM is modelled as in (2.11) in Section 2.1. This also requires to use the correlation structure given in (2.23) in the induced prior distribution under IBLnM.

The following sections will present simulations under different designs. Sections 4.2, 4.3 and 4.4 describe simulating ImpLogit model with two covariate fixed values, say $x_1$ and $x_2$, on different locations from $x = 0$. Having only two fixed covariate values leads to have only one selected pair of $\theta_j$ and $\theta_k$ at $x_1$ and $x_2$, respectively. Both of $x_1$ and $x_2$ will be selected to surround or to be on the right and left sides of the true median value. Fixing only two values allows to search the relation between the selected distance between $x_1$ and $x_2$ values and the produced interval estimates. Sections 4.5, 4.6 and 4.7 are for ImpLogit simulation with multiple covariate fixed values, $x_1, \ldots, x_m$.

For each design, ImpLogit model will be fitted over 2000 generated samples. Having 2000 generated samples will result in computing 2000 single interval estimates for $\beta_0$ and $\beta_1$. Averages of 2000 lower estimates and 2000 upper estimates will be calculated to find
what will be called as the averaged interval estimate. From here and on, an averaged interval estimate is abbreviated by AIE.

The simulation study results will be plotted graphically from page 108 to page 119 for the case of two fixed covariate values and pages 120 and 129 for the case of multiple fixed values.

4.2 Symmetric Two Values

In sections 4.2, 4.3 and 4.4, the covariate $x_i$ will be fixed at the following cases,

$$
\begin{align*}
  x_1 = & -0.5 \ and \ x_2 = 0.5, \\
  x_1 = & -1 \ and \ x_2 = 1, \\
  x_1 = & -2.5 \ and \ x_2 = 2.5,
\end{align*}
$$

(4.4)

Start with $x_1 = -0.5$ and $x_2 = 0.5$ so that they are located on similar distance from $x = 0$. Table 4.2 (page 104) gives the simulation computations obtained for lower and upper estimates of regression parameters.

Distance between $x_1$ and $x_2$ can be lengthened in order to investigate it’s impact on AIEs. Tables 4.3 and 4.4 give computational results when $x_1 = -1$, $x_2 = 1$ and $x_1 = -2.5$, $x_2 = 2.5$, respectively. The AIEs in tables 4.2, 4.3 and 4.4 are plotted in figures 4.1 and 4.2 for $\beta_0$ and $\beta_1$, respectively. Lower and upper estimates are solidly marked by a closed circle $\bullet$ when $x_1 = -0.5$ and $x_2 = 0.5$, a closed square $\blacksquare$ when $x_1 = -1$ and $x_2 = 1$ and a closed triangle $\blacktriangle$ when $x_1 = -2.5$ and $x_2 = 2.5$. The true values of $\beta_0$ and $\beta_1$ are always marked by “×”.

Figures 4.1 and 4.2 provide useful information about the behaviour of AIEs of $\beta_0$ and $\beta_1$. Both of $H_j$ and $H_k$ are assigned values that were used in (3.57) and (3.58). The top plots in both figures are under IDM where $\nu = 1$ and $\nu = 2$, respectively. This aims to navigate the impact of changing the value of $\nu$ on the AIEs lengths of regression parameters.

The top-left plots of both figures, where $\nu = 1$, show that true values of $\beta_0$ and $\beta_1$ are falling inside AIEs for all choices of $x_i$. A similar thing happens in the top-right plot when $\nu = 2$. Increasing $\nu$ does not necessarily result in a noticeable difference in the AIEs.
of $\beta_0$ in terms of length and values of lower and upper estimates. A bigger difference can be seen for $\beta_1$ in Figure 4.2. AIEs of $\beta_1$ in the top-right plot are looking longer than the corresponding ones in the top-left plot. One more important thing that a shorter distance between $x_1$ and $x_2$ produces longer interval estimates which can clearly be noticed in top plots of both figures. This indicates, tentatively, that a shorter distance between $x_1$ and $x_2$ results in more imprecision.

For each single (not averaged) interval estimate, the lower value can be plotted versus the upper value. Figure 4.3 gives 6 different plots in which the left column corresponds to the top-left plot of Figure 4.1 while the right column is for the top-right plot. Each plot in Figure 4.3 includes 2000 points. Each point locates above the dashed diagonal line because the lower estimate must be less than the upper estimate in any single interval estimate. Left and right columns of Figure 4.4 give plots that correspond to the AIEs in top-left and top-right plots of Figure 4.2, respectively.

In each plot of figures 4.3 and 4.4, the diagonal, horizontal and vertical lines divide the dotted area into three smaller areas. The smaller areas look like a square and two equal triangles. The proportion of points out of all points in the smaller square in each plot is given as percentage value in top plots of figures 4.1 and 4.2. Any point in that square represents a single interval estimate that includes the true regression parameter value.

Percentage values in top plots are equal or higher than 95% and 90% for estimates of $\beta_0$ and $\beta_1$, respectively, which matches with plots in figures 4.3 and 4.4 where few points are located in the top-right and bottom-left triangles. In other words, few cases of AIEs with extreme near high or low values of lower and upper estimates were found because most points are in the top-left square. This means that true values of $\beta_0$ and $\beta_1$ were included in greater than or equal to 95% and 90% out of all AIEs of $\beta_0$ and $\beta_1$, respectively, under IDM. This can be noticed in top plots of figures 4.1 and 4.2 where “×” is almost at the middle of all AIEs. Percentage values will be presented in all figures of this simulation study.

In middle-left plots of figures 4.1 and 4.2, the fixed prior values of location and scale
parameters in IBLnM are

\[ \mu_j \in \{-0.35, -0.45\}, \mu_k \in \{0.35, 0.45\}, \]  

(4.5)

and

\[ \sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}, \]  

(4.6)

while the middle-right plot shows AIEs when

\[ \mu_j \in \{-0.5, -1\}, \mu_k \in \{0.5, 1\}, \]  

(4.7)

with similar values of \( \sigma_j \) and \( \sigma_k \) as in (4.6). Having the location parameters values as in (4.5) and (4.7) means that

\[
\begin{align*}
\text{if } \mu_i = -0.35 & \implies \mathbb{E}(\theta_i) \approx \frac{e^{-0.35}}{1 + e^{-0.35}} \approx 0.41, \\
\text{if } \mu_i = -0.45 & \implies \mathbb{E}(\theta_i) \approx \frac{e^{-0.45}}{1 + e^{-0.45}} \approx 0.39, \\
\text{if } \mu_i = 0.35 & \implies \mathbb{E}(\theta_i) \approx \frac{e^{0.35}}{1 + e^{0.35}} \approx 0.59, \\
\text{if } \mu_i = 0.45 & \implies \mathbb{E}(\theta_i) \approx \frac{e^{0.45}}{1 + e^{0.45}} \approx 0.61,
\end{align*}
\]

(4.8)

which means that, in parallel with (3.78), (3.79) and (4.8), the prior approximate belief about \( \theta \) ranges between 0.27 and 0.73. This prior belief of \( \theta \) range can be compared to the case of having \( x_1 = -2.5 \) and \( x_2 = 2.5 \) from (4.4), provided that \( \beta_0 = 0, \beta_1 = 0.4 \), as

\[
\begin{align*}
\theta_i = \theta (x_i = -2.5) &= \frac{e^{0+0.4(-2.5)}}{1 + e^{0+0.4(-2.5)}} = \frac{e^{-1}}{1 + e^{-1}} \approx 0.27, \\
\text{and } \theta_i &= \theta (x_i = 2.5) = \frac{e^{0+0.4(2.5)}}{1 + e^{0+0.4(2.5)}} = \frac{e}{1 + e} \approx 0.73,
\end{align*}
\]

(4.9)

which shows the closeness of both ranges.

Middle plots of figures 4.1 and 4.2 show that all AIEs included the true values of \( \beta_0 \) and \( \beta_1 \), with less noticeable difference in lengths of AIEs of \( \beta_0 \). The imprecise model behaviour, in terms of AIEs lengths, changes in the middle plots of Figure 4.2 for \( \beta_1 \). In Figure 4.2, imprecision of \( \beta_1 \) is impacted and reduced under longer distance between \( x_1 \) and \( x_2 \) which is the same in top plots. The impact of covariate values range on lengths of AIEs is less on \( \beta_0 \) in comparison to \( \beta_1 \).

The bottom plots in both figures are for IBLnM with values of \( \mu_j \) and \( \mu_k \) similar to those in the middle plots but with larger values of \( \sigma_j \) and \( \sigma_k \). The scaling prior parameters
are fixed as

$$\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}.$$  (4.10)

In bottom plots, larger $\sigma_j$ and $\sigma_k$ lead to shorter AIEs in $\beta_0$ and $\beta_1$, but with stronger impact on $\beta_1$. The scale parameters $\sigma_j$ and $\sigma_k$ in IBLnM tend to play a strong role as for $\nu$ in IDM. The true values of both of $\beta_0$ and $\beta_1$ are always included in the AIEs which is quite visibly seen in figures 4.1 and 4.2. The relation between lengths of AIEs and the distance between $x_1$ and $x_2$ agrees with the tentative results found in the top and middle plots in both figures. Generally speaking, a shorter distance between $x_1$ and $x_2$ does not necessarily produce less imprecision.

Increasing the difference between $\mu_j$ and $\mu_k$ results in changing the lengths of the AIEs. Figures 4.1 and 4.2 acknowledge that increasing $\sigma_j$ and $\sigma_k$ shortens the AIEs, while changing $\mu_j$ farther to the left and $\mu_k$ to the right, lengthens the AIEs.

This section summarizes as the following: imprecision of regression parameters estimates in ImpLogit model is increased if the distance between $x_1$ and $x_2$ is shortened. If IDM is applied, a larger $\nu$ results in longer AIEs. If IBLnM is considered as an imprecise prior, then smaller $\sigma_j$ and $\sigma_k$ lead to have longer AIEs. Dragging $\mu_j$ and $\mu_k$ farther from each other makes longer AIEs. If $x_i$’s is symmetrically allocated around the true median value, then true values of $\beta_0$ and $\beta_1$ are always included in the AIEs under both imprecise priors.

It is important to note there is no change in the values of $\nu$, $\mu_i$ and $\sigma_i$ among all figures in the rest of this simulation study. Section 4.3 suggests to drag both covariate fixed values to the right and left sides of $x = 0$.

### 4.3 Right and Left-Shifted Two Values

The symmetric design is not always the case where $x_1$ and $x_2$ can be allocated to the right or left of the true median $x = 0$. Figures 4.5 and 4.6 present simulated AIEs when $x_1 = 1$, $x_2 = 2, 3$ and 4. Right-shifted $x_1$ and $x_2$ design gives AIEs for $\beta_0$ that include its true value as being seen in Figure 4.5. The only thing that can be noticed is the change of “×” location on the AIEs plots, where “×” tends to move to the right side of each AIE.
Figure 4.6 shows that \( \beta_1 \) true value is always included in the AIEs. The imprecision produced in case of \( x_1 = 1 \) and \( x_2 = 2 \) is larger than the case of \( x_1 = 1 \) and \( x_2 = 3 \) or 4. That is, a shorter range results in larger imprecision even for right-shifted covariate values. This does not conflict with Section 4.2 regarding the relation between the lengths of the AIEs and the distance between \( x_1 \) and \( x_2 \). AIEs of \( \beta_1 \) are sometimes longer compared to those found in the symmetric design. Therefore, shifting \( x_1 \) and \( x_2 \) may create longer AIEs of \( \beta_1 \) which stands against the main goal of fitting and using ImpLogit model.

Increasing \( \nu \) in the top-right plot of Figure 4.6 under IDM makes longer AIEs for \( \beta_1 \). It can be seen by looking at top plots in both figures that increasing \( \nu \) has stronger impact on \( \beta_1 \) than \( \beta_0 \). Despite that imprecision is less when \( x_1 = 1 \) and \( x_2 = 3 \) or 4, the AIEs included \( \beta_0 \) and \( \beta_1 \) true values in all plots.

Comparing the middle and bottom plots of Figure 4.6 shows that larger values of \( \sigma_j \) and \( \sigma_k \) result in shorter AIEs for \( \beta_1 \). Enlarging \( \sigma_j \) and \( \sigma_k \) may and may not decrease imprecision of \( \beta_0 \) in Figure 4.5 which is different from the previous design where AIEs lengths in \( \beta_0 \) and \( \beta_1 \) were decreased with larger \( \sigma_j \) and \( \sigma_k \). Shifting the location parameters farther to both sides of \( \mu = 0 \) lengthens AIEs which matches with the case of symmetric located \( x_1 \) and \( x_2 \). Again, scale parameters in IBLnM prior are playing a strong role in lengthening or shortening the AIEs for \( \beta_1 \).

ImpLogit model with left-shifted covariate values is taken into account. The covariate selections are \( x_1 = -2, -3, -4 \) and \( x_2 = -1 \). Plots of AIEs are given in figures 4.7 and 4.8. In both figures, ImpLogit model behaviour is quite similar to the right-shifted case. Left-shifted design matches with both of symmetric and right-shifted ones in letting a wider distance between \( x_1 \) and \( x_2 \) to reduce the produced imprecision. True values of both of \( \beta_1 \) and \( \beta_0 \) are always included in the AIEs. The true value of \( \beta_0 \) marked by “×” tends to move toward the left side of each AIE plot in Figure 4.7 (it is to the right side in Figure 4.5).

### 4.4 Extremely Shifted Two Values

Both of \( x_1 \) and \( x_2 \) are now suggested to be located farther on both sides of \( x = 0 \). For the right side, there is \( x_1 = 3.5 \), \( x_2 = 4.5 \) and \( x_1 = 3 \), \( x_2 = 5 \) where the AIEs plots are
given in figures 4.9 and 4.10. This new situation finds that true $\beta_0$ fails to belong to any of the computed AIEs under both of IDM and IBLnM as being figured out in Figure 4.9. If $x_1 = 3.5$ and $x_2 = 4.5$, then longer AIEs are produced in comparison to those when $x_1 = 3$ and $x_2 = 5$. This happens in all plots of figures 4.9 and 4.10. This gives a stronger evidence that, with two fixed covariate values, longer AIEs come with shorter fixed ranges. In Figure 4.10, $\beta_1$ true value is included in the AIEs only if $x_1 = 3.5$ and $x_2 = 4.5$.

The last attempt with two fixed covariate values is by considering $x_1 = -4.5$, $x_2 = -3.5$ and $x_1 = -5$, $x_2 = -3$. Having $x_1 = -4.5$ and $x_2 = -3.5$ produces longer AIEs as being seen in figures 4.11 and 4.12 than if $x_1 = -5$, $x_2 = -3$. True $\beta_0$ is out of all AIEs in Figure 4.11. This is similar to the case of extreme right-shifted values. AIEs of $\beta_1$ do include the true value only when $x_1 = -4.5$, $x_2 = -3.5$ but not if $x_1 = -5$, $x_2 = -3$ which can be seen in Figure 4.12. Generally, with farther locations of $x_1$ and $x_2$ (not around $x = 0$ or near to $x = 0$ from both sides), true values of both of $\beta_0$ and $\beta_1$ tend to escape from the corresponding AIEs.

The impact of $\nu$ over the current covariate locations does not change from that in the previous two sections. Larger $\nu$ gives longer AIEs under IDM. Less values of scale parameters and smaller $\mu_j$ with larger $\mu_k$ under IBLnM give longer AIEs.

The simulation study can be extended and enhanced by fixing more than two values for $x_i$. Section 4.5 starts by fixing multiple values around $x = 0$ symmetrically.
4.5 Symmetric Covariate Multiple Values

To simulate ImpLogit under multiple values of the covariate $x_i$, the following fixed values are kept: $\beta_0 = 0$ and $\beta_1 = 0.4$, $m \times n_i = 40$, $n = n_i$, $\forall i = 1, \ldots, m$, but $m \in \{5, 10, 20\}$ instead of $m \in \{2, 10, 20\}$. The covariate ranges are $x_i \in [-0.5, 0.5]$, $x_i \in [-1, 1]$ and $x_i \in [-2.5, 2.5]$. Table 4.5 presents the multiple covariate values. They are selected to surround $x = 0$ in order to have symmetrical circulation around the true median value.

If $m = 5$, the selected binomial parameters that induce the regression parameters prior distribution are $\theta_2$ and $\theta_4$ (corresponding to $x_2$ and $x_4$). The selected parameters become $\theta_3$ and $\theta_8$ when $m = 10$ and $\theta_5$ and $\theta_{16}$ if $m = 20$. The given selections do not change in sections 4.6 and 4.7.

The AIEs are plotted in figures 4.13 and 4.14. Having multiple covariate values finds that short ranges of $x_i$ produce longer AIEs. The true parameters values are always falling within the AIEs for all plots and all values of $m$. Larger $\nu$ and smaller $\sigma_j$ and $\sigma_k$ result in longer AIEs. Simulation results under multiple covariate values don’t show a difference from Section 4.2. The main new thing is that AIEs are longer when $m = 10$ or 20.

4.6 Right and Left-Shifted Multiple Values

Figures 4.15 and 4.16 give plots when $x_i \in [1, 2]$, $x_i \in [1, 3]$ and $x_i \in [1, 4]$. AIEs always include true $\beta_0$ and $\beta_1$. Figures 4.15 and 4.16 shows that enlarging $m$ results in longer AIEs. Note that $\beta_0$ true value is located to the right side of each AIE in Figure 4.15 which has happened in Figure 4.5. Longer AIEs are obtained under $x_i \in [1, 2]$ for both of $\beta_0$ and $\beta_1$.

Figures 4.17 and 4.18 are for left-shifted covariate multiple values and the selected covariate ranges are $x_i \in [-4, -1]$, $x_i \in [-3, -1]$ and $x_i \in [-2, -1]$. Having $x_i \in [-2, -1]$ produced longer AIEs for both of $\beta_0$ and $\beta_1$ than the other two cases of covariate values ranges. True values of regression parameters are always included in AIEs. Larger $m$ gives longer AIEs.
4.7 Extremely Shifted Multiple Values

Figures 4.19 and 4.20 are with \( x_i \in [3.5, 4.5] \) and \( x_i \in [3, 5] \). In both figures, true \( \beta_0 \) falls outside all AIEs. Larger \( m \) gives longer AIEs for both of \( \beta_0 \) and \( \beta_1 \). The true \( \beta_1 \) is out of AIEs when \( x_i \in [3, 5] \).

Figures 4.21 and 4.22 provide plots when \( x_i \in [-4.5, -3.5] \) and \( x_i \in [-5, -3] \). Having \( x_i \in [-4.5, -3.5] \) results in longer AIEs than if \( x_i \in [-5, -3] \). It is important to see that \( “x” \) is out of AIEs of \( \beta_0 \) in all plots. Figure 4.22 shows that true \( \beta_1 \) is not included when \( x_i \in [-5, -3] \).

4.8 Closer \( \theta_j \) and \( \theta_k \)

The relation between the closeness of the sequential orders of the selected pair \( \theta_j \) and \( \theta_k \) and the imprecision of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) can be investigated. For example, if \( m = 20, n_i = 2, \forall i \), the question is : does ImpLogit fit give different results when the selected pair is \( \theta_5 = \theta(x_5) \) and \( \theta_{16} = \theta(x_{16}) \) than if it is \( \theta_{10} = \theta(x_{10}) \) and \( \theta_{11} = \theta(x_{11}) \)? In other words : can the selected priors retard the role of \( x_i \)’s values range on including or firing the true regression parameters values from their AIEs?

To answer the last question, figures E.1 and E.2 (in Appendix E) present plots of AIEs under the symmetric design when \( m = 20 \) and \( n_i = 2, \forall i \) and different selected \( \theta_j \) and \( \theta_k \). The \( x_i \)’s values are similar to those in Table 4.5 when \( m = 20 \). That is, Figure E.1 corresponds to Figure 4.13, while Figure E.2 corresponds to Figure 4.14 with similar corresponding values of \( \nu \) in IDM and \( \mu_j, \mu_k, \sigma_j \) and \( \sigma_k \) in IBLnM. The difference in figures is simplified as : Figure 4.13 shows AIEs taking into account different values of \( m \), while Figure E.1 shows AIEs with fixed \( m = 20 \) and selections of \( \theta_5 \) and \( \theta_{16}, \theta_8 \) and \( \theta_{13} \) and finally \( \theta_{10} \) and \( \theta_{11} \). Therefore, both figures are having identical AIEs only when \( m = 20 \) and \( \theta_5 \) and \( \theta_{16} \) (given in black in the top three AIEs of each plot). The same situation can be seen in figures E.2 and 4.14. It is important to note that graphs of this section and sections 4.9 and 4.10 are given in Appendix E.

Selecting different pairs of \( \theta_j \) and \( \theta_k \) does not show a clear difference in AIEs lengths under the symmetric design. The location of the true parameters values kept being inside
all computed AIEs, but the proportion of single interval estimates that include true values of \( \beta_0 \) and \( \beta_1 \) is increased compared to percentage values in figures 4.13 and 4.14. The same thing is applied for extreme \( x_i \)'s where figures E.3 and E.4 correspond to figures 4.19 and 4.20, respectively. In Figure E.3, selecting \( \theta_8, \theta_{13} \) or \( \theta_{10}, \theta_{11} \) does not result in true \( \beta_0 \) to fall in any of the computed AIEs. The main thing is that the proportion of single interval estimates that include true \( \beta_0 \) is increased. AIEs with \( \theta_5 \) and \( \theta_{16} \) are similar to those in Figure 4.19. Selecting \( \theta_8, \theta_{13} \) or \( \theta_{10}, \theta_{11} \) makes true \( \beta_1 \) nearer to the AIEs but not brought inside any AIE when \( x_i \in [3,5] \).

### 4.9 Optimal Frequentist Design

The study of \( x_i \)'s values design in ImpLogit model can be developed using the Fisher information matrix (FIM). This enables to find the \( x_i \)'s at which the asymptotic variances of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) are taken to their minimum values. The main goal is to build a design from FIM and compare it to the previous ImpLogit designs. The FIM of \( \beta_0 \) and \( \beta_1 \) is given as

\[
\begin{bmatrix}
\sum_{i=1}^{m} n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) \\
\sum_{i=1}^{m} x_i n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)
\end{bmatrix}
\]

The inverse of (4.11) has to be found in order to find approximate estimates of the asymptotic variances of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). The determinant of (4.11) is

\[
\begin{bmatrix}
\sum_{i=1}^{m} n_i \frac{e^{0.4x_i}}{1 + e^{0.4x_i}} \left( \frac{1}{1 + e^{0.4x_i}} \right) \\
\sum_{i=1}^{m} x_i n_i \frac{e^{0.4x_i}}{1 + e^{0.4x_i}} \left( \frac{1}{1 + e^{0.4x_i}} \right)
\end{bmatrix}
\]
\[ D(x) = \left[ \sum_{i=1}^{m} n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \right] \left[ \sum_{i=1}^{m} n_i x_i^2 \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \right] - \left[ \sum_{i=1}^{m} x_i n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \right]^2, \quad (4.13) \]

\[ \Rightarrow \quad I^{-1} = \left[ \begin{array}{cc} \sum_{i=1}^{m} n_i x_i^2 \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} & \sum_{i=1}^{m} x_i n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \\ \frac{\sum_{i=1}^{m} x_i n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, ..., x_m)} & \frac{\sum_{i=1}^{m} n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, ..., x_m)} \end{array} \right] \right]. \quad (4.14) \]

Now, \( x_i \)'s in (4.14) are considered as variables, then derivatives with respect to \( x_i \) are taken and equated to zero value to be solved numerically as

\[ \frac{\partial}{\partial x_i} \left[ \sum_{i=1}^{m} n_i x_i^2 \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2} \right] \left[ \frac{\sum_{i=1}^{m} x_i n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, ..., x_m)} \right] = 0, \quad (4.15) \]

and

\[ \frac{\partial}{\partial x_i} \left[ \frac{\sum_{i=1}^{m} n_i \frac{e^{0.4x_i}}{(1 + e^{0.4x_i})^2}}{D(x_1, ..., x_m)} \right] = 0. \quad (4.16) \]

Maximizing the determinant function in (4.13) or solving any of equations (4.15) and (4.16) can give \( x_i \)'s values that optimize the estimates of the predictive probabilities \( \theta_i \), \( \forall i \). The covariate values that come from maximizing (4.13) or solving (4.15) or (4.16) are not necessarily the same. The values can be compared to \( x_i \)'s ranges in ImpLogit design. In other way, the main question is : do such \( x_i \)'s values (provided that \( \beta_0 = 0 \) and \( \beta_1 = 0.4 \)) match with those that reduce imprecision in ImpLogit model according to previous ImpLogit designs?

To answer the last question : Table 4.1 presents the computed values of \( D(x) \) in (4.13) under different \( x_i \)'s values. Larger values of the determinant function come when
$x_1 = -2.5$, $x_2 = 2.5$ or $x_i \in [-2.5, 2.5]$ while values of $D(x)$ when $x_1 = -0.5$, $x_2 = 0.5$ or $x_i \in [-0.5, 0.5]$ were the smallest. This matches with previous ImpLogit designs results, because a covariate range with larger $D(x)$ in Table 4.1 corresponds to a covariate range in ImpLogit model with less imprecision in regression parameters estimates. Table 4.1 shows that increasing $m$ decreases $D(x)$ value which means that having larger $m$ produces longer interval estimates of $\beta_0$ and $\beta_1$. This is not different from what was found in the previous sections of this chapter.

**Table 4.1:** Computed $D(x)$ values using (4.13) under the symmetric design provided that $\sum_{i=1}^{m} n_i = 40$.

<table>
<thead>
<tr>
<th>Two $x_i$'s</th>
<th>Multiple $x_i$'s</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2$</td>
<td>$m = 5$</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>$m = 20$</td>
</tr>
<tr>
<td>$x_1 = -0.5$, $x_2 = 0.5$</td>
<td>24.5</td>
</tr>
<tr>
<td>$x_1 = -1$, $x_2 = 1$</td>
<td>92.4</td>
</tr>
<tr>
<td>$x_1 = -2.5$, $x_2 = 2.5$</td>
<td>386.6</td>
</tr>
<tr>
<td>$x_i \in [-0.5, 0.5]$</td>
<td>10.3</td>
</tr>
<tr>
<td>$x_i \in [-1, 1]$</td>
<td>36.9</td>
</tr>
<tr>
<td>$x_i \in [-2.5, 2.5]$</td>
<td>8.6</td>
</tr>
</tbody>
</table>

The maximum value of the determinant function in (4.13) occurs when $x_1 = -3.85$ and $x_2 = 3.85$. This covariate range includes the range of $x_1 = -2.5$, $x_2 = 2.5$ (that produced the shortest interval estimates). Therefore, it is interesting to fit ImpLogit model when $x_1 = -3.85$ and $x_2 = 3.85$ to see the reduction of imprecision in regression parameters estimates. Figures E.5 and E.6 add to Figures 4.1 and 4.2 plots of a fourth interval estimate under the range of $x_1 = -3.85$ and $x_2 = 3.85$. The ends of the new interval estimate plots are marked by ▽. Other plots of interval estimates (marked by ●, ■ or ▲) are copied from Figures 4.1 and 4.2. Figures E.5 and E.6 show that under $x_1 = -3.85$ and $x_2 = 3.85$, the interval estimates are shorter for both of $\beta_0$ and $\beta_1$ but with a stronger impact on $\beta_1$.

For $m = 2$, solving (4.15) gives $x_i$'s values at which the approximate asymptotic variance of $\hat{\beta}_0$ is minimized, in the inverse of FIM, ∀$i$, to be at $x_1 = x_2 = 0$. The approximate asymptotic variance of $\hat{\beta}_1$ takes its minimum value at $x_1 = -5.998$ and $x_2 = 5.998$ after solving (4.16). This covariate range includes $x_1 = -2.5$, $x_2 = 2.5$. 

101
4.10 Multiple Parameter ImpLogit Model

ImpLogit model can be extended and developed by considering multiple parameters, \( \beta_0, \beta_1, ..., \beta_p \). \((P+1)\)-parameter ImpLogit model gives the opportunity to study the effects of \( p \) covariates on the produced imprecision of \( p + 1 \) regression parameters. If

\[
\theta_i = \theta(x_i) = \frac{e^{x_i' \beta}}{1 + e^{x_i' \beta}} = \frac{e^{\beta_0 + x_i' \beta_1 + x_i' \beta_p}}{1 + e^{\beta_0 + x_i' \beta_1 + x_i' \beta_p}},
\]

then the prior distribution of \( \theta = [\theta_j, \theta_k, ..., \theta_l] \) is defined as

\[
f(\theta_j, \theta_k, ..., \theta_l) = w f_1(\theta_j, \theta_k, ..., \theta_l) + (1 - w) f_2(\theta_j, \theta_k, ..., \theta_l), \quad w \in [0, 1].
\]

To induce a prior distribution for \( \beta = [\beta_0, ..., \beta_p] \), the Jacobian determinant \( \frac{\partial \theta}{\partial \beta} \) is

\[
\begin{vmatrix}
\frac{\partial \theta_j}{\partial \beta_0} & \frac{\partial \theta_j}{\partial \beta_1} & \ldots & \frac{\partial \theta_j}{\partial \beta_p} \\
\frac{\partial \theta_k}{\partial \beta_0} & \frac{\partial \theta_k}{\partial \beta_1} & \ldots & \frac{\partial \theta_k}{\partial \beta_p} \\
\frac{\partial \theta_l}{\partial \beta_0} & \frac{\partial \theta_l}{\partial \beta_1} & \ldots & \frac{\partial \theta_l}{\partial \beta_p}
\end{vmatrix}
\]

\[
= \begin{vmatrix}
\frac{e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} & \frac{x_i e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} & \ldots & \frac{x_i e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} \\
\frac{e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} & \frac{x_i e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} & \ldots & \frac{x_i e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} \\
\frac{e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} & \frac{x_i e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2} & \ldots & \frac{x_i e^{x_i' \beta}}{(1 + e^{x_i' \beta})^2}
\end{vmatrix}.
\]

For example, take the case of 3 regression parameters, \( \beta_0, \beta_1 \) and \( \beta_2 \). There is

\[
\theta_i = \frac{e^{\beta_0 + x_i' \beta_1 + x_i' \beta_2}}{1 + e^{\beta_0 + x_i' \beta_1 + x_i' \beta_2}}, \quad -\infty < \beta_j < \infty, \quad j = 0, 1, 2,
\]

\[-\infty < x_{ij} < \infty, \quad i = 1, ..., m, \quad j = 1, 2. \]

Using the aggregation property of the Dirichlet distribution gives the joint prior distribution for \( f_1 (\theta_j, \theta_k, \theta_l) \) as

\[
(\theta_j, \theta_k - \theta_j, \theta_l - \theta_k, 1 - \theta_l) \sim \text{Dir}(\nu H_j, \nu H_k - \nu H_j, \nu H_l - \nu H_k, \nu - \nu H_l),
\]

and for \( f_2 (\theta_j, \theta_k, \theta_l) \) as

\[
(1 - \theta_j, \theta_k - \theta_j, \theta_l - \theta_k, \theta_l) \sim \text{Dir}(\nu - \nu H_j^*, \nu H_j^* - \nu H_k^*, \nu H_l^* - \nu H_k^*, \nu - \nu H_l^*).
\]
For $f_1(\cdot)$, let $\varphi_{jkl1} = H_j$, $\varphi_{jkl2} = H_k - H_j$, $\varphi_{jkl3} = H_l - H_k$ and $\varphi_{jkl4} = 1 - H_l$. Also, for $f_2(\cdot)$, let $\varphi_{jkl1}^* = 1 - H_j^*$, $\varphi_{jkl2}^* = H_j^* - H_k$, $\varphi_{jkl3}^* = H_k^* - H_l$ and $\varphi_{jkl4}^* = H_l^*$. If $x_{i1} = x_{i2}$, $\forall i = 1, \ldots, m$, or $x_{i1} \leq x_{i(i+1)1}$ and $x_{i2} \leq x_{i(i+1)2}$, $\forall i = 1, \ldots, m - 1$, then $\theta_j$, $\theta_k$, $\theta_l$ are selected based on the sequential order. The prior distribution of $\theta = [\theta_j, \theta_k, \theta_l]^\top$ is

$$f(\theta_j, \theta_k, \theta_l) = w f_1(\theta_j, \theta_k, \theta_l) + (1 - w) f_2(\theta_j, \theta_k, \theta_l), \ w \in [0, 1],$$

where $f_1$ and $f_2$ are densities for the increasing and decreasing logistic functions. For $\beta = [\beta_0, \beta_1, \beta_2]^\top$ and selected $\theta = [\theta_j, \theta_k, \theta_l]^\top$, to find

$$\pi(\beta) = f(\theta(\beta)) \left| \frac{\partial \theta}{\partial \beta} \right|,$$

the jacobian determinant is

$$\left| \begin{array}{ccc}
\frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} & x_{j1} \frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} & x_{j2} \frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} \\
\frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} & \frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} & \frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} \\
\frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} & \frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2} & \frac{e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2}}{(1 + e^{\beta_0 + x_{i1} \beta_1 + x_{i2} \beta_2})^2}
\end{array} \right|.$$
Table 4.2: AIEs with $x_1 = -0.5$ and $x_2 = 0.5$.

<table>
<thead>
<tr>
<th>IDM</th>
<th>$\nu = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2, n = 20$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td></td>
<td>-0.015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IDM</th>
<th>$\nu = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2, n = 20$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td></td>
<td>-0.015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBLnM</th>
<th>$\mu_j \in {-0.35, -0.45}, \mu_k \in {0.35, 0.45}$, $\sigma_j = \sigma_k \in {0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2, n = 20$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td></td>
<td>-0.015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBLnM</th>
<th>$\mu_j \in {-0.5, -1}, \mu_k \in {0.5, 1}$, $\sigma_j = \sigma_k \in {0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2, n = 20$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td></td>
<td>-0.016</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBLnM</th>
<th>$\mu_j \in {-0.35, -0.45}, \mu_k \in {0.35, 0.45}$, $\sigma_j = \sigma_k \in {1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2, n = 20$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td></td>
<td>-0.015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBLnM</th>
<th>$\mu_j \in {-0.5, -1}, \mu_k \in {0.5, 1}$, $\sigma_j = \sigma_k \in {1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m = 2, n = 20$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td></td>
<td>-0.014</td>
</tr>
</tbody>
</table>
Table 4.3: AIEs with $x_1 = -1$ and $x_2 = 1$.

<table>
<thead>
<tr>
<th></th>
<th>$\nu = 1$</th>
<th>$\nu = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IDM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_n$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td>$m = 2, n = 20$</td>
<td>-0.014</td>
<td>0.014</td>
</tr>
<tr>
<td><strong>IBLnM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_n$</td>
<td>$\hat{\beta}_0$</td>
</tr>
<tr>
<td>$m = 2, n = 20$</td>
<td>-0.015</td>
<td>0.015</td>
</tr>
</tbody>
</table>

|                |           |           |
| **IBLnM**      |           |           |
|                | $\mu_j \in \{-0.35, -0.45\}$, $\mu_k \in \{0.35, 0.45\}$, $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$ |           |
|                | $\hat{\beta}_n$ | $\hat{\beta}_0$ | imp ($\hat{\beta}_0$) | $\hat{\beta}_{1,\text{imp}}$ | $\hat{\beta}_1$ | imp ($\hat{\beta}_1$) |
| $m = 2, n = 20$ | -0.014   | 0.016    | 0.030   | 0.363  | 0.443   | 0.08     |
|                |           |           |
| **IBLnM**      |           |           |
|                | $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$ |           |
|                | $\hat{\beta}_n$ | $\hat{\beta}_0$ | imp ($\hat{\beta}_0$) | $\hat{\beta}_{1,\text{imp}}$ | $\hat{\beta}_1$ | imp ($\hat{\beta}_1$) |
| $m = 2, n = 20$ | -0.015   | 0.015    | 0.030   | 0.341  | 0.449   | 0.108    |
|                |           |           |
| **IBLnM**      |           |           |
|                | $\mu_j \in \{-0.35, -0.45\}$, $\mu_k \in \{0.35, 0.45\}$, $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$ |           |
|                | $\hat{\beta}_n$ | $\hat{\beta}_0$ | imp ($\hat{\beta}_0$) | $\hat{\beta}_{1,\text{imp}}$ | $\hat{\beta}_1$ | imp ($\hat{\beta}_1$) |
| $m = 2, n = 20$ | -0.014   | 0.015    | 0.029   | 0.363  | 0.427   | 0.064    |
|                |           |           |
| **IBLnM**      |           |           |
|                | $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48\}$ |           |
|                | $\hat{\beta}_n$ | $\hat{\beta}_0$ | imp ($\hat{\beta}_0$) | $\hat{\beta}_{1,\text{imp}}$ | $\hat{\beta}_1$ | imp ($\hat{\beta}_1$) |
| $m = 2, n = 20$ | -0.013   | 0.014    | 0.027   | 0.358  | 0.436   | 0.078    |
Table 4.4: AIEs with $x_1 = -2.5$ and $x_2 = 2.5$.

<table>
<thead>
<tr>
<th>IDM</th>
<th>$\nu = 1$</th>
<th>$\nu = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 2, n = 20$</td>
<td>$m = 2, n = 20$</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>$\beta_0$</td>
<td>imp ($\hat{\beta}_0$)</td>
</tr>
<tr>
<td>-0.013</td>
<td>0.014</td>
<td>0.027</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBLnM</th>
<th>$\mu_j \in {-0.35, -0.45}, \mu_k \in {0.35, 0.45}$, $\sigma_j = \sigma_k \in {0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 2, n = 20$</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>-0.013</td>
<td>0.015</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBLnM</th>
<th>$\mu_j \in {-0.5, -1}, \mu_k \in {0.5, 1}$, $\sigma_j = \sigma_k \in {0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 2, n = 20$</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>-0.014</td>
<td>0.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IBLnM</th>
<th>$\mu_j \in {-0.5, -1}, \mu_k \in {0.5, 1}$, $\sigma_j = \sigma_k \in {1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 2, n = 20$</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>-0.013</td>
<td>0.014</td>
</tr>
</tbody>
</table>
Table 4.5: Multiple covariate values under different ranges.

<table>
<thead>
<tr>
<th>$m = 5$</th>
<th>$x_i \in [-0.5, 0.5]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1 = -0.5$</td>
<td>$x_2 = -0.25$</td>
</tr>
<tr>
<td>$x_7 = 0.18$</td>
<td>$x_8 = 0.29$</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>$x_i \in [-1, 1]$</td>
</tr>
<tr>
<td>$x_1 = -0.5$</td>
<td>$x_2 = -0.39$</td>
</tr>
<tr>
<td>$x_7 = -0.2$</td>
<td>$x_8 = -0.15$</td>
</tr>
<tr>
<td>$x_{13} = 0.1$</td>
<td>$x_{14} = 0.15$</td>
</tr>
<tr>
<td>$x_{19} = 0.42$</td>
<td>$x_{20} = 0.5$</td>
</tr>
<tr>
<td>$m = 20$</td>
<td>$x_i \in [-2.5, 2.5]$</td>
</tr>
<tr>
<td>$x_1 = -2.5$</td>
<td>$x_2 = -1$</td>
</tr>
<tr>
<td>$x_7 = -1$</td>
<td>$x_8 = -0.9$</td>
</tr>
<tr>
<td>$x_{13} = -0.4$</td>
<td>$x_{14} = -0.3$</td>
</tr>
<tr>
<td>$x_{13} = 0.3$</td>
<td>$x_{14} = 0.4$</td>
</tr>
<tr>
<td>$x_{19} = 0.9$</td>
<td>$x_{20} = 1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m = 5$</th>
<th>$x_i \in [-2.5, 2.5]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1 = -2.5$</td>
<td>$x_2 = -2.25$</td>
</tr>
<tr>
<td>$x_7 = 1$</td>
<td>$x_8 = 1.5$</td>
</tr>
<tr>
<td>$m = 10$</td>
<td>$x_i \in [-2.5, 2.5]$</td>
</tr>
<tr>
<td>$x_1 = -2.5$</td>
<td>$x_2 = -2.25$</td>
</tr>
<tr>
<td>$x_7 = -1$</td>
<td>$x_8 = -0.75$</td>
</tr>
<tr>
<td>$x_{13} = 0.75$</td>
<td>$x_{14} = 0.1$</td>
</tr>
<tr>
<td>$x_{19} = 2.25$</td>
<td>$x_{20} = 2.5$</td>
</tr>
</tbody>
</table>
Figure 4.1: Plots of AIEs of $\beta_0$ ended by • when $x_1 = -0.5$ and $x_2 = 0.5$, ■ when $x_1 = -1$ and $x_2 = 1$ and ▲ when $x_1 = -2.5$ and $x_2 = 2.5$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.2: Plots of AIEs of $\hat{\beta}_1$ ended by $\bullet$ when $x_1 = -0.5$ and $x_2 = 0.5$, $\blacksquare$ when $x_1 = -1$ and $x_2 = 1$ and $\blacktriangle$ when $x_1 = -2.5$ and $x_2 = 2.5$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.3: Left-column: Plots of lower vs upper estimates of $\beta_0$ under symmetric two $x_i$’s design with $\nu = 1$, $m = 2$ and $n_1 = n_2 = 20$, that correspond to the AIEs in the top-left plot of Figure 4.1. Right-column: For the top-right plot of Figure 4.1 with $\nu = 2$, $m = 2$ and $n_1 = n_2 = 20$. 

[Diagram of plots showing lower vs upper estimates of $\beta_0$ with different parameter values]
Figure 4.4: Left-column: Plots of lower vs upper estimates of $\beta_1$ under symmetric two $x_i$’s design with $\nu = 1$, $m = 2$ and $n_1 = n_2 = 20$, that correspond to AIEs in the top-left plot of Figure 4.2. Right-column: For the top-right plot of Figure 4.2 with $\nu = 2$, $m = 2$ and $n_1 = n_2 = 20$. 

$\hat{\beta}_1$
Figure 4.5: Plots of AIEs of $\beta_0$ ended by $\bullet$ when $x_1 = 1$ and $x_2 = 2$, $\blacksquare$ when $x_1 = 1$ and $x_2 = 3$ and $\blacktriangle$ when $x_1 = 1$ and $x_2 = 4$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.6: Plots of AIEs of $\beta_1$ ended by ● when $x_1 = 1$ and $x_2 = 2$, ■ when $x_1 = 1$ and $x_2 = 3$ and ▲ when $x_1 = 1$ and $x_2 = 4$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.7: Plots of AIEs of $\beta_0$ ended by • when $x_1 = -2$ and $x_2 = -1$, ■ when $x_1 = -3$ and $x_2 = -1$ and ▲ when $x_1 = -4$ and $x_2 = -1$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. **Top-right**: $\nu = 2$. **Middle**: Under IBLnM. **Bottom**: Under IBLnM.
Figure 4.8: Plots of AIEs of $\beta_1$ ended by $\bullet$ when $x_1 = -2$ and $x_2 = -1$, $\blacksquare$ when $x_1 = -3$ and $x_2 = -1$ and $\blacktriangle$ when $x_1 = -4$ and $x_2 = -1$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.9: Plots of AIEs of $\beta_0$ ended by $\bullet$ when $x_1 = 3.5$ and $x_2 = 4.5$, $\blacksquare$ when $x_1 = 3$ and $x_2 = 5$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.10: Plots of AIEs of $\beta_1$ ended by • when $x_1 = 3.5$ and $x_2 = 4.5$, ■ when $x_1 = 3$ and $x_2 = 5$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.11: Plots of AIEs of $\beta_0$ ended by ● when $x_1 = -4.5$ and $x_2 = -3.5$, ■ when $x_1 = -5$ and $x_2 = -3$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.

AIEs Under IDM

$v = 1$

$m = 2$

$n = 20$

$\beta_0$

$\hat{\beta}_0$

AIEs Under IDM

$v = 2$

$m = 2$

$n = 20$

$\beta_0$

$\hat{\beta}_0$

AIEs Under IBLnM

$\mu_j = -0.35, -0.45$, $\mu_k = 0.35, 0.45$

$\sigma_j = \sigma_k = 0.05, 0.2, 0.35, 0.5, 0.8, 0.95$

$\beta_0$

$\hat{\beta}_0$

AIEs Under IBLnM

$\mu_j = -0.5, -1$, $\mu_k = 0.5, 1$

$\sigma_j = \sigma_k = 0.05, 0.2, 0.35, 0.5, 0.8, 0.95$

$\beta_0$

$\hat{\beta}_0$

AIEs Under IBLnM

$\mu_j = -0.35, -0.45$, $\mu_k = 0.35, 0.45$

$\sigma_j = \sigma_k = 1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48$

$\beta_0$

$\hat{\beta}_0$

AIEs Under IBLnM

$\mu_j = -0.5, -1$, $\mu_k = 0.5, 1$

$\sigma_j = \sigma_k = 1.0, 1.08, 1.16, 1.24, 1.32, 1.40, 1.48$

$\beta_0$

$\hat{\beta}_0$
Figure 4.12: Plots of AIEs of $\beta_1$ ended by • when $x_1 = -4.5$ and $x_2 = -3.5$, ■ when $x_1 = -5$ and $x_2 = -3$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.13: Plots of AIEs of $\beta_0$ ended by • when $x_i \in [-0.5, 0.5]$, $i = 1, \ldots, m$, ■ when $x_i \in [-1, 1]$ and ▲ when $x_i \in [-2.5, 2.5]$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.14: Plots of AIEs of $\beta_1$ ended by $\bullet$ when $x_i \in [-0.5, 0.5]$, $i = 1, \ldots, m$, ■ when $x_i \in [-1, 1]$ and ▲ when $x_i \in [-2.5, 2.5]$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.15: Plots of AIEs of $\beta_0$ ended by • when $x_i \in [1, 2], i = 1, \ldots, m$, ■ when $x_i \in [1, 3]$ and ▲ when $x_i \in [1, 4]$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.16: Plots of AIEs of $\beta_1$ ended by $\bullet$ when $x_i \in [1, 2]$, $i = 1, \ldots, m$, $\blacksquare$ when $x_i \in [1, 3]$ and $\blacktriangle$ when $x_i \in [1, 4]$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.17: Plots of AIEs of $\beta_0$ ended by $\bullet$ when $x_i \in [-2, -1]$, $i = 1, \ldots, m$, ■ when $x_i \in [-3, -1]$ and ▲ when $x_i \in [-4, -1]$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.18: Plots of AIEs of $\beta_1$ ended by • when $x_i \in [-2, -1]$, $i = 1, \ldots, m$, ■ when $x_i \in [-3, -1]$ and ▲ when $x_i \in [-4, -1]$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.19: Plots of AIEs of $\beta_0$ ended by $\bullet$ when $x_i \in [3.5, 4.5]$, $i = 1, \ldots, m$, $\blacksquare$ when $x_i \in [3, 5]$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.20: Plots of AIEs of $\beta_1$ ended by $\bullet$ when $x_i \in [3.5, 4.5]$, $i = 1, \ldots, m$, $\blacksquare$ when $x_i \in [3, 5]$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.21: Plots of AIEs of $\beta_0$ ended by $\bullet$ when $x_i \in [-4.5, -3.5]$, $i = 1, ..., m$, $\times$ when $x_i \in [-5, -3]$. Percentage values are for single interval estimates that include true $\beta_0$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
Figure 4.22: Plots of AIEs of $\beta_1$ ended by $\bullet$ when $x_i \in [-4.5, -3.5]$, $i = 1, \ldots, m$, ■ when $x_i \in [-5, -3]$. Percentage values are for single interval estimates that include true $\beta_1$. Top-left: Under IDM with $\nu = 1$. Top-right: $\nu = 2$. Middle: Under IBLnM. Bottom: Under IBLnM.
CHAPTER 5

CONCLUSION AND FUTURE PLAN

This chapter presents a conclusion about 2-parameter ImpLogit model. The conclusion, presented in Section 5.1, will give recommendations for reducing imprecision of regression parameters estimates by describing a recommended design for 2-parameter ImpLogit model. Future research plan is given in Section 5.2.

5.1 Conclusion

The first step in building the imprecise logistic regression model is made. ImpLogit model applies probabilistic imprecision to regression parameters. ImpLogit model is a new imprecise logistic regression model that produces interval estimates for regression parameters.

In Chapter 1, the concept of probabilistic imprecision was presented. This required to have a quick look at the robust and hierarchical Bayesian methods. A general look at the historical development of Bayesian methods was given to understand circumstances that led statisticians to adopt the imprecise probabilistic approach. Chapter 1 introduced the imprecise Dirichlet model (IDM). Examples of applications of imprecise probability models were provided.

In Chapter 2, the successive differences of probabilities of occurrence in logistic regression model were assumed to follow the Dirichlet distribution. A new method for inducing a prior distribution to regression parameters in 2-parameter logistic regression model was given. This required to use the aggregation property in Dirichlet distribution to find the joint distribution for any selected $\theta_j$ and $\theta_k$. The new method was applied to both cases of increasing and decreasing logistic regression models. The logit-normal distribution and its multivariate form were introduced. The bivariate logit-normal distribution was fitted
by inducing a prior distribution for regression parameters, subject to the new method. Prior and posterior densities in univariate and multivariate logit-normal distributions were shown, in appendices B and C, to belong to an exponential family of distributions.

In Chapter 3, the shape of 2-parameter ImpLogit model was studied in the first section. It was found that logistic functions under lower and upper values of $\beta_0$ and $\beta_1$ intersect at a negative covariate value. Also, lower and upper values of $\beta_0$ and $\beta_1$ do not always correspond to lower and upper logistic functions. The correspondence occurs over specific parts of the covariate values as being shown in Section 3.1. Chapter 3 presented the relation between $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ and $\boldsymbol{\theta} = [\theta_j, \theta_k]'$, for any $j, k$. This meant to show that lower and upper $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ do not necessarily correspond to lower and upper $\boldsymbol{\theta} = [\theta_j, \theta_k]'$.

In Chapter 3, the 2-parameter ImpLogit model was established, developed and fitted. A new imprecise probability model, called the imprecise bivariate logit-normal model (IBLnM), was built. The 2-parameter ImpLogit model was fitted under IDM and IBLnM. Having a smaller sample size was found to produce higher imprecision in regression parameters estimates.

An advanced simulation study for 2-parameter ImpLogit model was given in Chapter 4. The simulation study was made under several designs of 2-parameter ImpLogit model. The simulation study showed that different factors can associate to increasing (or decreasing) interval estimates lengths in regression parameters. The factors include changes in prior parameters values, locations and spread of fixed covariate values and the number of covariate values ($m$).

In the 2-parameter ImpLogit model, changes in ranges of $\nu$ in IDM, $\mu_i$ and $\sigma_i$ in IBLnM and fixed covariate $x_i$ values, $\forall i = 1, \ldots, m$, affect the lengths of the produced interval estimates of regression parameters. This motivates to think in establishing a better and recommended design. The simulation results in Chapter 4 can be basically used to build and visualize a better design for 2-parameter ImpLogit model.

The ranges of the covariate $x_i$ are suggested to be lengthened and fixed around the prior believed median value of the logistic distribution. Having larger number of covariate values ($m$), provided that the sample size of binary observations is fixed, results in retarding the effect of lengthening $x_i$ ranges on the lengths of interval estimates of $\beta_0$ and $\beta_1$.
\( \beta_1 \). That is, with larger \( m \), the interval estimates are longer. Therefore, short covariate ranges and large \( m \) result in more imprecision in interval estimates of \( \beta_0 \) and \( \beta_1 \). Applying the 2-parameter ImpLogit model, say for a number of dose levels in a dose-response experiment, becomes better by having a small number of dose measured levels.

An important point in fitting 2-parameter ImpLogit model, under IDM, is to recommend \( \nu \) ranges to shorten the interval estimates. The same thing is for IBLnM imprecise prior where \( \sigma_i \), \( i = 1, 2 \), is found to play a stronger role than \( \mu_i \) in controlling the interval estimates lengths of regression parameters. It is recommended for \( \nu \) to belong to [1, 2]. For 2-parameter ImpLogit model under IBLnM, increasing values of the scale parameters \( \sigma_i \) gives less imprecision in the regression parameters estimates. The preference for each \( \sigma_i \) is to take values greater than 1 to shorten the interval estimates. For the location parameters \( \mu_i \), they are not suggested to be far from each other. In bivariate logit-normal distribution, \( \text{E} \left[ \log \left( \frac{\theta_i}{1-\theta_i} \right) \right] = \mu_i \), and if \( \theta_i = 0.5 \), there is \( \log \left( \frac{\theta_i}{1-\theta_i} \right) = 0 \). Then \( \mu_i \)'s are suggested to be closed to zero value from both sides.

In 2-parameter ImpLogit model, the lower and upper estimates of \( \beta_0 \) and \( \beta_1 \) do not necessarily correspond to lower and upper estimates of \( \theta \). The correspondence occurs for all \( x_i \geq 0 \), then the covariate \( x_i \) is better to have positive values.

The previous mentioned factors are found to have a greater impact on \( \beta_1 \) than \( \beta_0 \). Changes in lengths of \( \beta_1 \) interval estimates are larger than those for \( \beta_0 \). This approves the strong effect of the selected covariate ranges on the produced imprecision, since the covariate \( x_i \) associates with \( \beta_1 \) in the model structure.

The study of 2-parameter ImpLogit model was enhanced by using the Fisher information matrix (FIM) in Chapter 4. The covariate values that optimize the estimates of the predictive probabilities \( \theta_i \), \( \forall i \), were computed. An FIM design was compared to the previous ImpLogit designs in Chapter 4. FIM indicated that a covariate range of \( x_1 = -3.85 \) and \( x_2 = 3.85 \) optimizes the 2-parameter logistic regression model. This covariate range was applied to minimize interval estimates lengths. The covariate range of \( x_1 = -3.85 \) and \( x_2 = 3.85 \) included the range of \( x_1 = -2.5 \), \( x_2 = 2.5 \) at which shortest interval estimates were produced for \( \beta_0 \) and \( \beta_1 \).

Multiple parameter ImpLogit model was applied in Chapter 4 for 3 parameters \( \beta_0 \), \( \beta_1 \) and \( \beta_2 \) and two covariates \( x_{i1} \) and \( x_{i2} \). The behaviour of 3-parameter ImpLogit model
does not differ from the 2-parameter case. Interval estimates of $\beta_0$, $\beta_1$ and $\beta_2$ become longer with larger $m$, given that $m$ is the same for both covariates $x_{i1}$ and $x_{i2}$. Having longer ranges of $x_{i1}$ and $x_{i2}$ produced shorter interval estimates.

5.2 Future Plan

A future plan focuses on multiple parameters ImpLogit model. The question is: how each regression parameter, out of $p + 1$ parameters, affects the produced imprecision? It is important to see how ranges of all covariates can be fixed to have shorter interval estimates of all regression parameters. Therefore, Section 4.10 can be extended to study the relation among all covariates and their effect on the produced imprecision of $\beta_0$, $\ldots$, $\beta_p$ estimates.

Another attractive direction is to model the binomial parameter $\theta$ as the cdf of the normal distribution

$$\theta_i = \Phi \left( x_i' \beta \right), \quad i = 1, \ldots, m.$$  \hfill (5.1)

This model is called the “probit” model. Probit model was first presented by Bliss (1935). For more information about the probit model, McCullagh and Nelder (1989) and Albert and Chib (1993) are recommended.

Probabilistic imprecision can be involved in the probit model to be referred to by ImpProbit model. A comparison between ImpProbit and ImpLogit models allows to look at conditions that control interval estimates lengths of regression parameters. ImpProbit model helps to look at the recommended designs in imprecise logistic regression in parallel with ImpLogit model.
Bibliography


138


# Appendix A

## R Code

```r
# This program is for MCMC method to estimate regression
# is logit model.
require(xtable);library(MCMCpack);library(MASS);
library(coda);library(lattice) # some required packages.

### required info
par(mfrow=c(3,3),mar=c(5,4.3,.31,1),oma=c(2,2,1,1))
nu = c(0.5,1,2) ### concentration parameter value in IDM.
df = 5 ### degrees of freedom of t-distribution.
numdraws = 2000 ### number of MCMC draws.
firstdose = 0; seconddose = 10.2 ### first and second dose
numinterval = 5 ### m = 6 dose levels

### the jacobian of the prior induced pdf
prior5 <- function (b0,b1,xf,xs) {
  z = abs(xs-xf)*(exp(b0+xf*b1)/((1+exp(b0+xf*b1))**2))
  w = (exp(b0+xs*b1)/((1+exp(b0+xs*b1))**2))
  z*w}

### amounts of doses.
x = c(0, 2.6, 3.8, 5.1, 7.7, 10.2) ### amounts of doses.
n = c(49, 50, 48, 46, 49, 50) ### number of experimented insects.
yobs = c(0, 6, 16, 24, 42, 44) ### number of affected insects.
m = length(n) ### m value.

### to solve the data in classical approach
probobs = yobs/n
```

143
d = data.frame(cbind(probobs,x))
logit.fit = glm(probobs~x, data = d, family=binomial("logit"))
logit.fit$coefficients[1]
logit.fit$coefficients[2]

q = c(0.15,.3,.45);
f = rep(0,length(q)); s = rep(0,length(q))
b0hat = matrix(rep(0,length(nu)*length(q)),length(nu), length(q))
b1hat = matrix(rep(0,length(nu)*length(q)),length(nu), length(q))
corr = matrix(rep(0,length(nu)*length(q)),length(nu), length(q))
sigma1 = 0.2; sigma2 = c(0.3,0.3,0.3)
for (h in 1:length(nu)){
u1 = rep(0,length(q)); nu2 = rep(0,length(q));
for (k in 1:length(q)){
u1[k]=nu[h]*q[k] ; nu2[k]=nu[h]*(1-q[k])
f[k] = round(q[k]*m) ### the sequential order for theta j
s[k] = round((1.1-q[k])*m) ### the sequential order for theta k
b0cand = numeric(numdraws)
b1cand = numeric(numdraws)
randomt1 = numeric(numdraws)
randomt2 = numeric(numdraws)
jointcand = numeric(numdraws)
randomt1[1] = rt(1,df)
randomt2[1] = rt(1,df)
b0cand[1] = randomt1[1]*sigma1+logit.fit$coefficients[1]
b0 = numeric(numdraws)
b1 = numeric(numdraws)
b0[1] = b0cand[1]
b1[1] = b1cand[1]

# main body of the code
#
# to calculate the joint value for the initial point
#
likecand = numeric(length(x))
priorcand = numeric(numdraws)
likelihood = numeric(numdraws)
comthet1 = numeric(length(x)); comthet2 = numeric(length(x))
comthet = numeric(length(x))
for (i in 1:m){
  comthet[a][i] = exp(b0cand[1] + x[i]*b1cand[1])
  comthet[b][i] = (1+exp(b0cand[1] + x[i]*b1cand[1]))
  comthet[c][i] = comthet[a][i]/comthet[b][i]
  likecand[i] = dbinom(yobs[i],n[i],comthet[c][i])
}
d1 = exp(b0cand[1] + x[f[k]]*b1cand[1])
d2 = (1+exp(b0cand[1] + x[f[k]]*b1cand[1]))
d3 = exp(b0cand[1] + x[s[k]]*b1cand[1])
d4 = (1+exp(b0cand[1] + x[s[k]]*b1cand[1]))
e1 = d1/d2
e2 = d3/d4
e3 = e2-e1
e4 = 1 - e2
ee = c(e1,e3,e4)
density = ddirichlet(ee,c(nu1[k],nu2[k]-nu1[k],nu[h]-nu2[k]))
priorcand[1] = density*(prior5(b0cand[1],b1cand[1],x[f[k]],x[s[k]]))
jointcand[1] = prod(likecand)*priorcand[1]

# the rest of the chain

alphamcmc1 = numeric(numdraws)
alphamcmc2 = numeric(numdraws)
alphamcmc = numeric(numdraws)
for (j in 2:numdraws){
  randomt1[j] = rt(1,df)
  b0cand[j] = randomt1[j]*sigma1+logit.fit$coefficients[1]
  randomt2[j] = rt(1,df)
  b1cand[j] = randomt2[j]*sigma2[h]+logit.fit$coefficients[2]
  for (i in 1:m){
    comptheta1[i] = exp(b0cand[j]+x[i]*b1cand[j])
    comptheta2[i] = (1+exp(b0cand[j]+x[i]*b1cand[j]))
    comptheta[i] = comptheta1[i]/comptheta2[i]
    likecand[i] = dbinom(yobs[i],n[i],comptheta[i])
  }
  d1 = exp(b0cand[j]+x[f[k]]*b1cand[j])
  d2 = (1+exp(b0cand[j]+x[f[k]]*b1cand[j]))
  d3 = exp(b0cand[j]+x[s[k]]*b1cand[j])
  d4 = (1+exp(b0cand[j]+x[s[k]]*b1cand[j]))
  e1 = d1/d2
  e2 = d3/d4
  e3 = e2-e1
  e4 = 1 - e2
  ee = c(e1,e3,e4)
  density = ddirichlet(ee,c(nu1[k],nu2[k]-nu1[k],nu[h]-nu2[k]))
  priorcand[j]=density*(prior5(b0cand[j],b1cand[j],x[f[k]],x[s[k]]))
  likelihood[j] = prod(likecand)
  jointcand[j] = likelihood[j]*priorcand[j]
  alphamcmc1[j]=jointcand[j]
  alphamcmc2[j]=jointcand[j-1]
  ### to find alpha value in Metropolis Hastings.
  alphamcmc[j]=alphamcmc1[j]/alphamcmc2[j]
if (runif(1,0,1) < min(alphamcmc[j],1)){
b0cand[j] = b0cand[j]
b1cand[j] = b1cand[j]
jointcand[j] = jointcand[j]
}
else{
b0cand[j] = b0cand[j-1]
b1cand[j] = b1cand[j-1]
jointcand[j] = jointcand[j-1]
}
b0[j] = b0cand[j]
b1[j] = b1cand[j]
likecand = numeric(length(x))
comptheta = numeric(length(x))
}

############################ Metropolis Hastings ends##############################
b0hat[h,k] = mean(b0)
b1hat[h,k] = mean(b1)

################################################################################ the plots  ################################################################################
x1 = c(seq(min(x)-2,max(x)+2,0.1))
P = matrix(rep(0,length(x1)))
plot(x,yobs/n,pch=20,xlim=c(-2,13),ylim=c(0,1),las=1,ylab="")
x1 = c(seq(min(x)-2,max(x)+2,0.05))
P = matrix(rep(0,length(x1)))
thetaestimate1 = rep(0,length(x1));
thetaestimate2 = rep(0,length(x1))
thetaestimate = rep(0,length(x1))
for (i in 1:length(x1)){
thetaestimate1[i] = (exp(b0hat[h,k]+x1[i]*b1hat[h,k]))
thetaestimate2[i] = (1+exp(b0hat[h,k]+x1[i]*b1hat[h,k]))
thetaestimate[i] = thetaestimate1[i]/thetaestimate2[i]
}
lines(x1,thetaestimate,pch=20,lty=1,col=2,xlim=c(-2,13),ylim=c(0,1))
text(-1,0.9,expression(hat(beta)[0]))
text(.5,0.9,"="

text(3.5,0.9,round(b0hat[h,k],3))
text(-1,0.7,expression(hat(beta)[1]))
text(.5,0.7,"="

text(3.3,0.7,round(b1hat[h,k],3))

text(expression(nu),at=-2,3,cex=1.5)

text("=",at=0,3,lwd=1.5)

text(nu[h],at=3,3,lwd=1.5)

text("Pair:",at=8,3,lwd=1.5,cex=1)

text(f[k],at=11,3,lwd=1.5)

text("",at=12,3,lwd=1.5)

text(s[k],at=13,3,lwd=1.5)

############# code is finished##################

table1 = xtable(cbind(x,n,yobs), digits=3)
colnames(table1) <- c("Dose","n","y_{obs}"

print(table1, sanitize.text.function = function(x) { x })

table2 = xtable(b0hat, digits=3)
c1 = c("\theta_{1}, \theta_{6}\left(x_{1},x_{6}\right)"
c2 = c("\theta_{2}, \theta_{5}\left(x_{2},x_{5}\right)"
c3 = c("\theta_{3}, \theta_{4}\left(x_{3},x_{4}\right)"

148
rownames(table2) <- c(c1,c2,c3)
colnames(table2) <- c("$\nu=0.5$","$\nu=1$","$\nu=2$")
table2 = xtable(table2, digits=3)
align(table2) = "cccc"
print(table2, sanitize.text.function = function(x) { x })
table3 = xtable(b1hat, digits=3)
c1 = c("$\theta_{1}$, $\theta_{6}\left(x_{1},x_{6}\right)$")
c2 = c("$\theta_{2}$, $\theta_{5}\left(x_{2},x_{5}\right)$")
c3 = c("$\theta_{3}$, $\theta_{4}\left(x_{3},x_{4}\right)$")
rownames(table3) <- c(c1,c2,c3)
colnames(table3) <- c("$\nu=0.5$","$\nu=1$","$\nu=2$")
table3 = xtable(table3, digits=3)
align(table3) = "cccc"
print(table3, sanitize.text.function = function(x) { x })
Appendix B

Posterior logit-normal density

The logit-normal distribution is related to the normal distribution as follows

\[ \text{if } \theta \sim \text{logit-normal } (\mu, \sigma^2), \quad (B.1) \]

and

\[ \text{if } \theta = \frac{e^w}{1 + e^w}, \quad (B.2) \]

then

\[ w = \log \left( \frac{\theta}{1 - \theta} \right) \sim \text{N} (\mu, \sigma^2). \quad (B.3) \]

To find the distribution of \( w \) conditional on \( y \), where \( y \) comes from a binomial random variable \( \text{bin}(n, \theta) \), the posterior pdf is derived as follows

\[
\begin{align*}
\pi(w|y) &= \frac{f(y|w) \cdot \pi(w)}{m(y)} = \frac{\text{binomial} \times \text{normal}}{\text{marginal}} \\
&= \left[ \binom{n}{y} \left( \frac{e^w}{1 + e^w} \right)^y \left( 1 - \frac{e^w}{1 + e^w} \right)^{n-y} \right] \cdot \left\{ \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(w - \mu)^2}{2\sigma^2} \right] \right\} \\
&\propto \left[ \left( \frac{e^w}{1 + e^w} \right)^y \left( 1 - \frac{e^w}{1 + e^w} \right)^{n-y} \right] \left[ \exp \left( -\frac{(w - \mu)^2}{2\sigma^2} \right) \right] \\
&= \frac{e^{wy}}{(1 + e^w)^n} \left\{ \exp \left[ -\frac{(w - \mu)^2}{2\sigma^2} \right] \right\} = \frac{\exp \left[ -\frac{(w - \mu)^2 + 2\sigma^2 wy}{2\sigma^2} \right]}{(1 + e^w)^n} \\
&= \frac{\exp \left[ -\frac{(w^2 + \mu^2 - 2w\mu + 2\sigma^2 wy}{2\sigma^2} \right]}{(1 + e^w)^n}
\end{align*}
\]
\[
\exp \left( \frac{-w^2 - \mu^2 + 2w\mu + 2\sigma^2 w y}{2\sigma^2} \right) \\
\frac{(1 + e^w)^n}{(1 + e^w)^n} \\
\exp \left\{ \frac{-w^2 - \mu^2 + 2w\mu + 2\sigma^2 w y - 2\mu\sigma^2 y - (\sigma^2 y)^2}{2\sigma^2} \right\} \\
\alpha \\
\exp \left\{ \frac{-[w - (\mu + \sigma^2 y)]^2}{2\sigma^2} \right\} \\
\frac{(1 + e^w)^n}{(1 + e^w)^n}.
\]

(B.4)

Now, the posterior distribution of \( \theta \) is found by writing (B.4) in terms of \( \theta \), where based on (B.3), there is

\[
\pi (\theta | y) \propto \exp \left\{ \frac{- \left[ \log \left( \frac{\theta}{1-\theta} \right) - (\mu + \sigma^2 y) \right]^2}{2\sigma^2} \right\} \\
\frac{1}{\theta (1-\theta)} \cdot \frac{1}{\theta (1-\theta)} \\
\exp \left\{ \frac{- \left[ \log \left( \frac{\theta}{1-\theta} \right) - (\mu + \sigma^2 y) \right]^2}{2\sigma^2} \right\} \\
\alpha \cdot \frac{1}{\theta (1-\theta)} \\
\alpha \exp \left\{ \frac{- \left[ \log \left( \frac{\theta}{1-\theta} \right) - (\mu + \sigma^2 y) \right]^2}{2\sigma^2} \right\} \\
\frac{(1 - \theta)^{n-1}}{\theta}.
\]

(B.5)

Both of the prior and posterior distributions in (2.40) and (B.5) belong to exponential families. This can be shown by taking the log of the prior density in (2.40) so that it is
is written as

\[
\log [ f (\theta | \mu, \sigma)] = \log \left\{ \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ - \frac{\left[ \log \left( \frac{\theta}{1 - \theta} \right) - \mu \right]^2}{2\sigma^2} \right\} \cdot \left[ \frac{1}{\theta (1 - \theta)} \right] \right\}
\]

\[
= \log \left( \frac{1}{\sigma \sqrt{2\pi}} \right) - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1 - \theta} \right) \right]^2 - \frac{\mu^2}{2\sigma^2} + \frac{2\mu}{2\sigma^2} \left[ \log \left( \frac{\theta}{1 - \theta} \right) \right]
\]

\[
+ \log \left[ \frac{1}{\theta (1 - \theta)} \right]
\]

\[
= \log \left( \frac{1}{\sigma \sqrt{2\pi}} \right) - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1 - \theta} \right) \right]^2 - \frac{\mu^2}{2\sigma^2}
\]

\[
+ \frac{2\mu}{2\sigma^2} \left[ \log (\theta) - \log (1 - \theta) \right] - \log (\theta) - \log (1 - \theta)
\]

\[
= \frac{\log \left( \frac{1}{\sigma \sqrt{2\pi}} \right) - \frac{\mu^2}{2\sigma^2} + \frac{2\mu}{2\sigma^2} \left[ \log \left( \frac{\theta}{1 - \theta} \right) \right]}{\eta_0} + \left\{ - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1 - \theta} \right) \right]^2 \right\} \frac{1}{\eta_1}
\]

\[
+ \left[ \frac{\mu}{\sigma^2} - 1 \right] \log (\theta) \frac{1}{\eta_2} + \left[ - \frac{\mu}{\sigma^2} + 1 \right] \log (1 - \theta) \frac{1}{\eta_3}
\]

\[
= \eta_0 (\mu, \sigma) + \eta_1 (\mu, \sigma) \tau_1 (\theta) + \eta_2 (\mu, \sigma) \tau_2 (\theta) + \eta_3 (\mu, \sigma) \tau_3 (\theta) . \quad \text{(B.6)}
\]

It is clear that (B.6) forms an exponential family since \( \eta_0, \eta_1, \eta_2 \) and \( \eta_3 \) are the
The log of the posterior pdf given in (B.5) is also taken and compared to the log of the prior density in (B.6). That is

$\log [\pi (\theta|y)] = \log \left\{ \exp \left\{ - \left[ \log \left( \frac{\theta}{1-\theta} \right) - (\mu + \sigma^2 y) \right]^2 \right\} \frac{(1-\theta)^{n-1}}{\theta} \right\}$

$= K + \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]^2 - (\mu + \sigma^2 y)^2 + 2(\mu + \sigma^2 y) \log \left( \frac{\theta}{1-\theta} \right) + \log \left( \frac{(1-\theta)^{n-1}}{\theta} \right)$

$= K - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]^2 - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]^2 + \frac{\mu + \sigma^2 y}{\sigma^2} \log \left( \frac{\theta}{1-\theta} \right)$

$+ (n-1) \log (1-\theta) - \log (\theta)$

$= K - \frac{(\mu + \sigma^2 y)^2}{2\sigma^2} - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]^2 + \frac{\mu + \sigma^2 y}{\sigma^2} \log \left( \frac{\theta}{1-\theta} \right)$

$- \frac{\mu + \sigma^2 y}{\sigma^2} \log (1-\theta)$

$+ (n-1) \log (1-\theta) - \log (\theta)$

$= K - \frac{(\mu + \sigma^2 y)^2}{2\sigma^2} - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]^2 + \left( \frac{\mu + \sigma^2 y}{\sigma^2} - 1 \right) \log (\theta)$

$+ \left( -\frac{\mu + \sigma^2 y}{\sigma^2} + n-1 \right) \log (1-\theta)$
\[
\begin{align*}
&= \left[ K - \frac{(\mu + \sigma^2 y)^2}{2\sigma^2} \right] + \left\{ \frac{1}{2\sigma^2} \log \left( \frac{\theta}{1 - \theta} \right)^2 \right\} \\
&+ \left[ \frac{\mu}{\sigma^2} - 1 + y \right] \log (\theta) + \left[ \frac{-\mu}{\sigma^2} - 1 + n - y \right] \log (1 - \theta) \\
&= \eta_0^* (\mu, \sigma) + \eta_1^* (\mu, \sigma) \tau_1 (\theta) + \eta_2^* (\mu, \sigma) \tau_2 (\theta) + \eta_3^* (\mu, \sigma) \tau_3 (\theta). 
\end{align*}
\]  
(B.8)

Comparing both of (B.6) and (B.8) shows that
\[
\begin{align*}
\eta_1^* &= \eta_1, \\
\eta_2^* &= \eta_2 + y, \\
\eta_3^* &= \eta_3 + n - y.
\end{align*}
\]  
(B.9)
Appendix C

Posterior multivariate logit-normal density

Taking the log of the posterior multivariate logit-normal density gives

\[
\log [\pi (\theta | y)] = K - \frac{1}{2} [\log (\frac{\theta}{1-\theta}) - (\mu + \Sigma y)]^\prime \Sigma^{-1} [\log (\frac{\theta}{1-\theta}) - (\mu + \Sigma y)] \\
+ 1^\prime \log \left[ \frac{(1 - \theta)^{n-1}}{\theta} \right] \\
= K - \frac{1}{2} [\log (\frac{\theta}{1-\theta})]^\prime \Sigma^{-1} [\log (\frac{\theta}{1-\theta})] + \frac{1}{2} [\log (\frac{\theta}{1-\theta})]^\prime \Sigma^{-1} (\mu + \Sigma y) \\
+ \frac{1}{2} (\mu + \Sigma y)^\prime \Sigma^{-1} \left[ \log (\frac{\theta}{1-\theta}) \right] - \frac{1}{2} (\mu + \Sigma y)^\prime \Sigma^{-1} (\mu + \Sigma y) \\
+ 1^\prime \log [(1 - \theta)^{n-1}] - 1^\prime \log (\theta) \\
= K - \frac{1}{2} (\mu + \Sigma y)^\prime \Sigma^{-1} (\mu + \Sigma y) - \frac{1}{2\sigma^2} \left[ \log (\frac{\theta}{1-\theta}) \right]^\prime R^{-1} \left[ \log (\frac{\theta}{1-\theta}) \right] \\
+ \frac{1}{2\sigma^2} \left[ \log (\frac{\theta}{1-\theta}) \right]^\prime R^{-1} (\mu + \Sigma y) + \frac{1}{2\sigma^2} (\mu + \Sigma y)^\prime R^{-1} \left[ \log (\frac{\theta}{1-\theta}) \right] \\
+ 1^\prime \log [(1 - \theta)^{n-1}] - 1^\prime \log (\theta) \\
= K - \frac{1}{2\sigma^2} (\mu + \Sigma y)^\prime R^{-1} (\mu + \Sigma y) - \frac{1}{2\sigma^2} \left[ \log (\frac{\theta}{1-\theta}) \right]^\prime R^{-1} \left[ \log (\frac{\theta}{1-\theta}) \right] \\
+ \frac{1}{2\sigma^2} [\log (\theta) - \log (1 - \theta)]^\prime R^{-1} (\mu + \Sigma y) + \frac{1}{2\sigma^2} (\mu + \Sigma y)^\prime R^{-1} [\log (\theta) - \log (1 - \theta)] \\
+ (n - 1)^\prime \log [(1 - \theta)] - 1^\prime \log (\theta) \\
= K - \frac{1}{2\sigma^2} (\mu + \Sigma y)^\prime R^{-1} (\mu + \Sigma y) - \frac{1}{2\sigma^2} \left[ \log (\frac{\theta}{1-\theta}) \right]^\prime R^{-1} \left[ \log (\frac{\theta}{1-\theta}) \right] \\
+ \frac{1}{2} [\log (\theta) - \log (1 - \theta)]^\prime R^{-1} \left( \frac{1}{\sigma^2} \mu + R y \right) + \frac{1}{2} \left( \frac{1}{\sigma^2} \mu + R y \right)^\prime R^{-1} [\log (\theta) - \log (1 - \theta)] \\
+ (n - 1)^\prime \log [(1 - \theta)] - 1^\prime \log (\theta)
\]
\[ K - \frac{1}{2\sigma^2} (\mu + \Sigma y) R^{-1} (\mu + \Sigma y) - \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]' R^{-1} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right] \\
+ \left[ \left( \frac{1}{\sigma^2} \mu + R y \right)' R^{-1} - 1 \right]' \log(\theta) - \left[ \left( \frac{1}{\sigma^2} \mu + R y \right)' R^{-1} - (n - 1)' \right] \log(1 - \theta) \\
= \left[ K - \frac{1}{2\sigma^2} (\mu + \Sigma y)' R^{-1} (\mu + \Sigma y) \right] \eta_0 + \left\{ \frac{1}{2\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]' R^{-1} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right] \right\} \tau_1 \\
+ \left[ \left( \frac{1}{\sigma^2} \mu' R^{-1} + y' - 1 \right)' \log(\theta) \right] \eta_2 + \left\{ \frac{1}{\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]' R^{-1} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right] \right\} \tau_2 \\
+ \left[ \left( \frac{1}{\sigma^2} \mu' R^{-1} + n' - y' - 1 \right)' \log(\theta) \right] \eta_3 + \left\{ \frac{1}{\sigma^2} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right]' R^{-1} \left[ \log \left( \frac{\theta}{1-\theta} \right) \right] \right\} \tau_3 \\
= \eta_0 (\mu, \sigma, R) + \eta_1 (\mu, \sigma, R) \tau_1 (\theta) + \eta_2 (\mu, \sigma, R) \tau_2 (\theta) + \eta_3 (\mu, \sigma, R) \tau_3 (\theta). \quad (C.1) \]

From (C.1), the hyperparameters are

\[
\begin{align*}
\eta_1^* &= -\frac{1}{2\sigma^2}, \\
\eta_2^* &= \frac{1}{\sigma^2} \mu' R^{-1} + y' - 1', \\
\eta_3^* &= -\frac{1}{\sigma^2} \mu' R^{-1} + n' - y' - 1'.
\end{align*} \quad (C.2)
\]
Appendix D
Calculations and graphs from Chapter 3

Table D.1: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to top plots in Figure D.1 for 2-parameter mixed ImpLogit model with $w = 0.5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>-3.231</td>
<td>-3.216</td>
<td>0.015</td>
<td>0.584</td>
<td>0.618</td>
<td>0.034</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>-3.232</td>
<td>-3.217</td>
<td>0.015</td>
<td>0.580</td>
<td>0.619</td>
<td>0.039</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>-3.233</td>
<td>-3.218</td>
<td>0.015</td>
<td>0.578</td>
<td>0.619</td>
<td>0.041</td>
</tr>
</tbody>
</table>

Table D.2: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to middle plots in Figure D.1 for 2-parameter mixed ImpLogit model with $w = 0.8$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>-3.224</td>
<td>-3.222</td>
<td>0.002</td>
<td>0.596</td>
<td>0.619</td>
<td>0.023</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>-3.225</td>
<td>-3.219</td>
<td>0.006</td>
<td>0.594</td>
<td>0.621</td>
<td>0.027</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>-3.227</td>
<td>-3.218</td>
<td>0.009</td>
<td>0.592</td>
<td>0.622</td>
<td>0.030</td>
</tr>
</tbody>
</table>

Table D.3: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to bottom plots in Figure D.1 for 2-parameter mixed ImpLogit model with $w = 0.9$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>-3.227</td>
<td>-3.220</td>
<td>0.007</td>
<td>0.601</td>
<td>0.615</td>
<td>0.014</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>-3.226</td>
<td>-3.221</td>
<td>0.005</td>
<td>0.598</td>
<td>0.618</td>
<td>0.020</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>-3.227</td>
<td>-3.219</td>
<td>0.008</td>
<td>0.597</td>
<td>0.619</td>
<td>0.022</td>
</tr>
</tbody>
</table>
**Table D.4:** Interval estimates of $\theta$ for Table D.1 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\tilde{\theta}$</th>
<th>imp($\hat{\theta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>0.423</td>
<td>0.469</td>
<td>0.046</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>0.418</td>
<td>0.470</td>
<td>0.052</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>0.415</td>
<td>0.469</td>
<td>0.054</td>
</tr>
</tbody>
</table>

**Table D.5:** Interval estimates of $\theta$ for Table D.2 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\tilde{\theta}$</th>
<th>imp($\hat{\theta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>0.439</td>
<td>0.468</td>
<td>0.029</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>0.437</td>
<td>0.472</td>
<td>0.035</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>0.434</td>
<td>0.473</td>
<td>0.039</td>
</tr>
</tbody>
</table>

**Table D.6:** Interval estimates of $\theta$ for Table D.3 when $x = 5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\theta}$</th>
<th>$\tilde{\theta}$</th>
<th>imp($\hat{\theta}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>0.445</td>
<td>0.464</td>
<td>0.019</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>0.441</td>
<td>0.468</td>
<td>0.027</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>0.440</td>
<td>0.470</td>
<td>0.030</td>
</tr>
</tbody>
</table>
Figure D.1: Top: Plots of mixed ImpLogit model with $\nu = 1.5$, $w = 0.5$ and selected pairs $\theta_1, \theta_6$ (top-left), $\theta_2, \theta_5$ (top-middle), $\theta_3, \theta_4$ (top-right). Middle: $\nu = 1.5$, $w = 0.8$ and selected pairs $\theta_1, \theta_6$ (middle-left), $\theta_2, \theta_5$ (middle-middle) and $\theta_3, \theta_4$ (middle-right). Bottom: $\nu = 1.5$, $w = 0.9$ and selected pairs $\theta_1, \theta_6$ (bottom-left), $\theta_2, \theta_5$ (bottom-middle) and $\theta_3, \theta_4$ (bottom-right).
Table D.7: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to top plots in Figure D.2 for 2-parameter ImpLogit model under IBLnM.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>-3.225</td>
<td>-3.22</td>
<td>0.005</td>
<td>0.591</td>
<td>0.614</td>
<td>0.023</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>-3.227</td>
<td>-3.223</td>
<td>0.004</td>
<td>0.585</td>
<td>0.614</td>
<td>0.029</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>-3.227</td>
<td>-3.223</td>
<td>0.004</td>
<td>0.583</td>
<td>0.614</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Table D.8: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to middle plots in Figure D.2 for 2-parameter ImpLogit model under IBLnM.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>-3.225</td>
<td>-3.219</td>
<td>0.006</td>
<td>0.594</td>
<td>0.622</td>
<td>0.028</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>-3.228</td>
<td>-3.219</td>
<td>0.009</td>
<td>0.594</td>
<td>0.624</td>
<td>0.030</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>-3.229</td>
<td>-3.218</td>
<td>0.011</td>
<td>0.590</td>
<td>0.623</td>
<td>0.033</td>
</tr>
</tbody>
</table>

Table D.9: Interval estimates of $\beta_0$ and $\beta_1$ that correspond to bottom plots in Figure D.2 for 2-parameter ImpLogit model under IBLnM.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1$, $\theta_6(x_1, x_6)$</td>
<td>-3.229</td>
<td>-3.220</td>
<td>0.009</td>
<td>0.595</td>
<td>0.623</td>
<td>0.028</td>
</tr>
<tr>
<td>$\theta_2$, $\theta_5(x_2, x_5)$</td>
<td>-3.229</td>
<td>-3.215</td>
<td>0.014</td>
<td>0.596</td>
<td>0.626</td>
<td>0.030</td>
</tr>
<tr>
<td>$\theta_3$, $\theta_4(x_3, x_4)$</td>
<td>-3.230</td>
<td>-3.216</td>
<td>0.014</td>
<td>0.591</td>
<td>0.626</td>
<td>0.035</td>
</tr>
</tbody>
</table>
Figure D.2: Top: Plots of ImpLogit model with $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{0.55, 0.7, 0.85, 1, 1.15, 1.3, 1.45\}$ and selected pairs $\theta_1, \theta_6$ (top-left), $\theta_2, \theta_5$ (top-middle), $\theta_3, \theta_4$ (top-right). Middle: $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{0.3, 0.45, 0.6, 0.75, 0.9, 1.05, 1.2\}$. Bottom: $\mu_j \in \{-0.5, -1\}$, $\mu_k \in \{0.5, 1\}$, $\sigma_j = \sigma_k \in \{0.05, 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$. 
Table D.10: Interval estimates of $\beta_0$ and $\beta_1$ for 2-parameter increasing ImpLogit model under a small sample and when $\nu = 0.5$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>-3.246</td>
<td>-3.201</td>
<td>0.045</td>
<td>0.571</td>
<td>0.632</td>
<td>0.061</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>-3.249</td>
<td>-3.207</td>
<td>0.042</td>
<td>0.568</td>
<td>0.633</td>
<td>0.065</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>-3.248</td>
<td>-3.205</td>
<td>0.043</td>
<td>0.565</td>
<td>0.633</td>
<td>0.068</td>
</tr>
</tbody>
</table>

Table D.11: Interval estimates of $\beta_0$ and $\beta_1$ for 2-parameter increasing ImpLogit model under a small sample and when $\nu = 1$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>-3.246</td>
<td>-3.199</td>
<td>0.047</td>
<td>0.575</td>
<td>0.640</td>
<td>0.065</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>-3.247</td>
<td>-3.198</td>
<td>0.049</td>
<td>0.571</td>
<td>0.642</td>
<td>0.071</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>-3.248</td>
<td>-3.198</td>
<td>0.050</td>
<td>0.571</td>
<td>0.642</td>
<td>0.071</td>
</tr>
</tbody>
</table>

Table D.12: Interval estimates of $\beta_0$ and $\beta_1$ for 2-parameter increasing ImpLogit model under a small sample and when $\nu = 2$.

<table>
<thead>
<tr>
<th>selected pair</th>
<th>$\hat{\beta}_0$</th>
<th>$\bar{\beta}_0$</th>
<th>imp ($\hat{\beta}_0$)</th>
<th>$\hat{\beta}_1$</th>
<th>$\bar{\beta}_1$</th>
<th>imp ($\hat{\beta}_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1, \theta_6(x_1, x_6)$</td>
<td>-3.247</td>
<td>-3.199</td>
<td>0.048</td>
<td>0.573</td>
<td>0.642</td>
<td>0.069</td>
</tr>
<tr>
<td>$\theta_2, \theta_5(x_2, x_5)$</td>
<td>-3.249</td>
<td>-3.197</td>
<td>0.052</td>
<td>0.572</td>
<td>0.645</td>
<td>0.073</td>
</tr>
<tr>
<td>$\theta_3, \theta_4(x_3, x_4)$</td>
<td>-3.250</td>
<td>-3.195</td>
<td>0.055</td>
<td>0.572</td>
<td>0.646</td>
<td>0.074</td>
</tr>
</tbody>
</table>
Figure E.1: Plots of AIEs of $\beta_0$ when $m = 20$ and $n_i = 2$, $\forall i$, and selections of $\theta_5$ and $\theta_{16}$, $\theta_8$ and $\theta_{13}$ and $\theta_{10}$ and $\theta_{11}$. Percentage values are for single interval estimates that include true $\beta_0$. 
Figure E.2: Plots of AIEs of $\beta_1$ when $m = 20$ and $n_i = 2, \forall i$, and selections of $\theta_5$ and $\theta_{16}$, $\theta_8$ and $\theta_{13}$ and $\theta_{10}$ and $\theta_{11}$. Percentage values are for single interval estimates that include true $\beta_1$. 

164
Figure E.3: Plots of AIEs of $\beta_0$ when $m = 20$ and $n_i = 2$, $\forall i$, and selections of $\theta_5$ and $\theta_{16}$, $\theta_8$ and $\theta_{13}$ and $\theta_{10}$ and $\theta_{11}$. Percentage values are for single interval estimates that include true $\beta_0$. 

165
AIEs Under IDM

\[ \hat{\beta}_1 \]

**Figure E.4:** Plots of AIEs of \( \beta_1 \) when \( m = 20 \) and \( n_i = 2, \forall i \), and selections of \( \theta_5 \) and \( \theta_{16}, \theta_8 \) and \( \theta_{13} \) and \( \theta_{10} \) and \( \theta_{11} \). Percentage values are for single interval estimates that include true \( \beta_1 \).
Figure E.5: Plots of AIEs of $\beta_0$ ended by • when $x_1 = -0.5$ and $x_2 = 0.5$, ■ when $x_1 = -1$ and $x_2 = 1$, ▲ when $x_1 = -2.5$ and $x_2 = 2.5$ and ▽ when $x_1 = -3.85$ and $x_2 = 3.85$. Percentage values are for single interval estimates that include true $\beta_0$. 

167
**Figure E.6:** Plots of AIEs of $\beta_1$ ended by • when $x_1 = -0.5$ and $x_2 = 0.5$, ■ when $x_1 = -1$ and $x_2 = 1$, ▲ when $x_1 = -2.5$ and $x_2 = 2.5$ and ▽ when $x_1 = -3.85$ and $x_2 = 3.85$. Percentage values are for single interval estimates that include true $\beta_1$. 

- AIEs Under IDM
  - $\nu = 1$
  - $\beta_1$
  - $n = 20$
  - $m = 2$
  - $95\%$
- AIEs Under IDM
  - $\nu = 2$
  - $\beta_1$
  - $n = 20$
  - $m = 2$
  - $97\%$

- AIEs Under IBLnM
  - $\mu_j = -0.35, -0.45$ and $\mu_k = 0.35, 0.45$
  - $\sigma_j = \sigma_k$
  - $n = 20$
  - $m = 2$
  - $95\%$
- AIEs Under IBLnM
  - $\mu_j = -0.5, -1$ and $\mu_k = 0.5, 1$
  - $\sigma_j = \sigma_k$
  - $n = 20$
  - $m = 2$
  - $98\%$

- AIEs Under IBLnM
  - $\mu_j = -0.35, -0.45$ and $\mu_k = 0.35, 0.45$
  - $\sigma_j = \sigma_k$
  - $n = 20$
  - $m = 2$
  - $93\%$
- AIEs Under IBLnM
  - $\mu_j = -0.5, -1$ and $\mu_k = 0.5, 1$
  - $\sigma_j = \sigma_k$
  - $n = 20$
  - $m = 2$
  - $95\%$
Figure E.7: Plots of AIEs in 3-parameter ImpLogit model under IDM with $\nu = 1$, of $\beta_0$, $\beta_1$ and $\beta_2$ ended by • when $x_{i1} \in [-1, 1]$ and $x_{i2} \in [-1, 1]$, ■ when $x_{i1} \in [-2.5, 2.5]$ and $x_{i2} \in [-2.5, 2.5]$, $i = 1, \ldots, m$. 

169
Figure E.8: Plots of AIEs in 3-parameter ImpLogit model under IDM with $\nu = 1$, of $\beta_0$, $\beta_1$ and $\beta_2$ ended by • when $x_{i1} \in [-1, 1]$ and $x_{i2} \in [1, 3]$, ■ when $x_{i1} \in [-2.5, 2.5]$ and $x_{i2} \in [1, 6]$, $i = 1, \ldots, m$. 
Fisher Information Matrix for Logistic Regression

To find the information matrix, the likelihood function of the logistic regression model is

\[ L(\theta_1, \ldots, \theta_m | y) = \prod_{i=1}^{m} \left( \frac{n_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{n_i - y_i} \right) \]

\[ \Rightarrow \log [L(\theta_1, \ldots, \theta_m | y)] = \sum_{i=1}^{m} \log \left( \frac{n_i}{y_i} \right) + \sum_{i=1}^{m} y_i \log \theta_i + \sum_{i=1}^{m} (n_i - y_i) \log (1 - \theta_i) \]

\[ \Rightarrow \log [L(\beta_0, \beta_1 | y)] = \sum_{i=1}^{m} \log \left( \frac{n_i}{y_i} \right) + \sum_{i=1}^{m} y_i \log \left( \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) + \sum_{i=1}^{m} (n_i - y_i) \log \left( \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right). \]  

Then (F.1) is used in the information matrix of \( \beta_0 \) and \( \beta_1 \) as

\[ I(\beta_0, \beta_1) = \begin{bmatrix} -E \left[ \frac{\partial^2}{\partial \beta_0^2} \log L(\beta_0, \beta_1 | y) \right] & -E \left[ \frac{\partial^2}{\partial \beta_0 \partial \beta_1} \log L(\beta_0, \beta_1 | y) \right] \\ -E \left[ \frac{\partial^2}{\partial \beta_0 \partial \beta_1} \log L(\beta_0, \beta_1 | y) \right] & -E \left[ \frac{\partial^2}{\partial \beta_1^2} \log L(\beta_0, \beta_1 | y) \right] \end{bmatrix}, \]  

(F.2)

where from (F.2),

\[ \frac{\partial^2}{\partial \beta_0^2} \log L(\beta_0, \beta_1 | y) = \frac{\partial^2}{\partial \beta_0^2} \left[ \sum_{i=1}^{m} y_i \log \left( \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right] + \frac{\partial^2}{\partial \beta_0^2} \left[ \sum_{i=1}^{m} (n_i - y_i) \log \left( \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right] \]

\[ = \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^{m} \frac{y_i}{e^{\beta_0 + \beta_1 x_i}} \right\} + \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^{m} \frac{\partial}{\partial \beta_0} \left[ \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right] \right\} + \sum_{i=1}^{m} (n_i - y_i) \frac{\partial}{\partial \beta_0} \left[ \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right]. \]
The second derivative with respect to \( \beta \)
\[
\frac{\partial^2}{\partial \beta_0^2} \left\{ \sum_{i=1}^{m} \left( \frac{(1 + e^{\beta_0 + \beta_1 x_i}) e^{\beta_0 + \beta_1 x_i} - (e^{\beta_0 + \beta_1 x_i})^2}{(1 + e^{\beta_0 + \beta_1 x_i})^2} \right) \right\}
\]
\[
+ \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^{m} (n_i - y_i) \frac{-e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right\}
\]
\[
= \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^{m} \frac{y_i}{1 + e^{\beta_0 + \beta_1 x_i}} \right\} + \frac{\partial}{\partial \beta_0} \left\{ \sum_{i=1}^{m} \left( \frac{n_i - y_i}{1 + e^{\beta_0 + \beta_1 x_i}} \right) \right\}
\]
\[
= \sum_{i=1}^{m} \frac{y_i}{1 + e^{\beta_0 + \beta_1 x_i}} - \sum_{i=1}^{m} \frac{-e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \left( \frac{1 + e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i}} \right)
\]
\[
= \sum_{i=1}^{m} \frac{y_i}{1 + e^{\beta_0 + \beta_1 x_i}} + \sum_{i=1}^{m} \frac{(n_i - y_i)}{1 + e^{\beta_0 + \beta_1 x_i}} \left( \frac{1 + e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i}} \right)
\]
\[
= \sum_{i=1}^{m} \frac{y_i}{1 + e^{\beta_0 + \beta_1 x_i}} - \sum_{i=1}^{m} \frac{n_i}{1 + e^{\beta_0 + \beta_1 x_i}} \left( \frac{1 + e^{\beta_0 + \beta_1 x_i}}{e^{\beta_0 + \beta_1 x_i}} \right)
\]
\[
= -\sum_{i=1}^{m} \frac{y_i}{1 + e^{\beta_0 + \beta_1 x_i}} \left( \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = -\sum_{i=1}^{m} n_i \theta_i (1 - \theta_i).
\]

(F.3)

The second derivative with respect to \( \beta_1 \) is
\[
\frac{\partial^2}{\partial \beta_1^2} \log L (\beta_0, \beta_1 | y) = -\sum_{i=1}^{m} \frac{n_i x_i^2}{1 + e^{\beta_0 + \beta_1 x_i}} \left( \frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right) = -\sum_{i=1}^{m} n_i x_i^2 \theta_i (1 - \theta_i),
\]

(F.4)

and
\[
\frac{\partial^2}{\partial \beta_0 \partial \beta_1} \log L (\beta_0, \beta_1 | y) = \frac{\partial^2}{\partial \beta_1 \partial \beta_0} \log L (\beta_0, \beta_1 | y)
\]
\[
= -\sum_{i=1}^{m} x_i n_i \frac{e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) = -\sum_{i=1}^{m} x_i n_i \theta_i (1 - \theta_i),
\]

(F.5)

then plugging (F.3), (F.4) and (F.5) in (F.2) gives FIM as
\[
\begin{bmatrix}
\sum_{i=1}^{m} \frac{n_i e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) & \sum_{i=1}^{m} x_i n_i e^{\beta_0 + x_i \beta_1} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) \\
\sum_{i=1}^{m} x_i n_i e^{\beta_0 + x_i \beta_1} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right) & \sum_{i=1}^{m} \frac{n_i x_i^2 e^{\beta_0 + x_i \beta_1}}{1 + e^{\beta_0 + x_i \beta_1}} \left( \frac{1}{1 + e^{\beta_0 + x_i \beta_1}} \right)
\end{bmatrix}.
\]

(F.6)

172