Optimum Design in Nonlinear and Generalized Linear Mixed Models

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von MSc. Parisa Parsamaram

geb. am 22.09.1989 in Tehran, Iran

genehmigt durch die Fakultät für Mathematik

der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Rainer Schwabe
Prof. Dr. Weng Kee Wong

eingereicht am: 24.08.2022

Verteidigung am: 28.11.2022

Abstract

Generalized linear and nonlinear mixed effects models have been used in many fields of application, such as psychology, medicine, and engineering, etc. There are multiple observations within subjects over time. Hence, the structure of the data in these models is longitudinal. This means that data within each subject are correlated and between the subjects are uncorrelated. In this dissertation, we treat the binary and ordinal mixed effects regression model, where the response variable includes two or more than two levels, respectively. Moreover, a nonlinear longitudinal Poisson regression model is considered to test the ability of the subjects.

For estimating the model parameters, we intend to use the maximum likelihood estimator of the parameters due to its well behaved asymptotic properties; however, because the form of the log-likelihood function does not have a closed form, we have to choose an alternative estimation method. The quasi maximum likelihood estimation method is a suitable suggestion for this aim. To determine this estimate, it is required to build the quasi log-likelihood function. This function depends on the marginal first and second order moments of the response variable. These moments do not have an explicit closed form in the binary and ordinal mixed effects models, either, and have to be approximated, too. In contrast to that in the longitudinal Poisson regression model the required moments have an explicit analytical Under sufficient conditions for the quasi maximum likelihood estimate of the parameters, we aim to achieve the D-optimum designs for the experimental settings. Therefore, we construct the quasi Fisher information matrix and establish the corresponding D-optimality criterion. In the binary and ordinal mixed effects models, the quasi Fisher information matrix lacks the analytical form. Hence, we approximate it for particular cases of the models. On the other hand, in the longitudinal Poisson regression model, the quasi Fisher information matrix has the closed form and can be used directly.

Finally, the D-optimum designs based on the quasi Fisher information matrix are computed and their sensitivity is investigated with respect to various values of model parameters is investigated. Further, in the longitudinal Poisson model, an equivalence theorem for the evaluation of D-optimum designs is derived and the efficiency of D-optimum designs with respect to parameter misspecification is computed. In this model, these designs are quite robust over the settings considered. In contrast, in binary mixed effects model they are truly sensitive with respect to some changes of model parameters. In ordinal mixed effects model, the two point D-optimum designs are transformed to one point D-optimum design under some initial values of model parameters.

Zusammenfassung

Verallgemeinerte lineare und nichtlineare Modelle mit gemischten Effekten werden in vielen Anwendungsbereichen wie Psychologie, Medizin, Ingenieurwesen usw. verwendet. Die Struktur der Daten in diesen Modellen ist longitudinal. Das bedeutet, dass dass wiederholt Daten innerhalb jedes Subjekts erhoben werden und damit die Daten innerhalb jedes Subjekts korreliert und zwischen den Subjekten unkorreliert sind. In dieser Dissertation untersuchen wir das binäre und ordinale Regressionsmodell mit gemischten Effekten, bei dem die Antwortvariable zwei beziehungsweise mehr als zwei Stufen umfasst. Außerdem wird ein nichtlineares longitudinales Poisson-Regressionsmodell betrachtet, um die Entwicklung Fähigkeit von Probanden über einen längeren Zeitraum zu testen.

Für die Schätzung der Modellparameter wird üblicherweise, die Maximum-Likelihood-Schätzung der Parameter aufgrund ihrer guten asymptotischen Eigenschaften verwendet. Da die Log-Likelihood-Funktion jedoch keine geschlossene Darstellung besitzt, müssen wir ein alternatives Schätzverfahren wählen. Die Methode der Quasi-Maximum-Likelihood-Schätzung ist ein geeignetes Verfahren für diesen Zweck. Um diese Schätzung zu bestimmen, ist es erforderlich, die Quasi-Log-Likelihood-Funktion aufzustellen. Diese Funktion ist abhängig von den marginalen Momenten erster und zweiter Ordnung der Antwortvariablen. Diese Momente haben in den binären und ordinalen Modellen mit gemischten Effekten ebenfalls keine explizite geschlossene Form. Im Gegensatz hierzu besitzen im betrachteten longitudinalen Poisson-Regressionsmodell die benötigten Momente eine explizite analytische Form, so dass die Quasi-Likelihood-Funktion direkt angegeben werden kann. Unter Annahmen für die Quasi-Maximum-Likelihood-Schätzung der Parameter können wir dann D-optimale Designs bestimmen. Dafür konstruieren wir die Quasi-Fisher-Informationsmatrix und führen das entsprechende D-Optimalitätskriterium ein. Bei binären und ordinalen Modellen mit gemischten Effekten hat die Quasi-Fisher-Informationsmatrix keine analytische Form; daher approximieren wir diese für einige Spezialfälle dieser Modelle. Die Quasi-Fisher-Informationsmatrix für das longitudinale Poisson-Regressionsmodell eine geschlossene Form was eine direkte bestimmung erlaubt.

Schließlich werden das D-optimale Design auf der Grundlage dervorliegenden Ergebnisse bestimmt und deren Sensitivität in Bezug auf verschiedene Werte der Modellparameter untersucht. Ferner wird für das longitudinale Poisson-Modell ein Äquivalenzsatz zur Valisierung D-optimalen Designs aufgestellt und die Effizienz optimaler Designs unter Misspezifikation der Modellparameter berechnet. In diesem Modell sind diese Designs über die betrachteten Einstellungen ziemlich robust. Im Gegensatz, in binären Mischwirkungen modellieren sie sind in Bezug auf einige Änderungen von Musterrahmen aufrichtig empfindlich. Im gemischten Wirkungsmodell der Ordnungszahl werden die zwei Punkt-D-Optimum-Designs in ein Punkt-D-Optimum-Design unter einigen Anfangswerten von Musterrahmen umgestaltet.

Dedication

To my family and friends

Contents

Chapter 1

Introduction

Generalized linear mixed effects models and nonlinear mixed effects models are applied in different fields of applications; such as psychology, medicine, etc. The data have clustered structure form, which means that the observations within the clusters are correlated and between the clusters are uncorrelated. Therefore, there is over-dispersion among outcomes and we use the clusters as independent subjects which impacts randomly on the response variable. In addition, there are fixed unknown parameters among all subjects, which are the coefficients of the explanatory variables. They affect on the response variable and we are interested in their estimation. The explanatory variables are design variables and they are chosen by the experimenter to take observation from. In generalized linear mixed effects models, the response variable does not follow the Normal distribution. In the maximum likelihood analysis of these models, the marginal likelihood function needs numerical integration with respect to the random effects. As the number of the random effects exceeds, the dimension of the integrals increases. Therefore, it requires higher numerical computations, which may decrease the overall certainty. As a result, the maximum likelihood estimation of the model will be gained numerically and the statistician should be concerned about the behavior of the estimation in extreme values of the random effects.

Breslow and Clayton 1993 utilized two procedures for the analysis of the generalized linear mixed models. The first methodology is considered as the penalized quasi-likelihood (PQL) method exploited by Green 1987 in semi-parametric regression analysis. This method is available for GLMM's where the shrinkage estimation of the random effects is on focus (Robinson 1991). The other method is the marginal quasi-likelihood (MQL), which was named by Breslow and Clayton (1993) and was proposed by Goldstein 1991. In this dissertation, we use the quasi-likelihood approach which was firstly represented by Wedderburn 1974. This procedure only needs the marginal first and second order moments of the response variable. The second order moment is required to construct the marginal variance of the response variable. However, this term in some types of GLMM's is not straightforward to achieve and we need some approximation procedures. Since in this dissertation we study two types of generalized linear mixed models, as binary mixed effects regression model and ordinal mixed effects regression model with the probit and logit link functions, the second order moment of the response variable does not have an analytical explicit form. As a result we approximate them with the approximation of the logistic distribution with the standard Normal distribution when considering the logit link functions. This method is the modification of the method in Zeger, Liang, and Albert 1988 used for the computation of the first order moment of the response variable in the binary mixed logistic model. Moreover, there is a final approximation of one dimensional integral with the Simpson's rule considering both link functions. Afterwards this new approximation of the variance is compared with the direct numerical approximation of the variance matrix, which is only based on the numerical Simpson's rule. In addition, in the binary mixed regression model which is less complicated than the ordinal mixed regression model, the third approximation is obtained by approximating the final integrand via the computation. This method will be discussed later in Section 3.1. The interesting point was that this approximation also worked well in the special case of the model in comparison to the direct numerical approximation.

According to the new approximations, we construct the quasi log-likelihood function of the model and maximize it with respect to the parameters. It was informed that in most family of distributions, the quasi log-likelihood behaves like the ordinary log likelihood function. By setting the quasi log-likelihood function equal to zero, the generalized estimating equation is obtained. The quasi maximum-likelihood estimate of the parameters is the root of the considered equation. McCullagh 1983 examined the asymptotic properties of the quasi-likelihood function and showed that the estimators enjoy certain asymptotic optimality properties in some linear regression models. Sutradhar 2004 demonstrated that the quasi maximum-likelihood estimates in GLMM's are consistent and highly efficient.

The third type of model, which is considered in this dissertation, is a nonlinear longitudinal Poisson regression model. It has the closed form of the first and second order moments of the response variable. Therefore, the quasi maximum likelihood estimate is based on the explicit forms.

One type of optimum design is the D-optimum design. It is obtained from the maximization of the D-optimality criterion (Silvey 1980) which is the determinant of the Fisher information matrix. The aim of this dissertation is to achieve D-optimum design for the considered three types of models. It is actually needed to calculate the Fisher information matrix at first. In general the Fisher information matrix is the measure of the amount of information the observable response variable carries about the unknown model parameters. One interest is to maximize the information of parameters. Then, one recommendation would be to take the determinant of the Fisher information matrix and maximize it with respect to covariates (experimental settings) and the weights relevant to each level of the covariate in the model. More generally, the Fisher information matrix includes design which covers the experimental settings (support points) and their relevant weights. The definition, theory and application of optimum design and its types can be found in V. V. Fedorov 1972 and Silvey 1980. Actually, there is no explicit closed form of the Fisher information matrix in the considered models. The reason for this is that in the longitudinal Poisson regression, the log-likelihood function of the response lacks an explicit form. This problem is discussed later in Section 2.1. Furthermore, in the binary and ordinal mixed regression models, the marginal likelihood function lacks the analytical closed form and hence the Fisher information lacks the analytical closed form as well.

Breslow and Clayton 1993 suggested an approximation of the Fisher information matrix based on the PQL method. Also, Mielke and Schwabe 2010 noted the approximation of the Fisher information matrix based on the linearization of the model function in the fixed effects of the mixed effects model. They showed that this approximation is not reliable. Quasi Fisher information matrix is the other approximation of the Fisher information. It is based on the quasi log-likelihood function and it was introduced firstly by Wedderburn 1974. The quasi Fisher information matrix is asymptotically equivalent to the inverse of variance of the quasi maximum-likelihood estimate of the model parameters (P. McCullagh and Nelder 1983). The relevant literature is found in P. McCullagh and Nelder 1983, Chapter 9. As it was mentioned above, the quasi likelihood function and hence the construction of the quasi Fisher information matrix needs the computation of the first and the second order moments of the response variable. There are several articles who considered the quasi Fisher information matrix for the type of generalized linear mixed effects (GLMM's) model. Niaparast 2010 noticeably utilized the quasi Fisher information matrix in the Poisson regression model with one and two random effects, which went beyond its use in the generalized linear models. Niaparast and Schwabe 2013 obtained the quasi Fisher information matrix for the estimation of parameters in the Poisson regression with random coefficients.

The structure of the dissertation is determined as follows: 1) In chapter 2, the three models, as binary mixed effects model, ordinal mixed effects model and nonlinear longitudinal Poisson regression model are introduced, and their properties are revealed. Moreover, different methods of parameter estimation, such as the maximum likelihood estimate, the generalized least squares estimate and the quasi maximum likelihood estimate are represented. Then, the quasi Fisher information

matrix, definition of design and the general equivalence theorem to evaluate the optimum design are stated.
2) In chapter 3, the binary mixed effects regression model, its further properties to form the quasi Fisher information matrix is obtained. The quasi Fisher information matrix for the binary mixed effects regression model with one and two random effects is approximated and in the random intercept binary regression model the D-optimum design with treatment-control group is computed for different values of the model parameters.
3) In chapter 4, the ordinal mixed effects regression model, its further properties to form the quasi Fisher information matrix is obtained. The quasi Fisher information matrix for the random intercept ordered regression model is approximated and the D-optimum design with control-treatment group for different values of the model parameters is subsequently obtained.
4) In chapter 5, the nonlinear longitudinal Poisson regression model, its further properties to form the quasi Fisher information matrix is obtained. The quasi Fisher information matrix is formulated, its properties are achieved and the general equivalence theorem is constructed to evaluate the D-optimum design. Finally, the D-optimum design is calculated for different initial values of model parameters and the D-efficiency criterion is calculated.
5) Chapter 6 is dedicated to the conclusion and further research related to the topic of the dissertation and the general possible methods would be proposed.

The three types of models are introduced briefly in the next subsections and their general applications are discussed.

Binary mixed effects regression model

The first considered model in the current project is the binary mixed regression model. This model is used to analyze binary data in the sense of calculation and estimation of the probability of success coded as: 1 and correspondingly the calculation of the probability of failure coded as: 0. This model can be used in different applications; such as the outcome of an experiment in consideration of remedy for such disease or the success of a particular method of teaching used at a school to enhance the efficiency of learning among students. The considered probability may depend on such population parameters which are unknown and fixed among all observations. However, there may be random effects which impact on the probability of success randomly. This effect is actually a latent variable which means that it is unknown to the experimenter. It may be due to some unknown features regarding each subject (patients, students) or each cluster (clinic, schools). Therefore, the experimenter assume a probability distribution for the considered parameter. In chapter 3, we assume Normal distribution for the random effect as the traditional methodology. We deal with the general form of the binary mixed effects regression model.

More information regarding the general form of the model is stated in Chapter 2. There are some works in the field of optimum design in the noted model. Tekle, Tan, and Berger 2008 obtained maximin D-optimum designs for binary longitudinal responses. They used the approximate Fisher information matrix under the penalized quasi-likelihood. Tommasi, Rodriguez-Diaz, and Santos-Martin 2014 constrained the assumptions on the repeated measurement logistic regression model. In other words, they assumed one observation within each cluster. Then, they calculated the Fisher information matrix. Also, they proved that the Fisher information matrix for the random effect logistic regression model is equivalent to the Fisher information matrix in the linearized model. The corresponding Fisher information matrix depends on some integrals. They obtained some algebraic approximations for these integrals, which are consistent. At the end, they computed D-, A- and c- optimum designs and also the optimum design to estimate the percentile. Abebe et al. 2014 obtained Bayesian D-optimum designs based on the Fisher information matrix which is approximated based on two approaches. The first one was the first order penalized quasi-likelihood and the second one was based on the extended version of the generalized estimating equation. Seurat, Nguyen, and Mentré 2020 computed several optimality criteria based on Fisher information matrix evaluated by the new method of Monte-Carlo/Hamiltonian Monte-Carlo.

In this dissertation, D-optimum design is calculated numerically based on the new approximation of the quasi Fisher information matrix in the random intercept binary model which has been mentioned previously.

Ordinal mixed effects regression model

The second type of model in this dissertation is the ordinal mixed effects regression model. The analysis of the ordinal variables have been considered in the second have of the 20th century. When the data are categorized or more clearly when an unobservable latent variable is linearly a function of a categorized observable random variable, it is of interest to analyse the categorized data. Since there is the transformation between the latent variable and the discrete random variable, the latter variable is ordinal (Winship and Mare 1984). In experiments, there are covariate(s) affecting the latent random variable, which means as the covariate changes in one scale how much will be the impact on the latent response variable. This variable is classified by thresholds and creates the sets of increments. As each increment happens, the observed ordered random variable with the index equivalent to the index of the increment takes 1 and the other categories of the ordered random variables with different index rather than the observed increment, take zero value. The experimenter is interested in the estimation of the probability of each increment. These probabilities depend on the covariates. Their effect on the probability is specified by parameters. The relation between the probabilities and the linear predictors is determined by a link function. Based on the choice of the logit or the probit function, the corresponding models are built as ordered probit and ordered logit models, respectively (Agresti 2003). Moreover, the models including the random effects in addition to the fixed effects are called ordinal mixed effects models. They include some effects which cannot be controlled by the experimenter. In this dissertation, we consider the ordinal models with random effects, and analyze them.

Winship and Mare 1984 reviewed methods of the analysis of mixtures of ordered, dichotomous, and continuous variables in structural equation models and used maximum likelihood estimation procedure. Tutz and Hennevogl 1996 considered the ordinal regression model including random effects in the linear predictor. They used three alternative estimation procedures according to the EM type algorithm and applied numerical integration techniques (Gauss-Hermit or Monte Carlo), and a EM type algorithm based on posterior modes. Agresti and Lang 1993 considered a proportional odds model with subject-specific effect for repeated ordered categorical responses. They found that the unconditional maximum likelihood estimate for the parameters is inconsistent. Therefore, they used the conditional maximum likelihood estimation of the parameters, given sufficient statistics for the random effect.

Nonlinear longitudinal Poisson regression model

The third type of model is named as nonlinear longitudinal Poisson regression model. In order to introduce the model, consider an experiment where each subject is tested several times using the same instrument, e.g. an intelligence test or a test for working memory. The results of the experiment enhances as the time increases. However, this reaches a plateau. Provided that a task is given at several times to each subject, then, the number of correct answers (scores) to the questions at each time could be of importance. In other words, we are interested in the prediction of the mean score of the number of correct answers at each time. This term is called the mean of the response variable. The ability of the subjects to do the task is rising up exponentially. Also, the expected number of scores is an exponential function of the ability parameter (See section 2.1). Each subject and each time affect as a random effect on the expectation of the response variable.

The response variable conditioned on the random ability parameter is assumed to follow the Poisson distribution. This model is called the longitudinal Poisson regression model. More details on this model and its properties are found in Section

2.1 and Chapter 5.

Chapter 2

Nonlinear and generalized linear mixed effects models

Generalized linear and Nonlinear mixed effects models have been widely used in many fields, such as biomedical medicine, psychology, etc. (see the works of Sheiner and Beal 1983, Bono, Alarcón, and Blanca 2021). In the literature, two groups of regression models with random effects have been used; the first is the generalized linear mixed effects model and the second is the Nonlinear mixed effects models. The types of the two models and their properties are specified in this chapter. Additionally, the definition of the quasi Fisher information matrix, the design and the relevant equivalence theorem to assess the D-optimality of D-optimum designs are represented briefly with reference to the literature.

2.1 Model specification

To introduce the models respectively, consider N individuals, $i = 1, ..., N$. J settings, as $\boldsymbol{x}_{ij}$'s or time points t_{ij}' s, where $j = 1, ..., J$. Also, n_{ij} is the number of replications ($k = 1, ..., n_{ij}$) of observations y_{ijk} as a realization for random variable Y_{ijk} for individual i at setting $\boldsymbol{x}_{ij}$. The number of observations within each individual is denoted as $n_i = \sum_{j=1}^{J} n_{ij}$.

Y_{ijk} is considered as the response variable for the kth replication at jth setting $\boldsymbol{x}_{ij}$ for subject i. Also, let $\boldsymbol{\zeta}_i = (\zeta_{i0}, ..., \zeta_{i(q-1)})^\top$ be the random effect with q dimension. We define $\boldsymbol{h}(\boldsymbol{x}) = (1, h_1(\boldsymbol{x}), ..., h_{q-1}(\boldsymbol{x}))^\top$ as the continuous function of experimental setting $\boldsymbol{x}$ regarding the random effect. We also assume that $\boldsymbol{\zeta_i} \sim F_{\boldsymbol{\zeta}}$; and Y_{ijk} given $\boldsymbol{\zeta}_i$'s are (conditionally) independent given $\boldsymbol{\zeta}_i$. Moreover, $\boldsymbol{f}(\boldsymbol{x}) = (1, f_1(\boldsymbol{x}), ..., f_{p-1}(\boldsymbol{x}))^\top$ is the continuous function of the experimental setting x

regarding the population parameter $\boldsymbol{\beta} = (\beta_0, ..., \beta_{p-1})^\top$ with p-dimension. $\mathfrak{B}$ is denoted as the parameter space of all $\boldsymbol{\beta}$.

Then, by setting $\eta_{ij} = \boldsymbol{f}^\top(\boldsymbol{x}_{ij})\boldsymbol{\beta} + \boldsymbol{h}^\top(\boldsymbol{x}_{ij})\boldsymbol{\zeta}_i$, as the individual linear component, the generalized linear mixed model is formulated as follows:

$$g(\mathrm{E}(Y_{ijk} \mid \boldsymbol{\zeta}_i = \boldsymbol{z}_i)) = \eta_{ij}; \tag{2.1}$$

where g is the link function. For a realization of $\boldsymbol{\zeta}_i$ at $\boldsymbol{z}_i$ means $g : \mathrm{E}(Y_{ijk} \mid \boldsymbol{\zeta}_i = \boldsymbol{z}_i) \to g(\mathrm{E}(Y_{ijk} \mid \boldsymbol{\zeta}_i = \boldsymbol{z}_i)) = \eta_{ij}$, that $\mathrm{E}(Y_{ijk} \mid \boldsymbol{\zeta}_i = \boldsymbol{z}_i)$ can be any potential mean value of Y_{ijk}, such as μ. Then, g has to be defined on the whole range of μ. This model is the general form of the models used in Chapters 3 and 5, where the response variable is a univariate random variable. However, in Chapter 4, which states the ordinal mixed effects models, we work with a multivariate response variable.

Let h be the inverse of link function g. Then, model (2.1), is reformed in the following form as:

$$\mathrm{E}(Y_{ijk} \mid \boldsymbol{\zeta}_i) = h(\eta_{ij}). \tag{2.2}$$

Two types of generalized linear mixed effects models as the binary and ordinal mixed effects models are discussed in the subsequent sections.

Binary mixed effects regression model

In binary models the response Y_{ijk} may only take values 0 and 1 which are often interpreted as "success" ($Y_{ijk} = 1$) or "failure" ($Y_{ijk} = 0$). In order to form the binary mixed effects regression model, let $Y_{ijk} \mid \boldsymbol{\zeta}_i \sim Ber(p_{ij}(\boldsymbol{\zeta}_i))$, where $p_{ij}(\boldsymbol{\zeta}_i) = P(Y_{ijk} = 1 \mid \boldsymbol{\zeta}_i)$. According to the choice of the link function g, the probit and the logit binary mixed effects models are used. Let the unobservable latent variable U_{ijk} be defined as:

$$U_{ijk} = \eta_{ij} + \epsilon_{ijk}, \tag{2.3}$$

as the linear mixed effects model, where ϵ_{ijk}'s are identically and independently distributed, with constant dispersion parameter $\sigma_\epsilon^2 = 1$.

Actually the link function is implied by the distribution of ϵ_{ijk}, stated in formula (2.3). The response variable $Y_{ijk} = 1$, if $U_{ijk} \geq 0$ and $Y_{ijk} = 0$, if $U_{ijk} < 0$. More precisely, the inverse link function is equal to the distribution function of the ϵ_{ijk}. In particular,

1) As $g(\mu) = \Phi^{-1}(\mu)$, the binary mixed effects model with the probit link function is assumed. This situation is when ϵ_{ijk} follows the Normal distribution.

2) As $g(\mu) = \log(\frac{\mu}{1-\mu})$, the binary mixed effects model with the logit link function

is assumed. This situation is when ϵ_{ijk} follows the Logistic distribution (Agresti 2003).

The two link functions are applied when the probability of success against the covariate rises (falls) slowly for small and high values of the covariate and rise (falls) fast in the corresponding middle values.

In these models, we assume $\boldsymbol{\zeta_i} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

In order to show the visual description of the special case of these models, let $\boldsymbol{\zeta}_i = \zeta_{i0}$, $\boldsymbol{f}(\boldsymbol{x}) = (1, \boldsymbol{x})^\top$, where $\boldsymbol{x}$ be one dimensional, $\boldsymbol{x} \in \mathbb{RR}$. $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top = (-1, 1)^\top$, which means $p = 2$, $q = 1$. $g(\mu) = \Phi^{-1}(\mu)$; then, the model is reduced to the random intercept probit regression model as:

$$g(\mathrm{E}(Y_{ijk} \mid \zeta_{i0})) = \Phi^{-1}(\mathrm{E}(Y_{ijk} \mid \zeta_{i0})) = \eta_{ij}, \tag{2.4}$$

where $\eta_{ij} = \beta_0 + \beta_1 \boldsymbol{x}_{ij} + \zeta_{i0}$; therefore, the conditional probability of success is defined as:

$$P(Y_{ijk} = 1 \mid \zeta_{i0}) = \Phi(\beta_0 + \beta_1 x_{ij} + \zeta_{i0}). \tag{2.5}$$

$\zeta_{i0} \sim N(0, \sigma^2)$; $\sigma^2 = 1$, for standardization. The conditional probability curves (mean response curves) for different individuals indicated by different values of ζ_{i0} against $\boldsymbol{x}$ are illustrated in Figure 2.1. For the realization of the random intercepts $\zeta_{10} = -1.5, \zeta_{20} = -1, \zeta_{30} = -.5, \zeta_{40} = 0, \zeta_{50} = .5, \zeta_{60} = 1, \zeta_{70} = 1.5$, From right to left, the right most red curve denotes the conditional cumulative probability conditioned on $\zeta_{10} = -1.5$, and the largest random effect is $\zeta_{70} = 1.5$, which is shown with brown curve. As the value of the random effect increases, the function is shifted to the left.

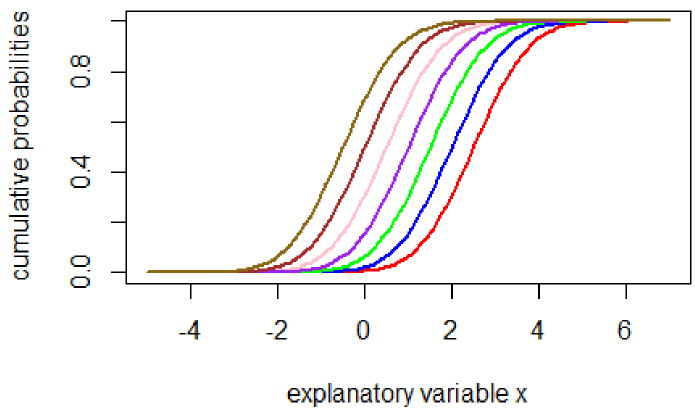

Figure 2.1: The inverse probit link function, $P(Y_{ijk} = 1 \mid \boldsymbol{\zeta}_i)$, against $\boldsymbol{x}$.

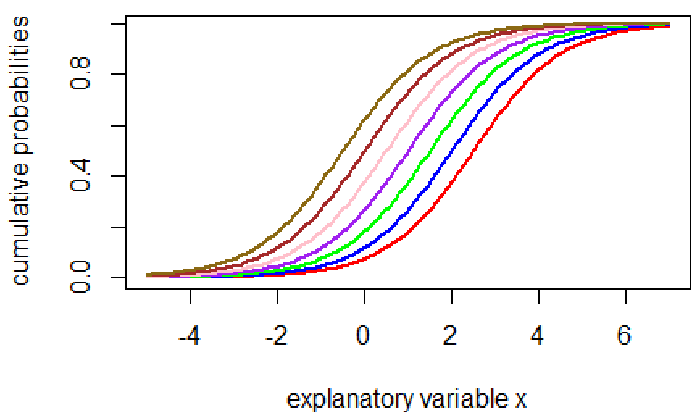

Figure 2.2: The inverse logit link function, $P(Y_{ijk} = 1 \mid \boldsymbol{\zeta}_i)$, against $\boldsymbol{x}$.

Moreover, considering the link function as:

$$g(\mathrm{E}(Y_{ijk} \mid \zeta_{i0})) = \log\left(\frac{p_{ij}(\zeta_{i0})}{1 - p_{ij}(\zeta_{i0})}\right), \tag{2.6}$$

the conditional probability curve $p_{ij}(\zeta_{i0})$ against x_{ij} is shown in Figure 2.2. It shows as the value ζ_{i0} increases, the function is shifted to the left".

We are interested in the estimation of the probability curve and also the estimation of the marginal first and second order moments of response variable.

The marginal first and second order moments of the response variable are shown as:

$$\mathrm{E}(Y_{ijk}) = \int_{\mathbb{R}^q} p_{ij}(\boldsymbol{\zeta}_i)\mathrm{f}(\boldsymbol{\zeta}_i)d\boldsymbol{\zeta}_i, \tag{2.7}$$

$$\mathrm{Var}(Y_{ijk}) = \mathrm{E}(Y_{ijk}^2) - \mathrm{E}^2(Y_{ijk}), \tag{2.8}$$

where $\mathrm{f}(\boldsymbol{\zeta}_i)$ is the density function of $\boldsymbol{\zeta}_i$.

In the binary probit mixed model, equations (2.7) and (2.8) have explicit closed forms (Zeger, Liang, and Albert 1988); however, $\mathrm{cov(Y_{ijk}, Y_{ijk'})}; k \neq k'$ in (3.3) and $\mathrm{cov(Y_{ijk}, Y_{ij'k'})}, j \neq j', \forall k, k'$ in (3.4) do not have any closed form and it needs to be approximated. In the Logistic mixed effects model, both first and second order moments of Y_{ijk} lacks the analytical closed form. The approach of the achievement of the above terms in the considered models is stated completely in Section 3.1.

To define the marginal likelihood function, let the vector of n_{ij} observations $\boldsymbol{y}_{ij} = (y_{ij1}, ..., y_{ijn_{ij}})^\top$ of subject i at setting $\boldsymbol{x}_{ij}$. All n_i observations $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^\top, ..., \boldsymbol{y}_{iJ}^\top)^\top$ of subject i, all $n = \sum_{i=1}^N n_i$ observations $\boldsymbol{y} = (\boldsymbol{y}_1^\top, ..., \boldsymbol{y}_N^\top)^\top$ over all subjects. Then,

$$L(\boldsymbol{\beta} \mid \boldsymbol{y}_i) = \int_{\mathbb{R}^q} \prod_{j=1}^{J} \prod_{k=1}^{n_{ij}} \mathrm{f}(y_{ijk} \mid \boldsymbol{\zeta}_i)\mathrm{f}(\boldsymbol{\zeta}_i)\mathrm{d}\boldsymbol{\zeta}_i, \tag{2.9}$$

where $\mathbb{R}^q$ is the defined domain of $\boldsymbol{\zeta}_i$ and $\mathrm{f}(y_{ijk} \mid \boldsymbol{\zeta}_i) = (p_{ij}(\boldsymbol{\zeta}_i))^{y_{ijk}}(1-p_{ij}(\boldsymbol{\zeta}_i))^{1-y_{ijk}}$ is the conditional density of Y_{ijk} conditioned on $\boldsymbol{\zeta}_i$; hence, the population marginal likelihood function is denoted as:

$$L(\boldsymbol{\beta} \mid \boldsymbol{y}) = \prod_{i=1}^{N} L(\boldsymbol{\beta} \mid \boldsymbol{y}_i), \tag{2.10}$$

In fact, in these models the marginal likelihood function may not have the explicit closed form, and we need to compute the integrals numerically.

Numerical integration

One numerical method of integration is the quadrature rule based on interpolating functions that are easy to integrate. If we are interested in integrating a well behaved function $s(.)$, (i.e.: piecewise continuous and bounded variation) against $t \in [a_1, a_2]$, where a_1 and a_2 are constant, the composite trapezoidal rule can be stated, which is denoted as follows (Jefferson 2019):

$$\int_{a_1}^{a_2} s(t)\mathrm{d}t \approx \frac{a_2 - a_1}{n}[\frac{s(a_1)}{2} + \sum_{k=1}^{n-1}(s(a_1 + k\frac{a_2 - a_1}{n})) + \frac{s(a_2)}{2}], \tag{2.11}$$

where we have assumed that the interval $[a_1, a_2]$ is broken up into n subintervals. The subintervals have the form $[a_1 + kh, a_1 + (k+1)h] \in [a_1, a_2]$, with $h = \frac{a_2 - a_1}{n}$ and $k = 1, ..., n-1$. Here, we used subintervals of the same length h but one could also use intervals of varying length $(h_k)_k$.

An alternative method for the numerical integration is the Simpson's rule. The Simpson's rules are several approximation for definite integrals, which we show two of them here. The most basic rule is called Simpson's 1/3 rule, or just Simpson's rule,

$$\int_{a_1}^{a_2} s(x)\mathrm{d}x \approx \frac{a_2 - a_1}{6}[s(a_1) + 4s(\frac{a_1 + a_2}{2}) + s(a_2)]. \tag{2.12}$$

Now, suppose that the interval $[a_1, a_2]$ is split up into n subintervals, with n an even number. Then, the composite Simpson's rule is given by:

$$\begin{aligned} \int_{a_1}^{a_2} s(x)\mathrm{d}x &\approx \tfrac{h}{3}\textstyle\sum_{j=1}^{n/2}[s(x_{2j-2}) + 4s(x_{2j-1}) + s(x_{2j})] \\ &= \frac{h}{3}[s(x_0) + 2\sum_{j=1}^{n/2-1} s(x_{2j}) + 4\sum_{j=1}^{n/2} s(x_{2j-1}) + s(x_n)], \end{aligned} \tag{2.13}$$

where $x_j = a_1 + jh$ for $j = 0, 1, ..., n-1, n$ with $h = \frac{a_2 - a_1}{n}$; $x_0 = a_1$ and $x_n = a_2$.

Moreover, Monte Carlo methods and quasi-Monte Carlo methods are easy to apply to multi-dimensional integrals. They may yield greater accuracy for the same number of function evaluations than repeated integration using one-dimensional methods.

In this dissertation, we use the Simpson's rule to compute the integral, where necessary, since we only come across one dimensional integrals in Sections 3.1, 4.1.

In the next subsection, the ordinal mixed effects regression model is introduced.

Ordinal mixed effects regression model

In the ordinal mixed effects regression model, the response variable is presumed to be ordinal with M categories based on an unobservable latent variable U_{ijk}, which is defined in (2.3).

The response variable is categorized by γ_m; $m = 1, ..., M-1$ thresholds; such that, $\gamma_1 < ... < \gamma_{M-1}$. Also, $\gamma_0 = -\infty$ and $\gamma_M = +\infty$.

Denote the M-dimensional random vector as $\boldsymbol{Y}_{ijk} = (Y_{ijk}^{(1)}, ..., Y_{ijk}^{(M)})^\top$, where

$$Y_{ijk}^{(m)} = \begin{cases} 1 & ; \quad \gamma_{m-1} \leq U_{ijk} < \gamma_m, \\ 0 & ; \quad o.w., \end{cases} \tag{2.14}$$

such that $m = 1, ..., M$. In fact, each $\boldsymbol{Y}_{ijk}$ is a vector consisting of one element equal to 1 and other elements equal to 0. The $\boldsymbol{Y}_{ijk}$ here is already a mathematical construct. In practice, people do not observe the vector $\boldsymbol{Y}_{ijk}$, but equivalently the value of the category, i.e. a (one-dimensional) random variable $\tilde{Y}_{ijk}$ which may attain values $m \in \{1, ..., M\}$. Then, the observations $\tilde{Y}_{ijk}$ are related to the latent variable U_{ijk} by the condition $\tilde{Y}_{ijk} = m$ if $\gamma_{m-1} \leq U_{ijk} < \gamma_m$. The combination of these two relations yield (2.14).

Moreover, $Y_{ijk} \mid \boldsymbol{\zeta}_i \sim Multinomial(p_{ij}^{(1)}(\boldsymbol{\zeta}_i), ..., p_{ij}^{(M)}(\boldsymbol{\zeta}_i))$; where

$$p_{ij}^{(m)}(\boldsymbol{\zeta}_i) := \mathrm{E}(Y_{ijk}^{(m)} \mid \boldsymbol{\zeta}_i), \tag{2.15}$$

with $\mathrm{E}(Y_{ijk}^{(m)} \mid \boldsymbol{\zeta}_i) = P(Y_{ijk}^{(m)} = 1 \mid \boldsymbol{\zeta}_i)$. Also, $\sum_{m=1}^{M} p_{ij}^{(m)}(\boldsymbol{\zeta}_i) = 1$.

Now, define

$$\delta_{ij}^{(m)}(\boldsymbol{\zeta}_i) = \gamma_m - \eta_{ij}. \tag{2.16}$$

Also, $Y_{ijk}^{(m)} = 1$ means that,

$$\gamma_{m-1} \leq U_{ijk} < \gamma_m \tag{2.17}$$

$$\gamma_{m-1} \leq \eta_{ij} + \epsilon_{ijk} < \gamma_m \tag{2.18}$$

$$\delta_{ij}^{(m-1)}(\boldsymbol{\zeta}_i) \leq \epsilon_{ijk} < \delta_{ij}^{(m)}(\boldsymbol{\zeta}_i), \tag{2.19}$$

where $\delta_{ij}^{(m)}(\boldsymbol{\zeta}_i) = \gamma_m - \eta_{ij}$. Then,

$$P(Y_{ijk}^{(m)} = 1 \mid \boldsymbol{\zeta}_i) = P(\epsilon_{ijk} < \delta_{ij}^{(m)}(\boldsymbol{\zeta}_i)) - P(\epsilon_{ijk} < \delta_{ij}^{(m-1)}(\boldsymbol{\zeta}_i)). \tag{2.20}$$

if g^{-1} is the distribution function of ϵ_{ijk}; then, this construction implies

$$p_{ij}^{(m)}(\boldsymbol{\zeta}_i) = g^{-1}(\delta_{ij}^{(m)}(\boldsymbol{\zeta}_i)) - g^{-1}(\delta_{ij}^{(m-1)}(\boldsymbol{\zeta}_i)). \tag{2.21}$$

In the special case $M = 2$, $\gamma_1 = 0$ leads to the binary case, which is treated in the previous subsection. Considering the special case of the ordered probit mixed effects regression model, as the random intercept model, let $M = 3$, $\boldsymbol{\zeta}_i = \zeta_{i0}$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top = (-1, 1)^\top$. For realization of random effect $\zeta_{10} = -.5, \zeta_{20} = 0, \zeta_{30} = .5$, Figure 2.3 shows $\Phi(\delta_{ij}^{(m)}(\boldsymbol{\zeta}_i))$ for each $m = 1, 2, 3$. Each plot demonstrates the considered cumulative distribution function conditioned on each ζ_{i0}. Figure 2.4 reveals the conditional probability $P(Y_{ijk}^{(m)} = 1 \mid \zeta_{i0})$; $m = 1, 2, 3$ for the same ζ_{i0}.

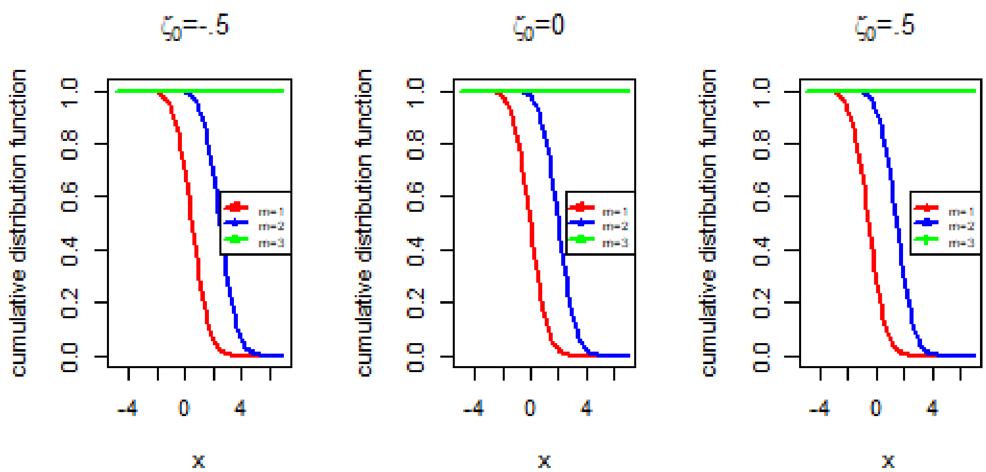

Figure 2.3: Cumulative probability curve, against $\boldsymbol{x}_{ij}$, the inverse probit link function.

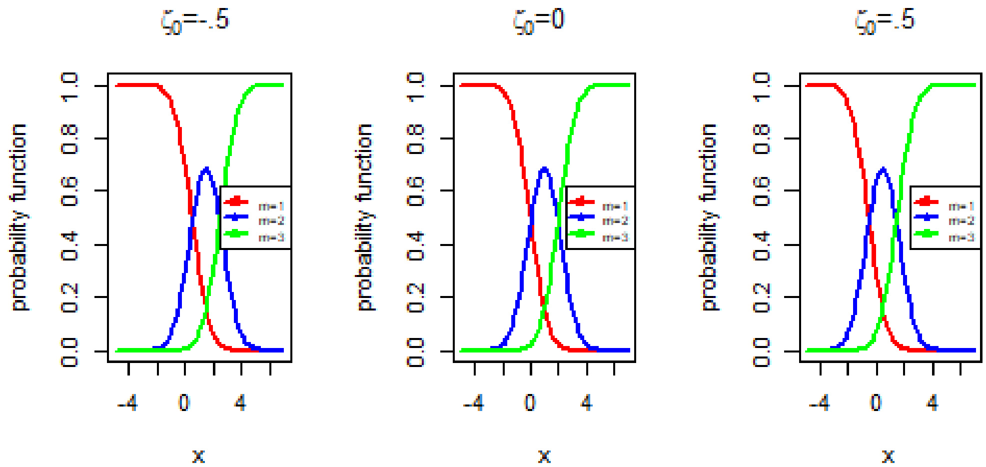

Figure 2.4: Conditional probability curve, $P(Y_{ijk}^{(m)} = 1 \mid \boldsymbol{\zeta}_i)$, against $\boldsymbol{x}_{ij}$, the inverse probit link function.

We are interested in the estimation of each curve. Therefore, it is required to estimate parameter $\boldsymbol{\beta}$. Moreover, similar to the binary mixed model, we are interested in the estimation of the marginal first and second order moments of the response variable. They are shown as:

$$\begin{aligned}\boldsymbol{\pi}_{ij} = \mathrm{E}(\boldsymbol{Y}_{ijk}) &= (\mathrm{E}(Y_{ijk}^{(1)}), \mathrm{E}(Y_{ijk}^{(2)}), ..., \mathrm{E}(Y_{ijk}^{(M)}))^\top \\ &= (\pi_{ij}^{(1)}, \pi_{ij}^{(2)}, ..., \pi_{ij}^{(M)})^\top, \qquad (2.22)\end{aligned}$$

where

$$\pi_{ij}^{(m)} = \int_{\mathbb{R}^q} \mathrm{E}(Y_{ijk}^{(m)} \mid \boldsymbol{\zeta}_i) \mathrm{f}(\boldsymbol{\zeta}_i) \mathrm{d}\boldsymbol{\zeta}_i. \qquad (2.23)$$

$$\mathrm{Var}(\boldsymbol{Y}_{ijk}) = [\mathrm{cov}(Y_{ijk}^{(m)}, Y_{ijk}^{(m')})]_{m,m'=1}^{M}, \qquad (2.24)$$

in which $\boldsymbol{\pi}_{ij}$ is with dimension $M \times 1$ and $\mathrm{Var}(\boldsymbol{Y}_{ijk})$ is with dimension $M \times M$. Similar to the binary mixed model, $\boldsymbol{\pi}_{ij}$ has the explicit closed form in the ordered mixed effects probit regression model; however, it does not form an explicit closed form in the ordered mixed effects logit regression model. Variance matrix (2.24) requires approximation in both considered models. More details on the achievement of general mean and variance matrix of $\boldsymbol{Y}$ are asserted in Section 4.1.

In general, the marginal likelihood function for the ordinal mixed effects model is regarded as:

$$L(\boldsymbol{\beta} \mid \boldsymbol{y}_i) = \int_{\mathbb{R}^q} \prod_{m=1}^{M} \prod_{j=1}^{J} \prod_{k=1}^{n_{ij}} [\Phi(\delta_{ij}^{(m)}(\boldsymbol{\zeta}_i)) - \Phi(\delta_{ij}^{(m-1)}(\boldsymbol{\zeta}_i))]^{y_{ijk}^{(m)}} \mathrm{f}(\boldsymbol{\zeta}_i)\mathrm{d}\boldsymbol{\zeta}_i, \qquad (2.25)$$

and

$$L(\boldsymbol{\beta} \mid \boldsymbol{y}) = \prod_{i=1}^{N} L(\boldsymbol{\beta} \mid \boldsymbol{y}_i). \qquad (2.26)$$

The marginal likelihood function lacks the analytical closed form. In order to obtain the integral, it is required to compute them numerically with the procedures which have already been stated in Section 2.1.

The third type of model, which is contemplated in this dissertation is the nonlinear longitudinal Poisson regression model. The definition, application and some properties of this model are stated in the next subsection. The process of the foundation of the model properties is different from the current stated models already.

Nonlinear longitudinal Poisson regression model

Consider a study where there are N subjects. Tasks are given at J time points to each subject. The sequence of the times is denoted as $\{t_j\}_{j=1}^{J}$; and at each time there are n_{ij} items.

At each item in the j^{th} time, the number of correct answers to the questions is of interest. We suppose that it follows the Poisson distribution and it is called the response variable. We are interested in the estimation of the ability of subjects in the response to the questionnaire. Therefore, we observe the outcomes in the line of the time. The mean ability is denoted by $\boldsymbol{\theta}(\boldsymbol{t}, \boldsymbol{\beta}) = (\theta(t_1, \boldsymbol{\beta}), ..., \theta(t_J, \boldsymbol{\beta}))^\top$ and it is presumed to increase and approach exponentially to a saturation level by increasing the time. It is also a function of unknown parameter $\boldsymbol{\beta}$. The attended relation regarding the jth time point is shown in the following:

$$\theta(t_j, \boldsymbol{\beta}) = \beta_0 - \beta_1 \exp\{-\beta_2 t_j\}, \qquad (2.27)$$

where β_0 represents the saturation level, $\beta_0 - \beta_1$ reveals the offset as $t_j \to 0$ and β_2 is the scale parameter. Figure 2.5 shows the mean ability $\theta(t_j, \boldsymbol{\beta})$ against time point t_j. Plot (a) shows the exponential curve of $\theta(t_j, \boldsymbol{\beta})$ for different values of β_2, as (—): $\beta_2 = 1$, (− −): $\beta_2 = 2$, (. . .) :, $\beta_2 = 3$, (− · −), $\beta_2 = 4$, when $\beta_0 = 3$, $\beta_1 = 2$. plot (b) shows the curve $\theta(t_j, \boldsymbol{\beta})$ for different values of β_1, as (—): $\beta_1 = 1$, (− −): $\beta_1 = 1.5$, (. . .) :, $\beta_1 = 2$, (− · −), $\beta_1 = 2.5$, when $\beta_0 = 3$, $\beta_2 = 1$. Plot (a) illustrates that as β_2 increases, the mean ability curve tends to the saturation level faster.

However, plot (b) shows as β_1 rises up, the range $\beta_0 - \beta_1$ as $t_j \to 0$ increases. In fact, the choice of the model parameters depends on the application in use.

In addition, mean ability is not only fixed among subjects and times, but it also varies randomly across subjects and times. The random part of the ability related to the ith subject and jth time point is denoted by Λ_{ij}, and the random effect related to the ith subject is denoted by $\mathbf{\Lambda}_i = (\Lambda_{i1}, ..., \Lambda_{iJ})^\top$. In this dissertation, it is assumed that $\mathbf{\Lambda}_i$ follows a multivariate Gamma distribution with correlated random elements.

The considered multivariate Gamma distribution is built up as follows (Mathai and Moschopoulos 1991):

Let

$$\Lambda_{ij} = S_{i0} + S_{ij}, \tag{2.28}$$

where S_{i0} is a kind of individual block effect being constant over time and S_{ij} is the random effect within the ith individual related to the jth time point. The elements of the sequence $\{S_{ij}\}_{j=0}^{J}$ are considered to be independent.

Suppose

$$S_{i0} \sim \Gamma(c_0, \tau); \tag{2.29}$$

$$S_{ij} \sim \Gamma(c, \tau); \tag{2.30}$$

where $c_0, c \geq 0$, $\tau > 0$. c_0 and c are shape parameter and τ is the common scale parameter. $\mathrm{E}(S_{i0}) = c_0\tau$, $\mathrm{E}(S_{ij}) = c\tau$. $\Lambda_{ij} \sim \Gamma(c_0 + c, \tau)$, with $\mathrm{E}(\Lambda_{ij}) = (c_0 + c)\tau$. The multivariate density function of the random effect $\mathbf{\Lambda}_i = (\Lambda_{i1}, ..., \Lambda_{iJ})^\top$ is obtained in Appendix A, Section A.7. Additionally, we presume the random effects Λ_{ij}'s are centered, which means $\mathrm{E}(\Lambda_{ij}) = 1$, then, it can be reparametrized in terms of the scale parameter τ and the intra-class correlation $\rho = c_0\tau$. Then, $c_0 + c = \frac{1}{\tau}$, and the second order moment of Λ_{ij} is:

$$\mathrm{Var}(\Lambda_{ij}) = \tau, \tag{2.31}$$

$$\mathrm{cov}(\Lambda_{ij}, \Lambda_{ij'}) = \mathrm{Var}(S_{i0}) = \rho\tau; j \neq j'. \tag{2.32}$$

where $\rho = \mathrm{corr}(\Lambda_{ij}, \Lambda_{ij'}); j \neq j'$ is the intraclass correlation.

In general, the longitudinal Poisson regression model for the considered study is constructed as follows:

$$\mathrm{E}(Y_{ijk} \mid \Lambda_{ij} = \lambda_{ij}) = \lambda_{ij} \exp(\theta(t_j, \boldsymbol{\beta}) - \sigma_{ij}), \tag{2.33}$$

Conditionally on the random ability $\Lambda_{ij} = \lambda_{ij}$, the response variable Y_{ijk} follows a Poisson distribution with mean $\mathrm{E}(Y_{ijk} \mid \Lambda_{ij} = \lambda_{ij})$ in (2.33). σ_{ij} is the difficulty

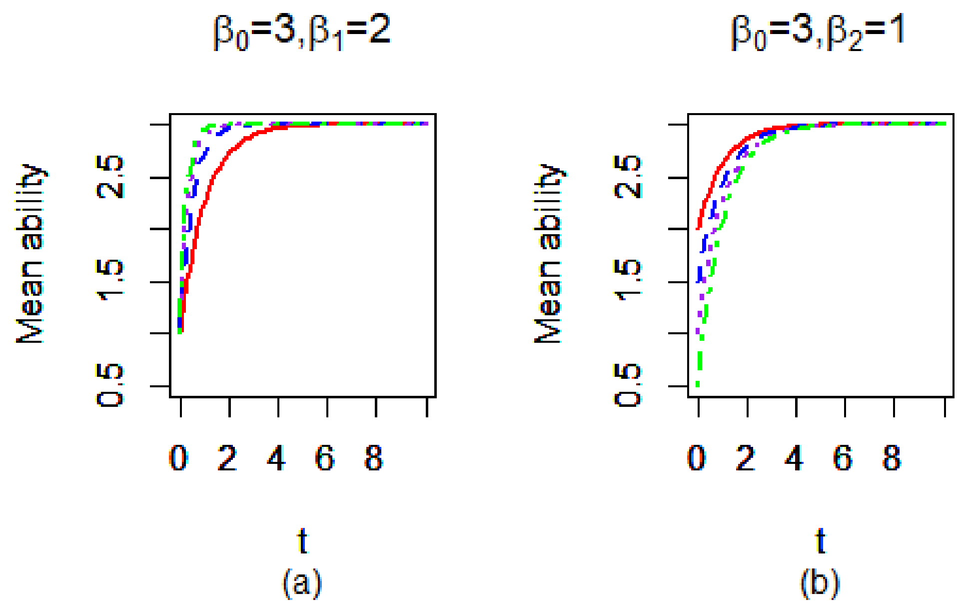

Figure 2.5: Exponential mean ability curves for various value of the parameters.

parameter and it is set to be $\sigma_{ij} = 0$. In this Chapter, we are interested in the estimation of parameter $\boldsymbol{\beta}$. The marginal first and second order moments of the response variable are obtained as follows:

$$\mathrm{E}(Y_{ijk}) = \exp(\theta(t_j, \boldsymbol{\beta})) \tag{2.34}$$

$$\begin{aligned}\mathrm{Var}(Y_{ijk}) &= \mathrm{Var}(\mathrm{E}(Y_{ijk} \mid \Lambda_{ij})) + \mathrm{E}(\mathrm{Var}(Y_{ijk} \mid \Lambda_{ij})) \\ &= \mu_j(\mu_j \tau + 1),\end{aligned} \tag{2.35}$$

where $\mu_j := \mathrm{E}(Y_{ijk})$. More information on the construction of variance matrix of the response variable is given in Chapter 5. The marginal likelihood function for the considered model is stated as:

$$L(\boldsymbol{\beta} \mid \boldsymbol{y}_i) = \int_{\mathbb{R}^J} \prod_{j=1}^{J} \prod_{k=1}^{n_{ij}} \mathrm{f}_{\mathbf{Y}_i \mid \boldsymbol{\Lambda}_i = \boldsymbol{\lambda}_i}(\boldsymbol{y}_i) \mathrm{f}_{\boldsymbol{\Lambda}_i}(\boldsymbol{\lambda}_i) \mathrm{d}\boldsymbol{\lambda}_i, \tag{2.36}$$

and

$$L(\boldsymbol{\beta} \mid \boldsymbol{y}) = \prod_{i=1}^{N} L(\boldsymbol{\beta} \mid \boldsymbol{y}_i), \tag{2.37}$$

where $\boldsymbol{\lambda}_i = (\lambda_{i1}, ..., \lambda_{iJ})^\top$. This integral has an explicit form. The procedure is stated in Appendix A, Section A.8, for $J = 3$. However, the analytical result cannot be applied for the calculation of the maximum likelihood estimation of $\boldsymbol{\beta}$ due to the complexity of the final formulation. In the following section, we state the definition of the maximum likelihood estimation of the model in general.

The difference between this model and the models in the binary and ordinal mixed effects models is that the model in (2.33) is based on the structure of the Nonlinear dependence of $\mathrm{E}(Y_{ijk} \mid \Lambda_{ij} = \lambda_{ij})$ on parameter $\boldsymbol{\beta}$; however, models of binary mixed effects model (2.1) and ordered mixed effects model (2.2) belong to the class of generalized linear mixed effects models.

2.2 Estimation of population parameter

In this section, some methods of parameter estimation, as the maximum likelihood estimation, generalized least squares estimation and the quasi maximum-likelihood estimation are represented. Some of their properties, such as asymptotic behaviour and unbiasedness are discussed. Then, in accordance with the structure and properties of the models, the quasi maximum likelihood estimation method is chosen to be used in the rest of the dissertation.

Maximum likelihood estimation

In accordance with the literature, the maximum likelihood estimate of model parameter is defined as follows:

Definition 2.2.1. *Maximum likelihood Estimation. $\hat{\boldsymbol{\beta}}(\boldsymbol{y})$ is a maximum-likelihood estimate given $\boldsymbol{y}$, if $L(\hat{\boldsymbol{\beta}}(\boldsymbol{y})|\boldsymbol{y}) \geq L(\boldsymbol{\beta}|\boldsymbol{y})$ for all $\boldsymbol{\beta} \in \mathfrak{B}$, where $\mathfrak{B}$ is the space of parameter $\boldsymbol{\beta}$.*

Indeed, maximum likelihood estimate of the parameter is a function over the elements $\boldsymbol{y}$ in the sample space and it involves maximizing the marginal likelihood function in order to search for the probability distribution and parameters that best fits the observed data. In some models the maximum likelihood estimates of the model parameters can be obtained analytically. However, in some other models this aim is not achievable and the numerical computation should be used. In this dissertation, for the models of consideration, the numerical integration techniques are also required, since the integral in (2.9), (2.25) and (2.36) do not have the analytical closed form (For example see the Markov chain Monte Carlo techniques (e.g., Zeger and Karim 1991; Booth and Hobert 1999)). Furthermore, Butler and Louis 1997 established the consistency of the maximum likelihood estimators of the fixed effects and the probability density function of the random effects in the binary mixed effects model. Cramér 1946 and Wald 1949 proved the consistency of maximum likelihood estimator when the observations are independent random samples from the same population, i.e. random samples are independent and identically distributed.

Bradley and Gart 1962 and Hoadley 1971 proved weak consistency when the observations are sampled from independent associated populations, i.e., random samples are independent but not identically distributed. Andrews 1987 established the asymptotic properties of maximum likelihood estimator using uniformly law of large numbers. Nie 2006 presented some easily verifiable conditions for the strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed effects models. Accordingly, they proved that the maximum likelihood estimator is consistent for some frequently used models such as mixed effects logistic regression models and growth curve models.

However, in the considered models in this dissertation the achievement of the maximum likelihood estimate of $\boldsymbol{\beta}$'s gets more complexity as the number of random effects rises up. Since the dimension of the integrals increases and more numerical computations are required. As a result, we encourage the other method of estimation, as the generalized least square estimation, which is stated in the next subsection.

Generalized least squares estimation

The construction of the generalized least square estimation goes back to Liang and Zeger 1986, who used a working generalized linear model for the marginal distribution of Y_{ijk}. They do not specify a form of the joint distribution of the repeated measurements. Instead, they introduced estimating equations that gave consistent estimates of the parameters and of their variances under weak assumptions about the joint distribution. They modeled the marginal rather than the conditional distribution given previous observations although the conditional approach may be more appropriate for some problems. In fact to find the generalized least square estimates of model parameters, it is needed to form the first and second order moments of the response variable.

For the complete illustration of finding the generalized least square estimate of parameters, we start by the exponential family of distribution with independent observations within each subject. This means that the random effect related to each subject or cluster is zero; then, the likelihood function does not rely on the unclosed integral over the random effect and it belongs to the exponential family of distributions. Indeed, we consider fixed effects model. More precisely, the elements of $\boldsymbol{Y}_{ij} = (Y_{ij1}, ..., Y_{ijn_{ij}})^\top$ are unconditionally independent. The general exponential family of distribution can be indicated in the form:

$$\text{f}(y_{ijk}) = \exp\{[y_{ijk}b_{ij} - a(b_{ij}) + b(y_{ijk})]\phi\}, \tag{2.38}$$

where $b_{ij} = r(\eta_{ij})$; $\eta_{ij} = \boldsymbol{f}^\top(\boldsymbol{x}_{ij})\boldsymbol{\beta}$ or it can be a Nonlinear function of $\boldsymbol{\beta}$ in Nonlinear models. $r(.)$, $a(.)$ and $b(.)$ are suitably chosen functions and ϕ is an additional parameter. By this formulation, the first two moments of Y_{ijk} are given by:

$$\mu_{ij}(\boldsymbol{\beta}) = \text{E}(Y_{ijk}) = a'(b_{ij}), \tag{2.39}$$

$$\text{Var}(Y_{ijk}) = a''(b_{ij})/\phi. \tag{2.40}$$

For simplicity, we set $\mu_{ij}(\boldsymbol{\beta}) = \mu_{ij}$. Consider $\boldsymbol{\mu}_{ij} = \text{E}(\boldsymbol{Y}_{ij}); j = 1, ..., J$, then, $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i1}^\top, ..., \boldsymbol{\mu}_{iJ}^\top)^\top$ is the mean vector of the response variable $\boldsymbol{Y}_i$ with $n_i \times 1$ dimension. The independence estimating equation in order to achieve $\hat{\boldsymbol{\beta}}_I$ as the least square estimate of $\boldsymbol{\beta}$ is constructed from the root of the underlying equation:

$$U_I(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{F}_i^\top \boldsymbol{\Delta}_i(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})) = 0, \tag{2.41}$$

where $\mathbf{F}_i = (\boldsymbol{f}(\boldsymbol{x}_{i1})\mathbf{1}_{n_{i1}}^\top, ..., \boldsymbol{f}(\boldsymbol{x}_{iJ})\mathbf{1}_{n_{iJ}}^\top)^\top$; $\boldsymbol{f}(\boldsymbol{x}) = (1, \boldsymbol{f}_1(\boldsymbol{x}), ..., \boldsymbol{f}_{p-1}(\boldsymbol{x}))^\top$ for the general experimental setting $\boldsymbol{x}$. $\boldsymbol{\Delta}_i = \text{diag}(r'(\eta_{ij}))$ is an $n_i \times n_i$ matrix and $\mathbf{1}_{n_{ij}}$ is the column vector of 1's with n_{ij} dimension. We further assume the identifiability, in the sense that distinct $\boldsymbol{\beta}$'s imply distinct $\boldsymbol{\mu}$'s. The rank of the model matrix $\mathbf{F} = (\mathbf{F}_1^\top, ..., \mathbf{F}_N^\top)^\top$ depends on the choice of $\boldsymbol{f}(\boldsymbol{x})$.

Theorem 2.2.1. *Liang and Zeger 1986. Let* $\boldsymbol{A}_i = \text{diag}(a''(b_{ij})\mathbf{I}_{n_{ij}})_{j=1,...,J}$*, with* $n_i \times n_i$ *dimension. The estimator* $\hat{\boldsymbol{\beta}}_I$ *of* $\boldsymbol{\beta}$ *is consistent and* $N^{1/2}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta})$ *is asymptotically multivariate Gaussian as* $N \to \infty$ *with zero mean and covariance matrix* $\mathbf{V}_I$ *given by*

$$\mathbf{V}_I = \lim_{N\to\infty} N(\sum_{i=1}^{N} \mathbf{F}_i^\top \boldsymbol{\Delta}_i \boldsymbol{A}_i \boldsymbol{\Delta}_i \mathbf{F}_i)^{-1}(\sum_{i=1}^{N} \mathbf{F}_i^\top \boldsymbol{\Delta}_i Cov(\boldsymbol{Y}_i)\boldsymbol{\Delta}_i \mathbf{F}_i)$$
$$(\sum_{i=1}^{N} \mathbf{F}_i^\top \boldsymbol{\Delta}_i \boldsymbol{A}_i \boldsymbol{\Delta}_i \mathbf{F}_i)^{-1} \tag{2.42}$$

where the moment calculations for the $\mathbf{Y}_i's$ *are taken with respect to the underlying model.*

The precision of $\hat{\boldsymbol{\beta}}_I$ is evaluated by $\text{Var}(\hat{\boldsymbol{\beta}}_I)$, which is consistent given only a correct specification of the regression. Moreover, $\hat{\boldsymbol{\beta}}_I$ can be shown to be reasonably efficient for a few simple designs, (Liang and Zeger 1986, section 5).

The principal disadvantage of $\hat{\boldsymbol{\beta}}_I$ is that it may not be highly efficient in situations where the correlation between repeated measurements is large and it is actually neglected. However, generalized estimating equation leads to estimators with higher efficiency, since they take the correlation of observations within each subject into account. The resulting estimator from the considered equation remains consistent (Liang and Zeger 1986). In addition, consistent variance estimates are available under the weak assumption that a weighted average of the estimated correlation matrices converges to a fixed matrix.

Let $\boldsymbol{R}(\boldsymbol{\alpha})$ be an $n_i \times n_i$ symmetric matrix which fulfills the requirement of the construction of a correlation matrix, and let $\boldsymbol{\alpha}$ be an $S \times 1$ vector which fully characterizes $\boldsymbol{R}(\boldsymbol{\alpha})$. Suppose matrix $\mathbf{V}_i$ can be decomposed as:

$$\mathbf{V}_i = \boldsymbol{A}_i^{1/2} R(\boldsymbol{\alpha}) \boldsymbol{A}_i^{1/2} / \phi, \tag{2.43}$$

with $n_i \times n_i$ dimension, which will be equal to $\text{Cov}(\boldsymbol{Y}_i)$ as the variance matrix for the ith subject, if $\boldsymbol{R}(\boldsymbol{\alpha})$ is indeed the true correlation matrix for the $\boldsymbol{Y}_i' s$.

We define the quasi-score function as:

$$\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \boldsymbol{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) \tag{2.44}$$

where

$$\boldsymbol{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{D}_i^\top \mathbf{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})), \tag{2.45}$$

in which $\boldsymbol{\alpha}$ is the parameter of correlation and

$$\mathbf{D}_i = (\mathbf{D}_{i1} \mathbf{1}_{n_{i1}}^\top, ..., \mathbf{D}_{iJ} \mathbf{1}_{n_{ij}}^\top)^\top, \tag{2.46}$$

that $\mathbf{D}_{ij} = (\frac{\partial \mu_{ij}}{\partial \beta_0}, ..., \frac{\partial \mu_{ij}}{\partial \boldsymbol{\beta}_{p-1}})^\top$; $j = 1, ..., J$ with dimension $p \times 1$. Further, $(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)$ is of order $n_i \times 1$ for the i^{th} subject. This equation does not depend on ϕ, hence, $\hat{\boldsymbol{\beta}}$ is the same whether ϕ is known or not.

The general estimating equation is presented as follows:

$$\boldsymbol{U}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{0}. \tag{2.47}$$

Two remarks are worth mentioning:

(1) If we specify $\boldsymbol{R}(\boldsymbol{\alpha})$ as the identity matrix, equation (2.44) reduces to independent equation (2.41).

(2) For each i, $\boldsymbol{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is analogous to the score function derived from the quasi-likelihood approach advocated by Wedderburn 1974, (see section 2.2) except that in this case $\mathbf{V}_i$ is not only a function of $\boldsymbol{\beta}$ but also of $\boldsymbol{\alpha}$ as well. We can replace $\boldsymbol{\alpha}$ with $\hat{\boldsymbol{\alpha}}(\boldsymbol{y}, \boldsymbol{\beta}, \phi)$, a $N^{1/2}$-consistent estimator of $\boldsymbol{\alpha}$ when $\boldsymbol{\beta}$ and ϕ are known, that $\hat{\boldsymbol{\alpha}}$ for which $N^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = O_p(1)$. Except for particular choices of $\boldsymbol{R}$ and $\hat{\boldsymbol{\alpha}}$, the scale parameter ϕ will generally remain in (2.44). To complete the process, we replace ϕ by $\hat{\phi}(\boldsymbol{y}, \boldsymbol{\beta})$, a $N^{1/2}$-consistent estimator when $\boldsymbol{\beta}$ is known. As a result, (2.44) has the form,

$$\sum_{i=1}^{N} \boldsymbol{U}_i[\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))] = \mathbf{0}, \tag{2.48}$$

and $\hat{\boldsymbol{\beta}}_G$ is defined to be the solution to equation (2.48). The large-sample property for $\hat{\boldsymbol{\beta}}_G$ is stated in the next theorem.

Theorem 2.2.2. *Liang and Zeger 1986 Under mild regularity conditions and given that:*
(i): $\hat{\boldsymbol{\alpha}}$ *is* $N^{1/2}$*-consistent given* $\boldsymbol{\beta}$ *and* ϕ
(ii): $\hat{\phi}$ *is* $N^{1/2}$*-consistent given* $\boldsymbol{\beta}$*; and*
(iii): $\mid \frac{\partial \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)}{\partial \phi} \mid \leq H(\boldsymbol{y}, \boldsymbol{\beta})$ *which is* $O_p(1)$*, for an optional function* $H(.,.)$*;*
then, $N^{1/2}(\hat{\boldsymbol{\beta}}_G - \boldsymbol{\beta})$ *is asymptotically multivariate Gaussian with zero mean and covariance matrix* $\mathbf{V}_G$ *given by*

$$\mathbf{V}_G = \lim_{N \to \infty} N(\sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1} \mathbf{D}_i)^{-1} \{\sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1} \mathrm{Cov}(\boldsymbol{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i\} (\sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1} \mathbf{D}_i)^{-1} \tag{2.49}$$

The variance estimate $\hat{\mathbf{V}}_G$ of $\hat{\boldsymbol{\beta}}_G$ can be obtained by replacing $\mathrm{Cov}(\boldsymbol{Y}_i)$ by $\boldsymbol{S}_i \boldsymbol{S}_i^\top$ as the sample variance and $\boldsymbol{\beta}$, ϕ, $\boldsymbol{\alpha}$ by their estimates in the expression $\mathbf{V}_G$.

Connection with Gauss-Newton method

To compute $\hat{\boldsymbol{\beta}}_G$, Liang and Zeger (1986) suggested a modified Fisher-scoring for $\boldsymbol{\beta}$ and moment estimation for $\boldsymbol{\alpha}$ and ϕ.

Given current estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\phi}$ of the nuisance parameters, the modified iterative procedure for $\boldsymbol{\beta}$ is formed as:

$$\hat{\boldsymbol{\beta}}(m+1) = \hat{\boldsymbol{\beta}}(m) - \{\sum_{i=1}^{N} \mathbf{D}_i^\top \tilde{\mathbf{V}}_i^{-1} \mathbf{D}_i\}_{(m)}^{-1} . \{\sum_{i=1}^{N} \mathbf{D}_i^\top \tilde{\mathbf{V}}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i)\}_{(m)} \tag{2.50}$$

where $\hat{\boldsymbol{\beta}}(m)$ is the given value at the mth iteration. $\{.\}_m$ indicates the term occurring at the mth iteration. $\tilde{\mathbf{V}}_i = \mathbf{V}_i[\boldsymbol{\beta}, \hat{\alpha}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta}))]$. This procedure is viewed as a modification of Fisher's scoring in that the limiting expectation value of the derivative of $\sum \boldsymbol{U}_i(\boldsymbol{\beta}, \hat{\alpha}(\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})))$ is used for correction.

Estimators of $\boldsymbol{\alpha}$ and ϕ

At a given iteration the correlation parameters $\boldsymbol{\alpha}$ can be estimated from the current Pearson residuals defined by

$$\hat{r}_{ijk} = \frac{y_{ijk} - a'(\hat{b}_{ij})}{(a''(\hat{b}_{ij}))^{1/2}}, \tag{2.51}$$

where $\hat{b}_{ij}$ depends upon the current value for $\boldsymbol{\beta}$. We can estimate the scale parameter ϕ by

$$\hat{\phi}^{-1} = \sum_{i=1}^{N} \sum_{j=1}^{J} \frac{\hat{r}_{ijk}^2}{n-p} \tag{2.52}$$

where $n = \sum_{i=1}^{N} n_i$. It is shown to be $N^{1/2}$-consistent given that the fourth moments of the y_{ij}'s are finite. The estimate of $\boldsymbol{\alpha}$ depends on the choice of $\boldsymbol{R}(\boldsymbol{\alpha})$ by a simple function of

$$\hat{R}_{uvk} = \sum_{i=1}^{N} \frac{\hat{r}_{iuk}\hat{r}_{ivk}}{n-p}, \tag{2.53}$$

Liang and Zeger 1986.

All in all, the variance matrix $\mathbf{V}_i$ is estimated as a function of $\boldsymbol{\beta}$.

In this dissertation, the nuisance parameter ϕ and the correlation parameter $\boldsymbol{\alpha}$ are assumed to be known. The variance matrix is calculated based on the first and second order moments of the response variable. Actually, by setting nuisance parameter $\sigma^2 = 1$ and $\boldsymbol{\beta}$ known in models (2.1), (2.2) and (2.33) for the computation of D-optimum design, the considered aim is fulfilled.

Quasi maximum-likelihood estimation

The quasi-likelihood function was firstly introduced by Wedderburn 1974. To define the likelihood function, it is needed to specify the form of the distribution of observations, but to define a quasi-likelihood function we need only specify a relation between the mean and variance of the observations. This function will then be applied for estimation. Wedderburn 1974 proposed that the quasi log-likelihood function for a one-parameter exponential family is the same as the log-likelihood function.

Assume there is one observation for each individual, and since individuals are assumed to be independent, all observations are independent. The quasi-likelihood function for each observation is defined by the relation in the following definition:

Definition 2.2.2. *For each observation y_i the quasi-likelihood function $ql(y_i, \mu_i)$ is defined by the relation*

$$\frac{\partial ql(y_i, \mu_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{\mathrm{V}_i}, \tag{2.54}$$

or equivalently,

$$ql(y_i, \mu_i) = \int^{\mu_i} \frac{y_i - t}{\mathrm{V}(t)} \mathrm{d}t + H(y_i), \tag{2.55}$$

where $H(y_i)$ is a function of y_i.

According to the above definition of $ql(.,.)$, has the following properties (Wedderburn 1974):

1) $\mathrm{E}[\frac{\partial ql(Y_i,\mu_i)}{\partial \mu_i}] = 0,$

2) $\mathrm{E}[\frac{\partial ql(Y_i,\mu_i)}{\partial \boldsymbol{\beta}_j}] = 0; j = 1, ..., J$

3) $\mathrm{E}[\frac{\partial ql(Y_i,\mu_i)}{\partial \mu_i}]^2 = -\mathrm{E}[\frac{\partial^2 ql(Y_i,\mu_i)}{\partial \mu_i^2}] = \frac{1}{\mathrm{V_i}}$

4) $\mathrm{E}[\frac{\partial ql(Y_i,\mu_i)}{\partial \beta_j} \frac{\partial ql(Y_i,\mu_i)}{\partial \beta_{j'}}] = -\mathrm{E}[\frac{\partial^2 ql(Y_i,\mu_i)}{\partial \beta_j \partial \beta_{j'}}] = \frac{1}{\mathrm{V_i}} \frac{\partial \mu_i}{\partial \beta_j} \frac{\partial \mu_i}{\partial \beta_{j'}}; j, j' = 1, ..., J$

Corollary 2.2.2.1. *[Wedderburn 1974] If the log-likelihood of observations y_i is defined as ℓ_i, then*

$$-\mathrm{E}(\frac{\partial^2 ql_i}{\partial \mu_i^2}) \leq -\mathrm{E}(\frac{\partial^2 \ell_i}{\partial \mu_i^2}). \tag{2.56}$$

Now, let

$$\boldsymbol{U}_i(\boldsymbol{\beta}) = \frac{\partial \mu_i}{\partial \boldsymbol{\beta}^\top} \frac{\partial ql(y_i, \mu_i)}{\partial \mu_i}, \tag{2.57}$$

and

$$\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{U}_i(\boldsymbol{\beta}). \tag{2.58}$$

Thus, the generalized estimating equation for the considered model is stated as:

$$\boldsymbol{U}(\hat{\boldsymbol{\beta}}) = 0. \tag{2.59}$$

Then the quasi maximum-likelihood estimate is obtained from the root of above equation (Wedderburn 1974). This definition was also stated in (2.45) and (2.46) with an additional parameter $\boldsymbol{\alpha}$ to state the generalized estimating equation, that here we assume it is hidden in the construction of $\mathbf{V}_i$. Similar to section 2.2, the quasi maximum likelihood estimate $\hat{\boldsymbol{\beta}}_{QL}$ can be obtained from the iterative equation (2.50), in which $\tilde{\mathbf{V}}_i$ is obtained from equations (3.1)-(3.7), Chapter 3.

In case $\mathbf{V}_i = 1$, the quasi-score function in (2.58) reduces to least square equation, and therefore, the quasi maximum-likelihood estimation is transformed to the least square estimation (see formula (2.41). It is well-known that if the likelihood function has the exponential form, maximum likelihood estimates of regression parameters will often be used to find the least square estimation.

In fact the method of weighted least squares can be applied to find maximum likelihood estimates even in some situations where the likelihood function does not have the exponential family form (Jørgensen 1983). McCullagh 1983 suggested a wider class of problems for which the least square approach can be used to maximize the likelihood function of the data, and he examined the asymptotic properties of the quasi-likelihood function. He also showed that the estimators enjoy a certain asymptotic optimality property.

In the following subsection, the properties of the quasi maximum-likelihood estimation are stated.

Properties of quasi maximum-likelihood estimation

Consider the general exponential family of distributions (2.38). The systematic component may be expressed as a regression function as follows:

$$\mathrm{E}(\boldsymbol{Y}_{ij}) = \boldsymbol{\mu}_{ij}(\boldsymbol{\beta}), \tag{2.60}$$

which is called the linear regression.

A more general regression function that implicitly involves dispersion parameter ϕ is denoted by McCullagh 1983:

$$\mathrm{E}[\mho(\boldsymbol{Y}_{ij})] = \psi_i(\boldsymbol{\beta}), \tag{2.61}$$

where $\mho(.)$ is a known Nonlinear function of the data.

If ϕ is known, the expression on the left side of (2.61) is a function of μ alone, and subject to reasonable monotonicity, we can express (2.61) in terms of (2.60). Weighted least squares may be applied to compute the maximum likelihood estimates in (2.60) and (2.61), (Bradley 1973).

The generalized least squares equation is shown in (2.59). The extension of the quasi log-likelihood definition to the vector $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^\top, ..., \boldsymbol{y}_{iJ}^\top)^\top$ is shown as (McCullagh 1983):

$$\frac{ql(\boldsymbol{y}_i, \boldsymbol{\mu}_i)}{\partial \boldsymbol{\mu}_i} = \mathbf{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i). \tag{2.62}$$

There is no guarantee that the quasi maximum-likelihood estimate of $\boldsymbol{\beta}$ is analogous to the maximum likelihood estimate of $\boldsymbol{\beta}$. Under weak conditions on the third derivative of (2.60) and assuming that $N^{-1}\mathbf{i}_{\boldsymbol{\beta}} = [-\frac{\partial^2 ql(\boldsymbol{y}_i,\boldsymbol{\mu}_i)}{\partial\beta_j\partial\beta_{j'}}]_{j,j'=1}^p$ has a positive definite limit, and that the third moments of $\boldsymbol{Y}$ are finite, the following asymptotic results apply:

$$N^{\frac{-1}{2}}U(\boldsymbol{\beta}) \sim N_p(\mathbf{0}, \frac{\mathbf{i}_{\boldsymbol{\beta}}}{N}) + O_p(N^{-\frac{1}{2}}), \tag{2.63}$$

$$\mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O(N^{-1}), \tag{2.64}$$

$$N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N_p(\mathbf{0}, N\mathbf{i}_{\boldsymbol{\beta}}^{-1}) + O_p(N^{-\frac{1}{2}}) \tag{2.65}$$

If the third moment is infinite, the error term in results (2.63) and (2.65) above are $O_p(1)$.

It can be shown that among all estimators of $\boldsymbol{\beta}$ for which the influence function is linear, i.e., estimator $\tilde{\boldsymbol{\beta}}$ satisfying

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} = L_{\boldsymbol{\mu}}(\boldsymbol{y} - \boldsymbol{\mu}) + O_p(N^{-1/2}), \tag{2.66}$$

where $L_{\boldsymbol{\mu}}$ is a $p \times n$ matrix of influences, quasi maximum-likelihood estimates have minimum asymptotic variance (McCullagh 1983).

Now, consider the generalized linear mixed model (2.1). The random effect $\boldsymbol{\zeta}_i \neq \mathbf{0}$. As it was mentioned, the observations within each cluster (subject) are correlated and they are independent between clusters. Moreover, an alternative approach to derive the considered moments, was a BLUP (best linear unbiased predictor approach). In this method, $\boldsymbol{\zeta}_i$ is assumed to be fixed like $\boldsymbol{\beta}$, and we try to estimate it along with $\boldsymbol{\beta}$, by using certain modified likelihood approaches, such as penalized quasi-likelihood approach of Breslow and Clayton 1993. Sutradhar 2004 considered the generalized linear mixed model with random intercept. He used the simulation approach of Jiang 1998, to compute the marginal first and second order moment of the response variable. This approach is different from the one which is introduced in this dissertation. Moreover, he assumed that each random effect $\boldsymbol{\zeta}_i; i = 1, ..., N$, has distinctive variance σ_i, and he was interested in the estimation of $\varphi = (\boldsymbol{\beta}, \sigma_1, ..., \sigma_q)$, with $(p+q)\times 1$ dimension. In this dissertation, $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, ..., \boldsymbol{Y}_N^\top)^\top$ is attended as the estimator for population parameter $\varphi = (\boldsymbol{\beta}, \sigma^2)^\top$ with $p \times 1$ dimension, and we utilized the considered vector estimator in generalized estimating equation (2.44). Due to the work of Jiang and Zhang 2001, Sutradhar 2004 utilized the base statistic based on the sum of the first and second order of y_{ijk}'s with coefficients as the design predictors. According to this assumption he formed the variance of the base statistic. He confirmed that it is not diagonal, since the two elemental statistics are correlated.

After the construction of the generalized estimating equation (2.47), he used the Newton-Raphson method to obtain the quasi maximum-likelihood estimate of φ and stated some general properties of the considered estimator.

Following the relevant results, the properties of $\hat{\boldsymbol{\beta}}_{QL}$ for model (2.1) is stated as:

1) $N^{1/2}(\hat{\boldsymbol{\beta}}_{QL} - \boldsymbol{\beta})$ is asymptotically (as $N \to \infty$) multivariate Normal, with zero mean and the covariance matrix that can be consistently estimated by:

$$\hat{\text{Var}}(\hat{\boldsymbol{\beta}}_{QL}) = \lim_{N\to\infty} N[\sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1}\mathbf{D}_i]_{\hat{\boldsymbol{\beta}}}^{-1}, \tag{2.67}$$

where $\mathbf{D}_i$ is the derivative of the mean of the base statistic with respect to φ and $[.]_{\hat{\boldsymbol{\beta}}_{QL}}$ denotes that the expression within the bracket is evaluated at $\hat{\boldsymbol{\beta}}_{QL}$.

2) Note that as the variance matrix $\mathbf{V}_i$ in QL estimating equation (2.47) is the true covariance matrix of the base statistic $\boldsymbol{Y}_i$, $\hat{\boldsymbol{\beta}}_{QL}$ is both consistent as well as highly efficient.

3) In an example of the binary logistic regression model, it is shown that the quasi maximum-likelihood estimate of φ has lower variance, (higher efficiency) than the variance of the improved method of moments (Jiang and Zhang 2001).

In Section 3.2, the quasi maximum likelihood estimate of the parameters in the specific case of the binary mixed effects model is evaluated by simulation.

2.3 Quasi Fisher information matrix

Since the Fisher information matrix in the three types of models (2.1), (2.2) and (2.33), does not have an analytical explicit form, we approximate it with the quasi Fisher information matrix.

The quasi Fisher information matrix is obtained from the definition of the quasi score function (2.44) as (P. McCullagh and Nelder 1983):

$$\text{Cov}(\boldsymbol{U}(\boldsymbol{\beta})) = \sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1}\mathbf{D}_i, \tag{2.68}$$

We set

$$\mathbf{M}_i^Q(\boldsymbol{\beta}) = \mathbf{D}_i^\top \mathbf{V}_i^{-1}\mathbf{D}_i, \tag{2.69}$$

as the individual quasi Fisher information matrix. Also,

$$\mathbf{M}^Q(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{M}_i^Q(\boldsymbol{\beta}), \tag{2.70}$$

where $\mathbf{M}^Q(\boldsymbol{\beta})$ is the population quasi Fisher information matrix.

Hence, the population quasi Fisher information with $\mathbf{D}^\top = (\mathbf{D}_1^\top, ..., \mathbf{D}_N^\top)$, and $\mathbf{V} = \text{diag}(\mathbf{V}_1, ..., \mathbf{V}_N)$ can be shown as:

$$\mathbf{M}^Q(\boldsymbol{\beta}) = \mathbf{D}^\top \mathbf{V}^{-1} \mathbf{D}. \tag{2.71}$$

Therefore, following equation (2.67) we obtain:

$$\text{Var}(\hat{\boldsymbol{\beta}}) \approx (\mathbf{M}^Q(\boldsymbol{\beta}))^{-1}. \tag{2.72}$$

Now, we present $\mathcal{M}$ as the set of quasi Fisher information matrices.

The set $\mathcal{M}$. Each element of $\mathcal{M}$ is a symmetric non-negative definite $p \times p$ matrix which can be represented by a point in $\mathbb{R}^{\frac{1}{2}p(p+1)}$, the points with coordinates $\{\mathbf{M}^Q_{ij}; 1 \leq i \leq j \leq p\}$, when $\mathbf{M}^Q = (\mathbf{M}^Q_{ij})$.

2.4 Design

In general, the discrete individual design $\xi_i^{(d)}$ is indicated as (Atkinson et al. 2007):

$$\xi_i^{(d)} = \begin{pmatrix} \boldsymbol{x}_{i1} & \ldots & \boldsymbol{x}_{iJ} \\ n_{i1} & \ldots & n_{iJ} \end{pmatrix}; \xi_i^{(d)} \in \Xi_i^{(d)}, \tag{2.73}$$

where $\boldsymbol{x}_{ij}; j = 1, ..., J$ is the jth experimental setting of the ith individual. Furthermore, $\Xi_i^{(d)}$ is the convex set of all possible designs for the ith individual. In model (2.33), the support of the design is denoted by t_{ij} as the jth time point, in the ith individual.

Moreover, the discrete population design is defined in the following:

$$\xi^{(d)} = \begin{pmatrix} \xi_1^{(d)} & \ldots & \xi_N^{(d)} \\ n_1 & \ldots & n_N \end{pmatrix}; \xi^{(d)} \in \Xi^{(d)}, \tag{2.74}$$

where each n_i represents the number of replication taken at the ith individual design and $\sum_{i=1}^N n_i = n$. Furthermore, $\Xi^{(d)}$ is the convex set of all population designs.

The quasi Fisher information matrix is not only a function of population parameter $\boldsymbol{\beta}$, but it is also a function of the discrete design $\xi^{(d)}$. In the next subsection the D-optimality criterion is presented.

D-optimum design

In general, the D-optimality criterion based on the population quasi Fisher information matrix is defined as follows (Atkinson et al. 2007):

$$\mathfrak{D}^Q(\xi^{(d)}, \boldsymbol{\beta}) = \log(\det(\mathbf{M}^Q(\xi^{(d)}, \boldsymbol{\beta}))). \tag{2.75}$$

Also, the individual D-optimality criterion is revealed based on the individual quasi Fisher information matrix, as

$$\mathfrak{D}^Q(\xi_i^{(d)}, \boldsymbol{\beta}) = \log(\det(\mathbf{M}^Q(\xi_i^{(d)}, \boldsymbol{\beta}))), \tag{2.76}$$

The D-optimum design $\xi^{\star(d)}$ is obtained by the maximization of D-optimality criterion (2.75) with respect to design, in other words,

$$\xi^{\star(d)} = \arg\max_{\xi^{(d)}}\{\mathfrak{D}^Q(\xi^{(d)}, \boldsymbol{\beta})\} \tag{2.77}$$

Equal individual designs

In this dissertation, we assume all individuals take the same design, i.e. $\xi_1^{(d)} = ... = \xi_N^{(d)} = \xi_0^{(d)}$ (see Appendix A, Section A.1). Therefore, the support points, $\boldsymbol{x}_{ij} = \boldsymbol{x}_j$, and $n_{ij} = n_j$. Also, $n_i = n; \forall i$.

$$\mathbf{M}^Q(\xi^{(d)}, \boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{M}^Q(\xi_i^{(d)}, \boldsymbol{\beta}) = N\mathbf{M}^Q(\xi_0^{(d)}, \boldsymbol{\beta}) \tag{2.78}$$

and also the population D-optimality criterion can be denoted as a function of the individual D-optimality criterion as follows:

$$\begin{aligned}\log(\det(\mathbf{M}^Q(\xi^{(d)}, \boldsymbol{\beta}))) &= \log(\det(N\mathbf{M}^Q(\xi_0^{(d)}, \boldsymbol{\beta}))) \\ &= p\log(N) + \log(\det(\mathbf{M}^Q(\xi_0^{(d)}, \boldsymbol{\beta})));\end{aligned} \tag{2.79}$$

hence, the maximization of the individual D-optimality criterion with respect to $\xi_0^{(d)}$ is equal to the maximization of the population D-optimality criterion with respect to $\xi^{(d)}$.

D-optimum design $\xi^{(d)}$ for models (2.1) and (2.2) determines that for each subject or cluster, how many replication we take at each experimental setting $\boldsymbol{x}_j; j = 1, ..., J$, in order that the determinant of the quasi Fisher information matrix $\mathbf{M}^Q(\xi^{(d)}, \boldsymbol{\beta})$ is maximized. In other words, $\det(\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{QL}))$ is minimized.

In model (2.33) the D-optimum design states that at each subject how many replications of the experiment is performed at each time, $t_j; j = 1, ..., J$ in order that the determinant of the quasi Fisher information matrix, $\mathbf{M}^Q(\xi^{(d)}, \boldsymbol{\beta})$ is maximized, in other words, $\det(\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{QL}))$ is minimized.

In this dissertation, we assume that the support points of the design are known; thus, to obtain the D-optimum design we only need to obtain the optimum n_j's. However, since n_j's are discrete, the achievement of the D-optimum design will be complex. Therefore, we need to represent the approximate design as (Silvey 1980):

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_J \\ w_1 & \dots & w_J \end{pmatrix}; \xi_i \in \Xi_i, \tag{2.80}$$

where $w_j = \frac{n_j}{n}$; $\sum_{j=1}^{J} w_j = 1$, $w_j \in (0,1)$. Ξ_i is the convex set of all designs ξ_i. However, still care has to be taken about the total number "n" of observations per individual.

Now finding D-optimum design $\xi^\star$ is equal to the achievement of $w_j^\star; j = 1, ..., J$, where $\mathbf{M}^Q(\xi, \boldsymbol{\beta})$ is maximized.

Equivalence theorem

The optimality of D-optimum design $\xi^\star$ can be checked by the equivalence theorem. Since D-optimality criterion (2.75) is a function of the approximate design ξ, the directional derivative $F_{\mathfrak{D}^Q}(\xi_1, \xi_2)$ is used to check the optimality. It is defined as follows:

$$F_{\mathfrak{D}^Q}(\xi_1, \xi_2) = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon}[\mathfrak{D}^Q(\mathbf{M}^Q((1-\epsilon)\xi_1 + \epsilon\xi_2)) - \mathfrak{D}^Q(\mathbf{M}^Q(\xi_1))]. \tag{2.81}$$

This function gives the rate of increase of $\mathfrak{D}^Q$ per unit of distance moved in the direction given by ξ_2.

Moreover, it can be written as,

$$F_{\mathfrak{D}^Q}(\xi_1, \xi_2) = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathfrak{D}^Q(\mathbf{M}^Q((1-\epsilon)\xi_1 + \epsilon\xi_2)) \mid_{\epsilon=0^+} . \tag{2.82}$$

For design measures ξ_1 and ξ_2, and $\epsilon \in (0,1)$, $\epsilon\xi_1 + (1-\epsilon)\xi_2$ is also a design measure. Therefore, the set of the design measures is a convex set. Based on the form of the quasi Fisher information matrix, we could conclude

$$\mathbf{M}^Q(\epsilon\xi_1 + (1-\epsilon)\xi_2) \geq \epsilon\mathbf{M}^Q(\xi_1) + (1-\epsilon)\mathbf{M}^Q(\xi_2). \tag{2.83}$$

Hence the quasi Fisher information matrix is included in a convex set of quasi Fisher information matrices.

Definition 2.4.1. *(Silvey 1980) Differentiablity of $\mathfrak{D}^Q$ at $\mathbf{M}^Q$ implies that, if $\sum_{i=1}^{N} a_i = 1$,*

$$F_{\mathfrak{D}^Q}(\xi_1, \sum_{i=1}^{N} a_i\xi_i) = \sum_{i=1}^{N} a_i F_{\mathfrak{D}^Q}(\xi_1, \xi_i). \tag{2.84}$$

If $\mathbf{M}^Q$ is convex then, $\mathfrak{D}^Q$ is differentiable. The differentiability is the assumption of the following theorem.

Theorem 2.4.1. ***Equivalence Theorem.*** *(Silvey 1980). If $\mathfrak{D}^Q$ is differentiable at all points of $\mathcal{M}^{Q^+} = \{\mathbf{M}^Q \in \mathcal{M}^Q, \mathfrak{D}^Q < \infty\}$, ξ_x is the one-point design which assigns weight 1 to x and a $\mathfrak{D}^Q-$ optimum measure exists, then $\xi^\star$ is $\mathfrak{D}^Q-$ optimum if and only if*

$$\max_{x\in\mathfrak{X}} F_{\mathfrak{D}^Q}(\xi^\star, \xi_{\boldsymbol{x}}) = \min_{\xi} \max_{x\in\mathfrak{X}} F_{\mathfrak{D}^Q}(\xi, \xi_{\boldsymbol{x}}), \tag{2.85}$$

where the minimum with respect to ξ is the minimum over $\{\xi : \mathrm{M}^Q(\xi) \in \mathcal{M}^{Q+}\}$.

D-efficiency criterion

In general, the efficiency of the D-optimum design is evaluated by the D-efficiency criterion as:

$$\mathrm{eff}_D(\xi;\boldsymbol{\beta}) = \left(\frac{\det(\mathbf{M}^Q(\xi;\boldsymbol{\beta}))}{\det(\mathbf{M}^Q(\xi^\star_{\boldsymbol{\beta}};\boldsymbol{\beta}))}\right)^{\frac{1}{p}}, \tag{2.86}$$

where $\xi^*_{\boldsymbol{\beta}}$ is the D-optimum design assuming the specific value of $\boldsymbol{\beta}$. Also, ξ is an arbitrary design. For example, if the D-efficiency is 0.5 for design ξ, we would take as twice as the sample size when the true D-optimum design ($\xi^\star_{\boldsymbol{\beta}}$) would be used.

In this chapter, some specific definitions and theories regarding the model equations, their properties and estimation have been discussed. Moreover, the reason for considering the quasi Fisher information matrix for the further analysis has been explained. The definition of D-optimum design and its brief theory have been provided. In addition, some references for further studies, new ideas and investigations have been given.

Chapter 3

Optimum Design in a Binary Mixed Effects Model

In this chapter, the D-optimum design is aimed to be obtained for the binary mixed effects model. For this aim, the Fisher information matrix of the model needs to be constructed. As the Fisher information matrix in the model of consideration does not have the analytical closed form (as indicated in Section 2.1), it is approximated by the quasi-Fisher information matrix (Wedderburn 1974). To derive this matrix, it is required to obtain the marginal first and second order moments of the response variable. These moments do not have the explicit form either, in some situations of the model. Thus, they are approximated, and they are evaluated by numerical computations. According to this, the D-optimum design is achieved for a treatment-control model for the two settings for the design. Since the individuals take the same design, the D-optimum design is obtained for individual and it is extended to the population D-optimum design. These designs are constructed for the quasi-maximum likelihood estimate of the model parameters. Therefore, the asymptotic behaviour of these estimates is evaluated by simulation. Because the quasi-Fisher information matrix does not have an explicit form, there cannot be made direct use of an equivalence theorem. Then, D-optimality of the considered designs are evaluated directly by the second derivative of the D-optimality criterion with respect to the design. At last, the D-optimum design for the random intercept binary model is obtained for different values of model parameters and the sensitivity of D-optimum design with respect to parameter changes is investigated.

3.1 Binary mixed effects model

The binary mixed effects model have been represented in Section 2.1. The response variable Y_{ijk} takes values 0 and 1. The success probability $p_{ij}(\boldsymbol{\zeta_i})$ for $Y_{ijk} = 1$ depends on the setting x_j and on the random effect $\boldsymbol{\zeta_i}$. It is assumed $\boldsymbol{\zeta_i} \sim N(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_q$, which means that the elements of the random effect $\boldsymbol{\zeta}_i$ are assumed to be independent and have the same distribution. The marginal density function of Y_{ijk} does not have the closed form. Therefore, marginal likelihood function of the considered model, shown in (2.10), does not have analytical closed form. According to this, the Fisher information matrix cannot be attained analytically.

As it was mentioned, the quasi Fisher information matrix is one suggestion of the approximation of Fisher information matrix, and therefore, it is needed to obtain the first and second order moments of the response variable; they are indicated as follows:

Let $\pi_j = \mathrm{E}(p_{ij}(\boldsymbol{\zeta}_i))$; where $p_{ij}(\boldsymbol{\zeta}_i) = P(Y_{ijk} = 1|\boldsymbol{\zeta}_i) = \mathrm{E}(Y_{ijk}|\boldsymbol{\zeta_i})$; then,

$$\pi_j := \mathrm{E}(Y_{ijk}). \tag{3.1}$$

Also, because Y_{ijk} is binary that

$$\mathrm{Var}(Y_{ijk}) = \pi_j(1 - \pi_j). \tag{3.2}$$

The marginal covariance elements of the response variable are written as:

$$\mathrm{cov}(Y_{ijk}, Y_{ijk'}) = \mathrm{cov}(\mathrm{E}(Y_{ijk}|\boldsymbol{\zeta}_i), \mathrm{E}(Y_{ijk'}|\boldsymbol{\zeta}_i)) + \mathrm{E}(\mathrm{cov}(Y_{ijk}, Y_{ijk'}|\boldsymbol{\zeta}_i)),$$

and the second term above is zero; therefore, by defining $v_j := \mathrm{Var}(p_{ij}(\boldsymbol{\zeta}_i))$,

$$\mathrm{cov}(Y_{ijk}, Y_{ijk'}) = v_j; k \neq k', \tag{3.3}$$

since $\mathrm{cov}(\mathrm{E}(Y_{ijk} \mid \boldsymbol{\zeta}_i), \mathrm{E}(Y_{ijk'} \mid \boldsymbol{\zeta}_i)) = \mathrm{Var}(p_{ij}(\boldsymbol{\zeta}_i))$. In addition,

$$\mathrm{cov}(Y_{ijk}, Y_{ij'k'}) = \mathrm{cov}(\mathrm{E}(Y_{ijk}|\boldsymbol{\zeta}_i), \mathrm{E}(Y_{ij'k'}|\boldsymbol{\zeta}_i)) + \mathrm{E}(\mathrm{cov}(Y_{ijk}, Y_{ij'k'}|\boldsymbol{\zeta}_i)),$$

and the second term above is zero; as a result, by defining $c_{jj'} := \mathrm{cov}(p_{ij}(\boldsymbol{\zeta}_i), p_{ij'}(\boldsymbol{\zeta}_i))$,

$$\mathrm{cov}(Y_{ijk}, Y_{ij'k'}) = c_{jj'}; j \neq j', k, k', \tag{3.4}$$

since $\mathrm{cov}(\mathrm{E}(Y_{ijk} \mid \boldsymbol{\zeta}_i), \mathrm{E}(Y_{ij'k'} \mid \boldsymbol{\zeta}_i)) = \mathrm{cov}(p_{ij}(\boldsymbol{\zeta}_i), p_{ij'}(\boldsymbol{\zeta}_i))$. The variance-covariance matrix in the same setting j is equal to:

$$\mathbf{V}_{jj} = v_j \mathbf{1}_{n_j} \mathbf{1}_{n_j}^\top + (\pi_j(1 - \pi_j) - v_j)\mathbf{I}_{n_j}, \tag{3.5}$$

with dimension $n_j \times n_j$ and the variance-covariance matrix between the jth and j'th experimental settings is given by:

$$\mathbf{V}_{jj'} = c_{jj'} \mathbf{1}_{n_j} \mathbf{1}_{n_{j'}}^\top. \tag{3.6}$$

with dimension $n_j \times n_{j'}$. In total, the individual variance-covariance matrix of $\boldsymbol{Y}_i$ is constructed as follows:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \dots & \mathbf{V}_{1J} \\ \mathbf{V}_{21} & \dots & \mathbf{V}_{2J} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{J1} & \dots & \mathbf{V}_{JJ} \end{pmatrix}, \tag{3.7}$$

with dimension $n \times n$.

In the next subsections, the variance matrix $\mathbf{V}$ is established for a specific case of the binary mixed effects model to form individual quasi Fisher information matrix (2.69).

Quasi Fisher information matrix in model with probit link function

In this subsection, we consider the probit link function $g(p) = \Phi^{-1}(p)$. In order to construct π_j in (3.1), we need to obtain the following integral:

$$\pi_j = \int_{\mathbb{R}^q} \Phi(\eta_{ij}) \mathrm{f}_{N(\mathbf{0}, \sigma^2 \mathbf{I}_q)}(\boldsymbol{\zeta}_i) \mathrm{d}\boldsymbol{\zeta}_i, \tag{3.8}$$

where $\eta_{ij} = \boldsymbol{f}^\top(\boldsymbol{x}_{ij})\boldsymbol{\beta} + \boldsymbol{h}^\top(\boldsymbol{x}_{ij})\boldsymbol{\zeta}_i$, $\mathrm{f}_{N(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$ denotes the density of the multivariate Normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. This integral was solved in Zeger, Liang, and Albert 1988, and the regarding result is shown in the following lemma.

Lemma 3.1.1. ***Zeger, Liang, and Albert 1988.*** *Set*

$$\alpha_j = (1 + \sigma^2 h^\top(\boldsymbol{x}_j) h(\boldsymbol{x}_j))^{-\frac{1}{2}} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}.$$

Then, in the binary mixed probit regression model,

$$\pi_j = \Phi(\alpha_j). \tag{3.9}$$

Furthermore, following (3.2),

$$\mathrm{Var}(Y_{ijk}) = \Phi(\alpha_j)(1 - \Phi(\alpha_j)). \tag{3.10}$$

In order to find off-diagonal elements of $\mathbf{V}_i$, we consider the special case of the binary mixed probit regression model with a random intercept.

Lemma 3.1.2. *Let* $\alpha_{j0} = (1+\sigma^2)^{-\frac{1}{2}} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}$ *and*

$$q_{j0}(z) := \frac{(1+\sigma^2)^{\frac{1}{2}}}{(1+2\sigma^2)^{\frac{1}{2}}} \left(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} - \frac{\sigma^2 z}{(1+\sigma^2)^{\frac{1}{2}}} \right). \tag{3.11}$$

In the random intercept binary regression model,

$$\text{cov}(Y_{ijk}, Y_{ijk'}) = \int_{-\infty}^{\alpha_{j0}} \phi(z)\Phi(q_{j0}(z))\text{d}z - (\Phi(\alpha_{j0}))^2; k \neq k', \tag{3.12}$$

and

$$\text{cov}(Y_{ijk}, Y_{ij'k'}) = \int_{-\infty}^{\alpha_{j0}} \phi(z)\Phi(q_{j'0}(z))\text{d}z - (\Phi(\alpha_{j0}))(\Phi(\alpha_{j'0})); j \neq j', k, k'. \tag{3.13}$$

Proof.
Consider the random intercept probit model,

$$\Phi^{-1}(p_{ij}(\zeta_{i0})) = \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} + \zeta_{i0}. \tag{3.14}$$

Firstly, it is attempted to obtain (3.13), then the result is reduced to the case, $j = j'$.

As $\text{cov}(Y_{ijk}, Y_{ij'k'}) = c_{jj'}$ in (3.4), it is needed to obtain $\text{E}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})]$; and, the first order moment of the response variable has already been obtained in Lemma 3.1.1.

$$\begin{aligned} \text{E}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] &= \textstyle\int_{-\infty}^{+\infty} \Phi(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} + \zeta_{i0})\Phi(f^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} + \zeta_{i0})\phi_{0,\sigma}(\zeta_{i0})\text{d}\zeta_{i0}, \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}} \phi_{-\zeta_{i0},1}(v)\text{d}v \right) . \\ &\qquad \left(\int_{-\infty}^{f^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta}} \phi_{-\zeta_{i0},1}(t)\text{d}t \right) \phi_{0,\sigma}(\zeta_{i0})\text{d}\zeta_{i0}, \end{aligned} \tag{3.15}$$

where $\phi_{\mu,\sigma}(.)$ indicates the density function of the Normal distribution. The last equality above is obtained by the variable change principle.

At last, by considering $q_{j'0}(z) = \frac{(\sigma^2+1)^{\frac{1}{2}}}{(2\sigma^2+1)^{\frac{1}{2}}} \left(f^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} - \frac{\sigma^2 z}{(\sigma^2+1)^{\frac{1}{2}}} \right)$,

$$\text{E}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] = \textstyle\int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}} \phi_{0,(1+\sigma^2)}(v) \int_{-\infty}^{f^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta}} \phi_{\frac{\sigma^2 v}{1+\sigma^2},(1+\frac{\sigma^2}{1+\sigma^2})}(t)\text{d}t\text{d}v$$

$$= \int_{-\infty}^{\alpha_{j0}} \phi(z)\Phi(q_{j'0}(z))\mathrm{d}z, \tag{3.16}$$

By the substitution of (3.16) for the first term in (3.4) and using lemma 3.1.1 for the replacement of π_j (3.13) is obtained.

Finally, (3.12) is gained by setting $j = j'$. ■

From the above lemma it is observed that $\mathrm{cov}(Y_{ijk}, Y_{ijk'})$ and $\mathrm{cov}(Y_{ijk}, Y_{ij'k'})$ do not have explicit closed form due to the unclosed integral in their first term. Later, in the next subsection, we utilize an additional approximation of the current unclosed integral in the first terms, and its outcomes are evaluated and compared with the results of Lemma 3.1.2.

Lemma 3.1.3. *The columns of matrix* $\mathbf{D}$ *in (2.46) for the binary mixed probit regression model is as follows:*

$$\mathbf{D}_j = (1 + \sigma^2 \boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{h}(\boldsymbol{x}_j))^{-\frac{1}{2}} \boldsymbol{f}(\boldsymbol{x}_j)\phi(\alpha_j); j = 1, ..., J. \tag{3.17}$$

Proof.

Since $\pi_j = \Phi(\alpha_j)$,

$$\frac{\partial}{\partial \boldsymbol{\beta}^\top}\pi_j = \frac{\partial}{\partial \boldsymbol{\beta}^\top}\Phi(\alpha_j), \tag{3.18}$$

and the result is obtained. ■

Remark. By considering the special case of the model as the random intercept probit regression model,

$$\mathbf{D}_j = (1 + \sigma^2)^{-\frac{1}{2}} \boldsymbol{f}^\top(\boldsymbol{x}_j)\phi(\alpha_{j0}) \tag{3.19}$$

Approximation of covariance matrix

In order to approximate the first term integral in (3.12) and (3.13) of Lemma 3.1.2, it is needed to approximate $\Phi(q_j(z)), \forall j$. One approximation of this function is stated in Lin 1989, which is used for Approximating the Normal tail probability and its inverse for use on a pocket calculator. It is denoted as:

$$\Phi^{\mathrm{Lin}}(z) = \frac{1}{2}\exp\{-\tau z - \alpha z^2\}, \tag{3.20}$$

where $\alpha = 0.416$, $\tau = 0.717$. This approximation is modified to the following term:

$$\Phi(z) \approx \Phi^*(z) = \begin{cases} 1 - \frac{1}{2}\exp(-\tau z - \alpha z^2) & ; \quad z \geq 0, \\ \frac{1}{2}\exp(-\tau(-z) - \alpha(-z)^2) & ; \quad z \leq 0, \end{cases} \tag{3.21}$$

which overlaps the true standard Normal cumulative distribution function on the interval $z \in [-3, 3]$. This approximation shares the symmetry property, i.e. $\Phi^*(z)+$

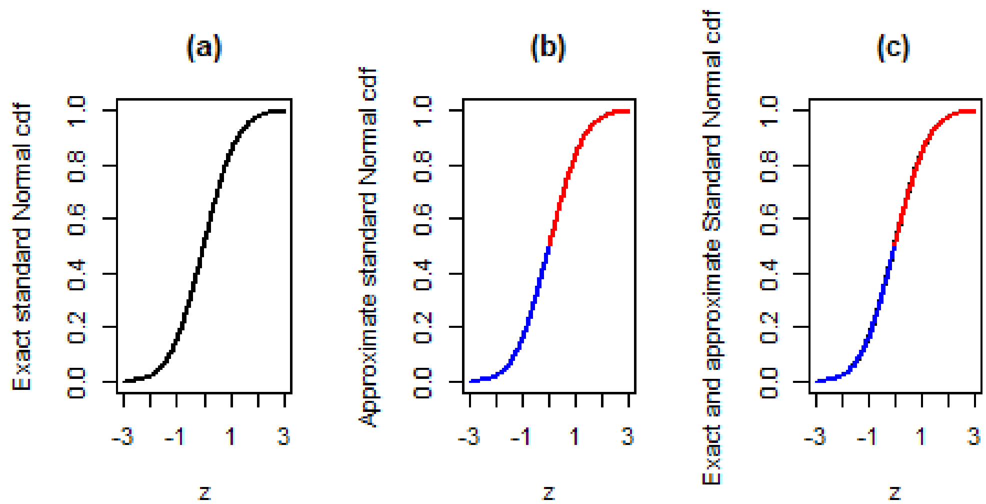

Figure 3.1: (a) Exact standard Normal Cumulative Distribution Function, (b) Modified Lin's approximation of standard Normal Cumulative Distribution Function, (c) Overlapping of the two functions.

$\Phi^*(-z) = 1$. Figure 3.1 shows the curve of $\Phi^\star(z)$ and $\Phi(z)$ in plot (a) and plot (b), respectively. Plot (c) shows that the two functions overlap against $z \in [-3, 3]$. Moreover, by considering interval $z \in [-3, 3]$ with 200 points and the calculation of $\Phi(z)$ and $\Phi^\star(z)$ at the corresponding points, we obtain $\frac{\sum_{i=1}^{200} |\Phi(z_i) - \Phi^\star(z_i)|}{200} = 0.002$. This term yields small value, which can be negligible.

Hence, we can substitute $\Phi^\star(q_j(z))$ for $\Phi(q_j(z))$ in (3.12) and (3.13), respectively and solve the integrals. The results are given in Appendix A, Section A.2. In the next subsection, the approximation of the covariance elements in a specific example of the binary random intercept probit model is evaluated.

Assessment of new approximations

In this section, an example of the binary probit mixed effects model is considered. Assume that $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, $\boldsymbol{\zeta}_i = \zeta_{i0}$, $\boldsymbol{f}(\boldsymbol{x}_j) = (1, x_j)^\top; j = 1, 2$, $\boldsymbol{h}(\boldsymbol{x}_j) = 1$; $(x_1, x_2) = (0, 1)$, $\sigma^2 = 1$. Then, the model is called the binary random intercept probit regression model. The marginal covariances of the response variable (3.3) and (3.4) are computed based on the three methods of approximations as:

1) **Numerical approximation.** The numerical computation of (3.3) and (3.4) based on the Simpson's rule.

2) **First new approximation.** Firstly, covariances (3.3) and (3.4) are obtained in correspondence with Lemma 3.1.2. Then, the results, which are in terms of unclosed integrals, are computed by the Simpson's rule.
3) **Second new approximation.** The unclosed integrals in (3.12) and (3.13) are approximated by **Rule 2** and **Rule 1** in Appendix A.2, respectively.

Figure 3.2 shows $\text{cov}(Y_{ijk}, Y_{ijk'}); j = 1, 2$ in Plots (a) and (b) against $\beta_0 \in [-3, 3]$ at $\beta_1 = 1$. The black curve indicates the numerical approximation method. The red curve indicates the first new approximation method, and the blue curve shows the second new approximate covariance. It illustrates that for most values of $\beta_0 \in [-3, 3]$, the three methods of the computation of the covariance overlap. However, for β_0 near to zero, the second approximation of the covariances $\text{cov}(Y_{i1k}, Y_{i1k'})$, $\text{cov}(Y_{i2k}, Y_{i2k'})$ and $\text{cov}(Y_{i1k}, Y_{i2k'})$ exceed the other two methods. The reason for this is due to the approximation of $\Phi(.)$ by $\Phi^\star(.)$ in (3.21).

Moreover, the corresponding covariance elements are computed against $\beta_1 \in [-3, 3]$ at $\beta_0 = 1$. They are shown in Figure 3.3. Plot (a) shows that $\text{cov}(\mathrm{Y_{i1k}}, \mathrm{Y_{i1k'}})$ is constant for all $\beta_1 \in [-3, 3]$. The reason for this is that $x_1 = 0$ and β_1 is the slope parameter. The numerical approximation and the first approximation of $\text{cov}(Y_{i1k}, Y_{i1k'})$ overlap, whereas, the second approximation deviates slightly from the first two methods, which takes slightly larger values. Plot (b) illustrates $\text{cov}(Y_{i2k}, Y_{i2k'})$. It is obvious that the three methods overlap in more frequent cases of $\beta_1 \in [-3, 3]$; however, similar to Figure 3.2 for β_1 close to zero the second approximations slightly shows different values than the first two approximations. This result also happens for $\text{cov}(Y_{i1k}, Y_{i2k'})$.

Furthermore, the determinant of variance matrix $\mathbf{V}$ is also computed based on the three methods. For this aim, we consider a fixed approximate design such as:

$$\xi_i = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}, \tag{3.22}$$

where $n = 50$, $n_1 = n_2 = 25$. The determinant of $\mathbf{V}$ in power of $\frac{1}{n} = \frac{1}{50}$ is computed against $\beta_0 \in [-3, 3]$ at $\beta_1 = 1$ according to the three methods. The results are displayed in Figure 3.4. It shows that the three methods overlap to high extend. Moreover, the determinant of the individual quasi Fisher information matrix against $\beta_0 \in [-3, 3]$ at $\beta_1 = 1$ is computed and it is shown in Figure 3.5. It shows that the first two approximations overlap to high extent. However, this is not right for the second approximation and it rises above the first two approximations for $\beta_0 \in (-1, 2)$.

Furthermore, the determinant of the individual variance matrix in power of $[\frac{1}{50}]$ against $\beta_1 \in [-3, 3]$ at $\beta_0 = 1$ is outlined in Figure 3.6. It shows that this quantity obtains its higher values at $\beta_1 \in (-2, 0)$. Figure 3.7 illustrates the determinant of the individual quasi Fisher information matrix against $\beta_1 \in [-3, 3]$

at $\beta_0 = 1$. This Figure shows that the numerical approximation and first new approximation overlap. However, the second new approximation deviates from the first two methods and it shows slightly a different trend.

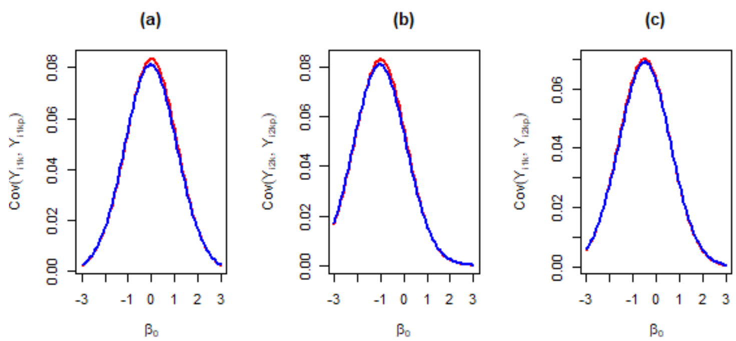

Figure 3.2: Entries of the covariance matrix in dependence on β_0 (probit link)

.

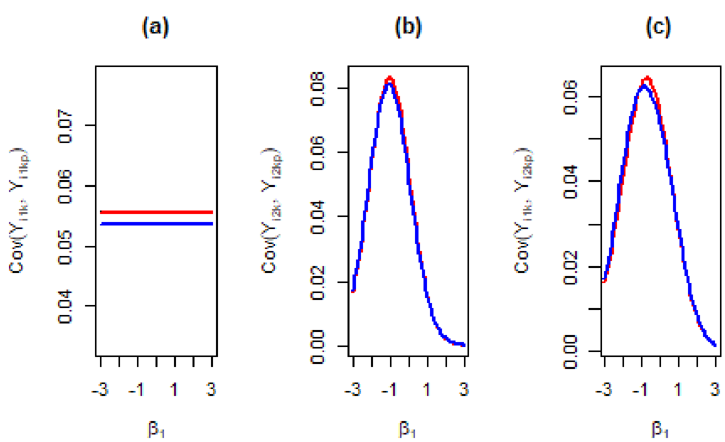

Figure 3.3: Entries of the covariance matrix in dependence on β_1 (probit link)

.

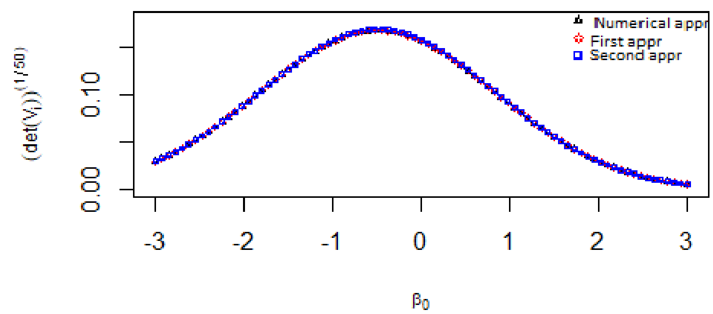

Figure 3.4: Variance matrix determinant in power of $\frac{1}{50}$ in dependence on β_0, (probit link)

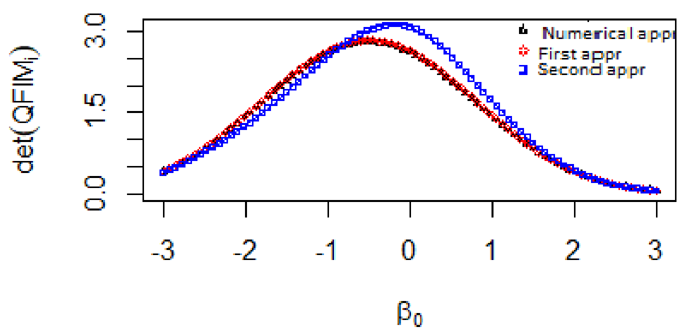

Figure 3.5: Quasi Fisher information matrix determinant in dependence on β_0 (probit link).

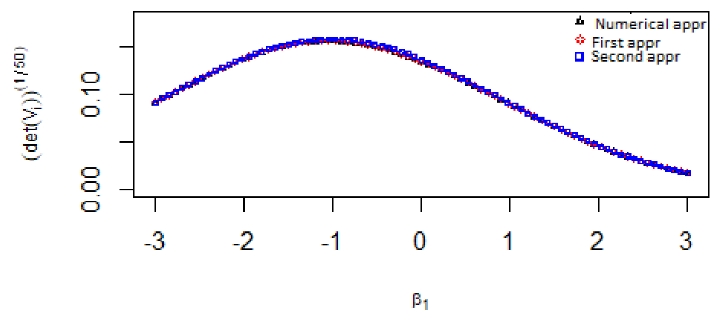

Figure 3.6: Variance matrix determinant in power of $\frac{1}{50}$ in dependence on β_1, (probit link)

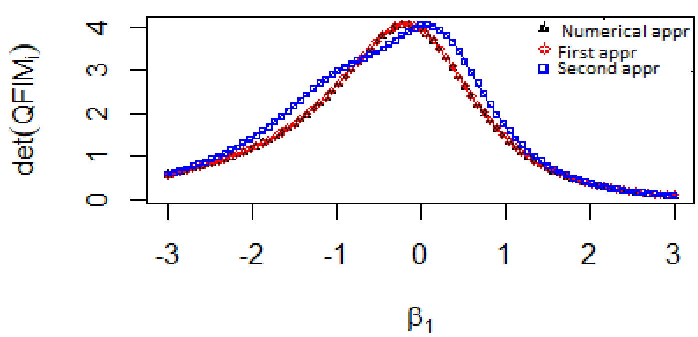

Figure 3.7: Quasi Fisher information matrix determinant in dependence on β_1 (probit link).

In general, we conclude that the first two approximations for the approximation of the quasi Fisher information matrix for the special case of the random intercept model work well. The third method deviates hardly in non extreme values of the population parameters.

In the next subsection, the components of the quasi Fisher information matrix are built for the binary random intercept logistic regression model.

Quasi Fisher information matrix in model with logit link function

In this section, the binary mixed effects model with the logit link function is considered. We demonstrate the methods to approximate the variance matrix and also the quasi Fisher information matrix. First of all, we deal with the achievement of the marginal first order moment of the response variable. This term was already discussed in Zeger, Liang, and Albert 1988 and was eventually approximated based on the following approach:

$$\begin{aligned}\pi_j &= \int_{\mathbb{R}^q} \frac{\exp(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+h^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i})}{1+\exp(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+h^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i})}\phi_{\mathbf{0},\sigma^2\mathbf{I}_q}(\boldsymbol{\zeta}_i)\mathrm{d}\boldsymbol{\zeta}_i \\ &\approx \int_{\mathbb{R}^q}\int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+h^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i}} \phi_{0,\frac{1}{c^2}}(u)\mathrm{d}u\phi_{\mathbf{0},\sigma^2\mathbf{I}_q}(\boldsymbol{\zeta}_i)\mathrm{d}\boldsymbol{\zeta}_i \\ &\approx \Phi(\alpha_{j(c)}). \end{aligned} \tag{3.23}$$

where $\alpha_{j(c)} = (1+c^2\sigma^2h^\top(\boldsymbol{x}_j)h(\boldsymbol{x}_j))^{-\frac{1}{2}}\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}$, in which $c=\frac{16\sqrt{3}}{15\pi}$. The second approximate equality is gained from the approximation of the logistic cumulative distribution function $\frac{\exp(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+h^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i})}{1+\exp(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i})}$ with the cumulative distribution function of the Normal distribution $\Phi_{0,\frac{1}{c}}(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+h^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i})$, with zero mean and standard deviation $\frac{1}{c}$. The third approximate equality is obtained from the direct calculation of the corresponding integral, in the second approximate equality. Further, $\mathrm{Var}(Y_{ijk}) \approx \pi_j(1-\pi_j)$ in (3.2). In order to approximate covariance elements (3.3) and (3.4), which are based on the second order moment of $p_{ij}(\boldsymbol{\zeta}_i)$, we concise the model to the binary random intercept Logistic model. The corresponding covariance elements are already obtained in Lemma 3.1.2, Appendix A.2 for the binary random intercept probit regression model. The only difference to this situation is that σ^2 in the previous results is transformed to $c^2\sigma^2$; however, the final form of the covariance elements in the binary logistic mixed effects regression model is stated in Appendix A, section A.3.

In addition, matrix $\mathbf{D}$ is constructed in correspondence with Lemma A.3.2, Appendix A. According to this lemma, matrix $\mathbf{D}$ for the specific case of the model as the random intercept logit model is obtained based on the form of $\mathbf{D}_j$ as follows:

$$\mathbf{D}_j \approx (c^2\sigma^2+1)^{-\frac{1}{2}} \boldsymbol{f}(\boldsymbol{x}_j)\phi(\alpha_{j(c)}). \tag{3.24}$$

Checking the properties of quasi Fisher information matrix

In each specific example of the considered models, it is required to evaluate the positive-definiteness of the individual variance matrix, $\mathbf{V}$. For the proof of the following lemma, let

$$\boldsymbol{\Upsilon} = \begin{pmatrix} \pi_1(1-\pi_1)+(n_1-1)v_1 & n_1c_{12} \\ n_2c_{21} & \pi_2(1-\pi_2)+(n_2-1)v_2 \end{pmatrix},$$

where $v_j = \text{cov}(Y_{ijk}, Y_{ijk'}); k \neq k'$ and $c_{jj'} = \text{cov}(Y_{ijk}, Y_{ij'k'}); j \neq j', k, k'$.

Lemma 3.1.4. *In binary mixed effects regression model, considering the two point design, individual variance matrix* $\mathbf{V}$ *is positive definite, if*
1) $\det(\boldsymbol{\Upsilon}) > 0$
2) For $j = 1, 2$, $(\pi_j(1-\pi_j) - v_j) > 0$

Proof.
The proof is stated in Appendix A, Section A.4 ■

In case condition (1) in Lemma 3.1.4 is satisfied for two point design; then, $\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta})) > 0$. Since $\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta}))$ is written in correspondence with

$$\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta})) = \frac{1}{[\det(\Upsilon)]^2} \det(\boldsymbol{Q})[\det(\mathbf{D}^{(0)})]^2,$$

where $\mathbf{D}^{(0)} = (\mathbf{D}_1, ..., \mathbf{D}_J)^\top$ with $p \times J$ dimension and $\mathbf{D}_j = (\frac{\partial \pi_j}{\partial \beta_0}, ..., \frac{\partial \pi_j}{\partial \beta_{p-1}})^\top$; $j = 1, ..., J$ and

$$\boldsymbol{Q} = n_1n_2 \begin{pmatrix} v_2 & -c_{12} \\ -c_{21} & v_1 \end{pmatrix} + \begin{pmatrix} n_1(\pi_2(1-\pi_2)-v_2) & 0 \\ 0 & n_2(\pi_1(1-\pi_1)-v_1) \end{pmatrix}, \tag{3.25}$$

Here, $p = 2$. $\det(\boldsymbol{Q}) = n_1n_2 \det(\boldsymbol{\Upsilon})$. Thus,

$$\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta})) = \frac{n_1n_2}{\det(\Upsilon)}[\det(\mathbf{D}^{(0)})]^2 > 0.$$

which is a 2×2 matrix.

On the other way around, to verify the positive definitness of the quasi Fisher information matrix, it is also possible to check the positiveness of the corresponding eigenvalues. Under condition (2) in Lemma 3.1.4, it is shown that the diagonal elements of the quasi Fisher information matrix are positive (See formula 3.25, Appendix A, Section A.5). Hence, we conclude that the eigenvalues of the quasi Fisher information matrix are positive and therefore the quasi Fisher information matrix is a positive definite matrix.

3.2 Evaluation of quasi maximum-likelihood estimation

In this section, the performance of the quasi maximum likelihood estimate of parameter $\boldsymbol{\beta}$ is evaluated in a special case of the binary random intercept probit regression model. For this aim, 10000 random samples are taken from model (2.4), with the basic assumptions of $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top = (1, -1)^\top$, $\boldsymbol{f}^\top(\boldsymbol{x}_j) = (1, x_j)$; $(x_1, x_2) = (0, 1)$. $\boldsymbol{h}^\top(x_j) = 1$, and $\boldsymbol{\zeta}_i = \zeta_{i0}$, $\sigma^2 = 1$. Furthermore, $i = 1, ..., 120$, which means that $N = 120$. Uniform design (3.22) is used with equal weights $\frac{1}{2}$ at support points 0 and 1. For instance, $n = 50$ is the number of replication within each subject. $n_1 = n_2 = 25$ are the number of replication at the experimental settings x_1 and x_2 respectively. Therefore, there are 120×50 observations at each time of sampling. The quasi maximum-likelihood estimate of $\boldsymbol{\beta}$ is obtained from the Newton-Raphson method. The mean and variance of $\hat{\boldsymbol{\beta}}_{QL}$ are obtained as follows:

$$\overline{\hat{\boldsymbol{\beta}}_{QL}} = \textstyle\sum_{i=1}^{10000} \frac{\hat{\boldsymbol{\beta}}_{QL(i)}}{10000} = (1.002, -1.003)^\top$$

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_{QL}) = \begin{pmatrix} 0.011 & -0.002 \\ -0.002 & 0.003 \end{pmatrix}$$

and $[\det(\mathrm{Var}(\hat{\boldsymbol{\beta}}_{QL}))]^{\frac{1}{2}} = [2.7 \times 10^{-5}]^{\frac{1}{2}} = 5.2 \times 10^{-3}$. Moreover, by setting $N^\star = nN$, $[\det(\mathrm{Var}(N^{\frac{1}{2}}\hat{\boldsymbol{\beta}}_{QL}))]^{\frac{1}{2}} = [10^4 \times 2.7 \times 10^{-5}] = 0.52$.

Figure 3.8 illustrates that the scatter plot of $\hat{\boldsymbol{\beta}}_{QL}$ follows an elliptical shape which calls for the bivariate Normal distribution. Moreover, the qq-plots of $\hat{\beta}_{QL0}$ and $\hat{\beta}_{QL1}$ in Figures 3.9 and 3.10 show that the 10000 simulated estimators are following the Normal distribution. Finally, Anderson-Darling test for Multivariate Normality (Thode 2002) is done and the result shows that it is not rejected that $\hat{\boldsymbol{\beta}}_{QL}$ does not follow the bivariate Normal distribution, as $p - value = 0.41$. The test is done by statistical software R, package mvnTest.

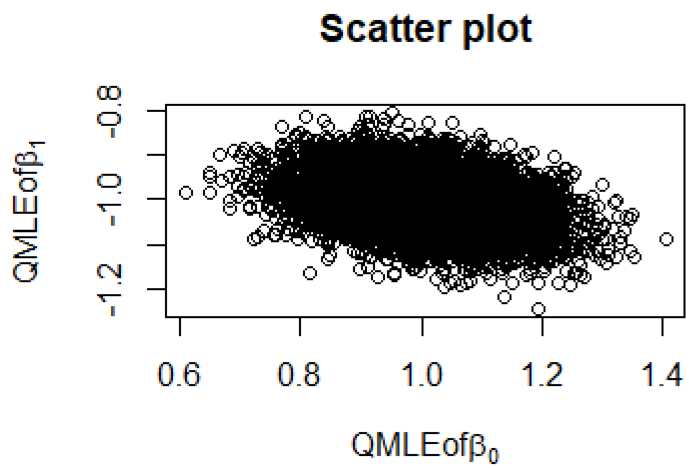

Figure 3.8: Scatter plot of the quasi maximum-likelihood estimates of $\boldsymbol{\beta}$ (probit link)

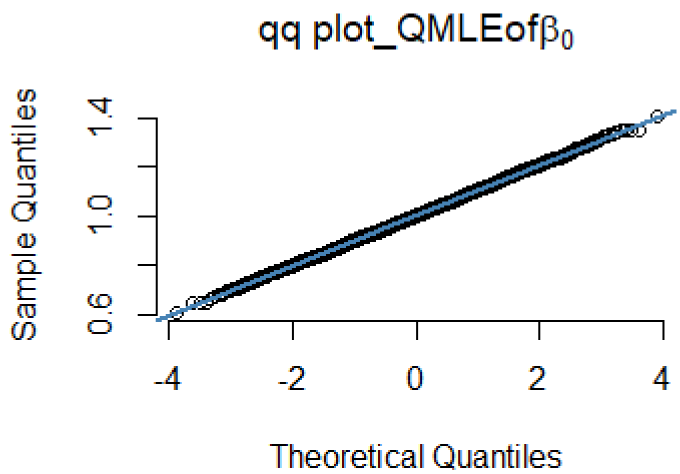

Figure 3.9: qq plot of the quasi maximum likelihood estimate of β_0 (probit link)

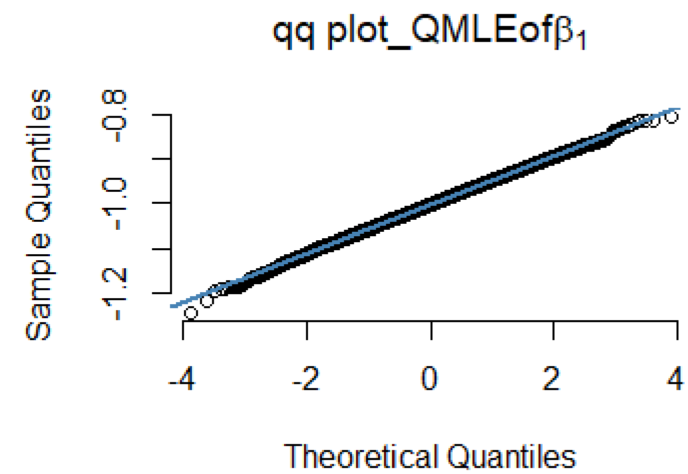

Figure 3.10: qq plot of the quasi maximum likelihood estimate of β_1 (probit link)

3.3 D-optimum design

In this section, we aim to obtain two point D-optimum design for the special case of the random intercept probit regression model. In the subsequent Theorem the direct approach from the mathematical calculus is attempted to gain the D-optimum design.

Theorem 3.3.1. *In the random intercept binary regression model, with two fixed experimental settings, as $x_1 = 0$ and $x_2 = 1$, and with sample size n, the D-optimum weight $w_1^\star$ at x_1 and $w_2^\star$ at x_2 are given by:*

$$\xi^\star = \begin{pmatrix} x_1 & x_2 \\ w_1^* & w_2^* \end{pmatrix}, \tag{3.26}$$

where

$$w_1^* = \frac{-\lambda_{n1} + \sqrt{\lambda_{n2}}}{n\lambda_d}, \; w_2^* = 1 - w_1^*, \tag{3.27}$$

with

$\lambda_{n1} = (\pi_1(1-\pi_1) - v_1)(\pi_2(1-\pi_2) + (n-1)v_2),$
$\lambda_{n2} = (\pi_1(1-\pi_1) - v_1)(\pi_2(1-\pi_2) - v_2)(\pi_1(1-\pi_1) + (n-1)v_1)(\pi_2(1-\pi_2) + (n-1)v_2),$
$\lambda_d = v_1\pi_2(1-\pi_2) - v_2\pi_1(1-\pi_1),$

where $\lambda_d \neq 0$ *and* $w_1^\star \in [0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$.

Moreover,

$$w_1^\star = \frac{1}{2}; w_2^\star = 1 - w_1^\star, \tag{3.28}$$

where $\lambda_d = 0$.

The above results show that the D-optimum weight does not depend on $c_{jj'}; j \neq j'$.

Proof.

The proof is stated in Appendix A, Section A.5. ■

Examples.

Considering the random intercept probit regression model with $\beta_0 \in [-3, 3]$, as $\beta_1 = 1$, $\sigma^2 = 1$, fixed at $x_1 = 0, x_2 = 1$; the D-optimum weights are aimed to be gained by the maximization of D-optimality criterion (2.75), with respect to the weight w_1. For this reason, Theorem 3.3.1 is applied. The three approximations of the covariance elements in Section 3.1 are utilized and the D-optimum weight $w_1^\star$ is sketched against $\beta_0 \in [-3, 3]$, fixed at $\beta_1 = 1$ in Figure 3.11. Furthermore, the D-optimum weight $w_1^\star$ against $\beta_1 \in [-3, 3]$, fixed at $\beta_0 = 1$ is shown in Figure 3.12. The three approximations are somehow overlay in both figures. The slight difference can be shown for the first approximation for higher values of β_0 and β_1 near to 3.

Moreover, the D-optimum design $\xi^\star$ is already calculated for different values of $n = \{50, 100, 300, 500, 800, 1000\}$ using Theorem 3.3.1. Figure 3.13 shows the D-optimum wight $w_1^\star$ against $\beta_0 \in [-3, 3]$, fixed at $\beta_1 = 1$. For $\beta_0 \in (-1, 1)$, the D-optimum weights are virtually overlapping for all n's. However, for more extreme values of $\beta_0 \in [-3, -1) \cup (1, 3]$, the D-optimum weights regarding different n give a slight different answer. For more extreme values of β_0 we consider that n can affect on the value of $w_1^\star$, which means that $w_1^\star$ is quite sensitive with respect to the changes of the sample size n for extreme values of $\beta_0 \leq 3$.

Figure 3.14 reveals the D-optimum weight against $\beta_1 \in [-3, 3]$ fixed at $\beta_0 = 1$. . For most values of $\beta_1 \in [-3, 1]$ the D-optimum weights coincide; however, as β_1 exceeds one, the D-optimum weight for larger n should be taken much less at $x_1 = 0$ and therefore more at $x_2 = 1$.

Finally, D-optimum weights are obtained For the considered different n and for different $\beta_0 = -3, -2, -1, 0, 1, 2, 3$ at $\beta_1 = 1$. The results are shown in Table 3.1. For all n as β_0 rises up, the D-optimum weight decreases and for larger n, the degree of the decrease is sharper. Moreover, $w_1^\star$ for $\beta_1 = \{-3, -2, -1, 1, 2, 3\}$, fixed at β_0 is obtained and it is illustrated in Table 3.2 for different n. The higher values of $w_1^\star$ is achieved in the middle values of β_1.

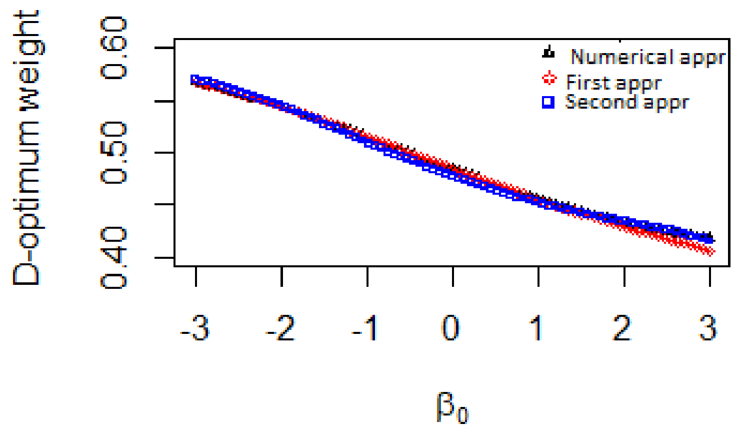

Figure 3.11: D-optimum weight $w_1^{\star}$ in dependence on β_0, (probit link)

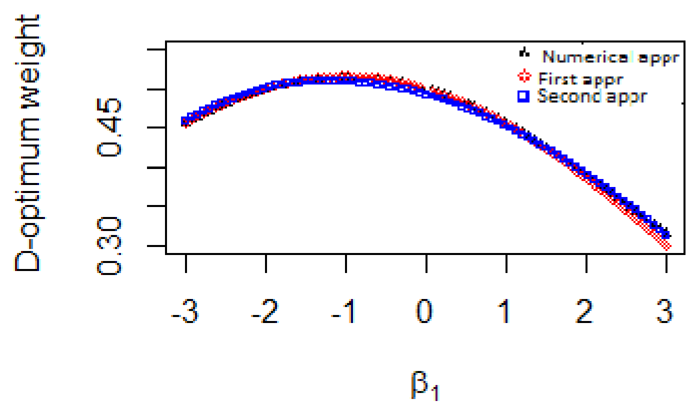

Figure 3.12: D-optimum weight $w_1^{\star}$ in dependence on β_1, (probit link)

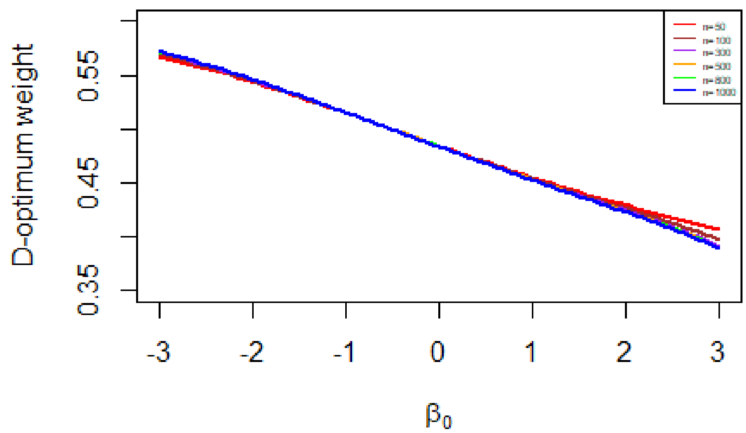

Figure 3.13: D-optimum weight $w_1^\star$ in dependence on β_0, (probit link)

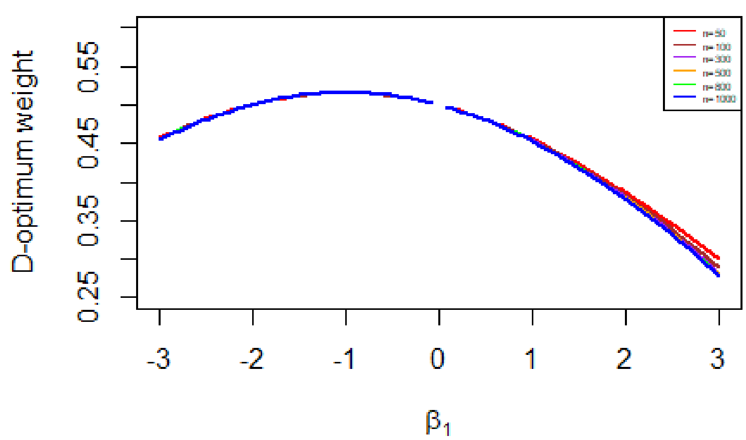

Figure 3.14: D-optimum weight $w_1^\star$ in dependence on β_1, (probit link)

Table 3.1: D-optimum weight, random intercept probit model

$n \mid \beta_0$	-3	-2	-1	0	1	2	3
50	0.567	0.544	0.515	0.484	0.455	0.429	0.406
100	0.570	0.545	0.515	0.484	0.454	0.426	0.398
300	0.572	0.546	0.516	0.484	0.453	0.424	0.391
500	0.572	0.546	0.516	0.484	0.453	0.423	0.390
800	0.573	0.546	0.516	0.484	0.453	0.423	0.389
1000	0.573	0.546	0.516	0.484	0.452	0.423	0.389

Table 3.2: D-optimum weight, $w_1^\star$, random intercept probit model

$n \mid \beta_1$	-3	-2	-1	1	2	3
50	0.457	0.501	0.516	0.455	0.385	0.300
100	0.456	0.501	0.516	0.454	0.381	0.289
300	0.455	0.501	0.516	0.453	0.378	0.281
500	0.455	0.501	0.516	0.453	0.378	0.279
800	0.455	0.501	0.516	0.453	0.377	0.278
1000	0.455	0.501	0.516	0.452	0.377	0.278

Further research

In this dissertation, we only consider the random intercept binary regression model and achieve the two point D-optimum design with fixed design points $\{0, 1\}$. In order to approximate the covariance elements and the quasi Fisher information matrix for the general case of the binary mixed effects model, we obtain the approximation of covariance elements $\text{cov}(Y_{ijk}, Y_{ijk'})$ and $\text{cov}(Y_{ijk}, Y_{ij'k'}); \forall j, j', k, k'$, in the binary mixed effects model with two random effects, as:

$$\Phi^{-1}(p_{ij}(\boldsymbol{\zeta}_i)) = \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} + \zeta_{i0} + \zeta_{i1}\boldsymbol{x}_j, \tag{3.29}$$

where $\boldsymbol{\zeta}_i = (\zeta_{i0}, \zeta_{i1})^\top$.

This target, may helps to have more precise result and a general overview of the covariance element in a more extended case of the binary mixed effects regression model.

Lemma 3.3.2. *Let* $\alpha_{j1} = (1 + (1 + \boldsymbol{x}_j^2)^{-\frac{1}{2}}\sigma^2)\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}$ *and* $\varphi_{j,j'} = (\boldsymbol{x}_j^2 + \boldsymbol{x}_{j'}^2 + \frac{1}{\sigma^2} - \frac{(\boldsymbol{x}_j + \boldsymbol{x}_{j'})^2}{2 + \frac{1}{\sigma^2}})$*. In the binary mixed effects regression model with two random effects,*

$\text{cov}(Y_{ijk}, Y_{ij'k'}) =$

$$\int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta}} \int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}} \frac{1}{2\pi} \frac{1}{\sqrt{\varphi_{j,j'}(2+\frac{1}{\sigma^2})}} \exp(u^2+v^2-\frac{(u+v)^2}{2+\frac{1}{\sigma^2}}-\frac{(a_{1,j,j'}u+a_{1,j',j}v)^2}{\varphi_{j,j'}})\mathrm{d}u\mathrm{d}v$$

$$-\Phi(\alpha_{j1})\Phi(\alpha_{j'1}), \tag{3.30}$$

where $a_{1,j,j'} = (\boldsymbol{x}_j(1+\frac{1}{\sigma^2}) - \boldsymbol{x}_{j'})$.

Proof.

Based on the form of $\mathrm{cov}(Y_{ijk}, Y_{ij'k'})$ in (3.4), $\mathrm{E}[p_{ij}(\boldsymbol{\zeta}_i)p_{ij'}(\boldsymbol{\zeta}_i)]$ is needed to be obtained. It is stated as:

$$\mathrm{E}[p_{ij}(\boldsymbol{\zeta}_i)p_{ij'}(\boldsymbol{\zeta}_i)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\zeta_{i0}+\zeta_{i1}\boldsymbol{x}_j)\Phi(\boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta}+\zeta_{i0}+\zeta_{i1}\boldsymbol{x}_{j'})\mathbf{f}_{\mathbf{0},\Sigma}(\boldsymbol{\zeta}_i)\mathrm{d}\zeta_{i0}\mathrm{d}\zeta_{i1}, \tag{3.31}$$

in which $\mathbf{f}(.)$ denotes the density of random effect $\boldsymbol{\zeta}_i$, $\Sigma = \sigma\mathbf{I}_2$, and

$$\Phi(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\zeta_{i0}+\zeta_{i1}\boldsymbol{x}_j) = \int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\zeta_{i0}+\zeta_{i1}\boldsymbol{x}_j} \phi(t)\mathrm{d}t.$$

$$= \int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}} \phi_{-(\zeta_{i0}+\zeta_{i1}\boldsymbol{x}_j),1}(u)\mathrm{d}u \tag{3.32}$$

The last equality above is obtained by changing variable, $u = t - \zeta_{i0} - \zeta_{i1}\boldsymbol{x}_j$.

By the substitution of (3.32) for $\Phi(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\zeta_{i0}+\zeta_{i1}\boldsymbol{x}_j); j, j'$ in (3.31),

$$\mathrm{E}[p_{ij}(\boldsymbol{\zeta}_i)p_{ij'}(\boldsymbol{\zeta}_i)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}}\int_{-\infty}^{\boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta}} \exp(r(\zeta_{i0},\zeta_{i1}))\mathbf{f}_{\mathbf{0},\Sigma}(\boldsymbol{\zeta}_i)\mathrm{d}u\mathrm{d}v\mathrm{d}\zeta_{i0}\mathrm{d}\zeta_{i1} \tag{3.33}$$

$$r(\zeta_{i0},\zeta_{i1}) = -\tfrac{1}{2}\zeta_{i1}^2(\boldsymbol{x}_j^2+\boldsymbol{x}_{j'}^2+\tfrac{1}{\sigma^2}) - \tfrac{1}{2}\zeta_{i0}^2(2+\tfrac{1}{\sigma^2}) - \tfrac{1}{2}2\zeta_{i1}(u\boldsymbol{x}_j+v\boldsymbol{x}_{j'})$$
$$-\tfrac{1}{2}2(u+v+\zeta_{i1}(\boldsymbol{x}_j+\boldsymbol{x}_{j'}))\zeta_{i0} - \tfrac{1}{2}(u^2+v^2)$$

By solving integral (3.33) with respect to ζ_{i0} and ζ_{i1}, the first term in (3.30) is obtained. Finally, considering the second term in (3.4), $\mathrm{E}(p_{ij}(\boldsymbol{\zeta}_i))$ is required to be obtained. Due to Lemma 3.1.1, $\mathrm{E}(p_{ij}(\boldsymbol{\zeta}_i)) = \Phi(\alpha_{j1})$. Then, The result in (3.30) is obtained. ■

Actually, by setting $j = j'$, $\mathrm{cov}(Y_{ijk}, Y_{ijk'})$ can be obtained. Due to Lemma (3.1.3), the elements of matrix $\mathbf{D}$ can be obtained and at last quasi information matrix can be approximated.

The last question is that whether it is possible to generalize (3.30) for the general binary mixed effects regression model. To my mind, there should be more effort to find a general form of the covariance element for the general form of the binary mixed effects regression model.

Chapter 4

Optimum Design in Ordinal Mixed Effects Model

In this chapter, the D-optimum design is aimed to be achieved in the ordinal mixed effects model. For this purpose, the Fisher information matrix of the considered model has to be constructed. Since the Fisher information matrix lacks the analytical closed form, the quasi-Fisher information matrix is suggested to be obtained. It was firstly introduced by Wedderburn 1974. This matrix is based on the construction of the quasi log-likelihood function, which is built up in terms of the marginal first and second order moments of the response variable. We attempt to obtain the closed form of the considered moments. However, they lack the analytical explicit form. Therefore, we consider new approximations for them.

At last, the approximation of the quasi Fisher information matrix is set up and the relevant D-optimum design is obtained. This approach parallels the approach in Chapter 3 withe difference that now observations are multinomial (multivariate) instead of binomial (univariate binary). Since the individuals take the same design, the D-optimum design is obtained for individual and it is extended to the population D-optimum design.

4.1 Ordinal mixed effects model

The ordinal mixed effects model and the regarding link functions have already been stated in Section 2.1 (See formula (2.21)). In this section, we present the properties of the model in order to form the quasi log-likelihood function and the quasi Fisher information matrix.

Consider the response variable $\boldsymbol{Y}_{ij} = (\boldsymbol{Y}_{ij1}^\top, ..., \boldsymbol{Y}_{ijn_j}^\top)^\top$, where $\boldsymbol{Y}_{ijk} = (Y_{ijk}^{(1)}, Y_{ijk}^{(2)},$

$..., Y_{ijk}^{(M)})^\top$. Therefore, the response variable $\mathbf{Y}_{ij}$ is with dimension $Mn_j \times 1$ and the response $\mathbf{Y}_i = (\mathbf{Y}_{i1}^\top, ..., \mathbf{Y}_{iJ}^\top)^\top$ is with dimension $Mn \times 1$.

We define $\tilde{\boldsymbol{\pi}}_j := \mathrm{E}(\boldsymbol{Y}_{ij})$; then,

$$\begin{aligned}\tilde{\boldsymbol{\pi}}_j &= (\mathrm{E}^\top(\boldsymbol{Y}_{ij1}), \mathrm{E}^\top(\boldsymbol{Y}_{ij2}), ..., \mathrm{E}^\top(\boldsymbol{Y}_{ijn_j}))^\top \\ &= (\boldsymbol{\pi}_j^\top, \boldsymbol{\pi}_j^\top, ..., \boldsymbol{\pi}_j^\top)^\top \\ &= \mathbf{1}_{n_j} \otimes \boldsymbol{\pi}_j. \end{aligned} \tag{4.1}$$

where $\boldsymbol{\pi}_j = (\pi_j^{(1)}, ..., \pi_j^{(M)})^\top$ is an M-dimensional vector of probabilities; such that, $\pi_j^{(m)} = E(Y_{ijk}^{(m)}) = P(\tilde{\mathbf{Y}}_{ijk} = m), \sum_{m=1}^{M} \pi_j^{(m)} = 1$. Then, considering $\tilde{\boldsymbol{\pi}} := \mathrm{E}(\boldsymbol{Y}_i)$, we write

$$\tilde{\boldsymbol{\pi}} = (\tilde{\boldsymbol{\pi}}_1^\top, \tilde{\boldsymbol{\pi}}_2^\top, ..., \tilde{\boldsymbol{\pi}}_J^\top)^\top. \tag{4.2}$$

In order to build up the individual variance matrix $\mathbf{V}$ of the response variable $\boldsymbol{Y}_i$, the variance components of the response variable are formulated as follows:

$$\mathrm{Var}(\boldsymbol{Y}_{ijk}) = [\mathrm{cov}(Y_{ijk}^{(m)}, Y_{ijk}^{(m')})]_{m,m'=1}^{M}, \tag{4.3}$$

$$\mathrm{Cov}(\boldsymbol{Y}_{ijk}, \boldsymbol{Y}_{ijk'}) = [\mathrm{Cov}(Y_{ijk}^{(m)}, Y_{ijk'}^{(m')})]_{m,m'=1}^{M}; k \neq k' \tag{4.4}$$

is an $M \times M$ covariance matrix of rank $M-1$ (as long as $\pi_j^{(m)} > 0$ for all m).

$$\mathrm{Cov}(\boldsymbol{Y}_{ijk}, \boldsymbol{Y}_{ij'k'}) = [\mathrm{Cov}(Y_{ijk}^{(m)}, Y_{ij'k'}^{(m')})]_{m,m'=1}^{M}; j \neq j', k, k' \tag{4.5}$$

where each matrix (2.24), (4.4) and (4.5) is with dimension $M \times M$. Each element of these matrices are obtained from the following terms:

For the computation of $\boldsymbol{V}_j := \mathrm{Var}(\boldsymbol{Y}_{ijk})$, the diagonal elements are obtained as:

$$\mathrm{Var}(Y_{ijk}^{(m)}) = \pi_j^{(m)}(1 - \pi_j^{(m)}); m = 1, ..., M, \tag{4.6}$$

and for $m \neq m'$ the off-diagonal elements are obtained as:

$$\mathrm{Cov}(Y_{ijk}^{(m)}, Y_{ijk}^{(m')}) = -\pi_j^{(m)}\pi_j^{(m')}, m, m' = 1, ..., M. \tag{4.7}$$

For the computation of $\boldsymbol{C}_j := \mathrm{Cov}(\boldsymbol{Y}_{ijk}, \boldsymbol{Y}_{ijk'})$; $k \neq k'$, each element is obtained as follows:

Consider $\mathrm{cov}(Y_{ijk}^{(m)}, Y_{ijk'}^{(m')}) := c_j^{(mm')}$; then,

$$c_j^{(mm')} = \mathrm{cov}(p_{ij}^{(m)}(\boldsymbol{\zeta}_i), p_{ij}^{(m')}(\boldsymbol{\zeta}_i)), m, m' = 1, ..., M. \tag{4.8}$$

For the computation of $\boldsymbol{C}_{jj'} := \text{Cov}(\boldsymbol{Y}_{ijk}, \boldsymbol{Y}_{ij'k'}); j \neq j', k, k'$, the diagonal and the off-diagonal blocks can be obtained from the following equation.

Consider $\text{cov}(Y_{ijk}^{(m)}, Y_{ij'k'}^{(m')}) := c_{jj'}^{(mm')}$; then,

$$c_{jj'}^{(mm')} := \text{cov}(p_{ij}^{(m)}(\boldsymbol{\zeta}_i), p_{ij'}^{(m')}(\boldsymbol{\zeta}_i)); m, m' = 1, ..., M, \tag{4.9}$$

in which $\text{cov}(p_{ij}^{(m)}(\boldsymbol{\zeta}_i), p_{ij'}^{(m')}(\boldsymbol{\zeta}_i))$. Then,

$$\mathbf{V}_{jj} = \mathbf{1}_{n_j}\mathbf{1}_{n_j}^{\top} \otimes \boldsymbol{C}_j + \mathbf{I}_{n_j} \otimes [\boldsymbol{V}_j - \boldsymbol{C}_j], \tag{4.10}$$

with $Mn_j \times Mn_j$ dimension and

$$\mathbf{V}_{jj'} = \mathbf{1}_{n_j}\mathbf{1}_{n_{j'}}^{\top} \otimes \boldsymbol{C}_{jj'}, \tag{4.11}$$

with dimension $Mn_j \times Mn_{j'}$. The marginal variance co-variance matrix of the response variable is stated as:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \dots & \mathbf{V}_{1J} \\ \mathbf{V}_{21} & \dots & \mathbf{V}_{2J} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{J1} & \dots & \mathbf{V}_{JJ} \end{pmatrix}, \tag{4.12}$$

where the dimension of the variance matrix is $Mn \times Mn$.

Since the variance-covariance matrix is singular, we remove the last element of the response variable $Y_{ijk}^{(M)} = 1 - \sum_{m=1}^{M-1} Y_{ijk}^{(m)}$. Then, it becomes non-singular and we could calculate the inverse of the reduced form of $\mathbf{V}$ with dimension $(M-1)n \times (M-1)n$, which is needed to obtain the quasi Fisher information matrix.

Referring to matrix $\mathbf{D}$ with $Mn \times p$ dimension as:

$$\mathbf{D} = (\mathbf{D}_1\mathbf{1}_{n_1}^{\top}, ..., \mathbf{D}_J\mathbf{1}_{n_j}^{\top})^{\top}, \tag{4.13}$$

such that, $\mathbf{D}_j = (\frac{\partial \boldsymbol{\pi}_j}{\partial \beta_0}, ..., \frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\beta}_{p-1}})^{\top}$; $j = 1, ..., J$ with dimension $p \times M$, and $\frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\beta}^{\top}} = (\mathbf{D}_j^{(1)\top}, ..., \mathbf{D}_j^{(M)\top})^{\top}$, where $\mathbf{D}_j^{(m)} = \frac{\partial \pi_j^{(m)}}{\partial \boldsymbol{\beta}^{\top}}$ with dimension $1 \times p$. Matrices $\mathbf{D}$ and $\mathbf{D}_j$ have to be reduced by the corresponding $\mathbf{D}_j^{(M)}$ in the last row, due to the singularity problem of the individual variance matrix.

Quasi Fisher information matrix in model with probit function

In this section, the quasi Fisher information matrix is set up for the ordinal mixed effects probit regression model. We begin by constructing the relevant first and second order moments of the response variables.

Let

$$\alpha_j^{(m)} = (1+\sigma^2 \boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{h}(\boldsymbol{x}_j))^{-\frac{1}{2}}(\gamma_m - \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}); m = 1, ..., M-1,$$

and $\alpha_j^{(0)} = -\infty, \alpha_j^{(M)} = +\infty$ The marginal expectation of the m^{th} level of the response variable is obtained in the subsequent lemma:

Lemma 4.1.1. *In ordinal mixed effects regression model with probit link function,*

$$\pi_j^{(m)} = \Phi(\alpha_j^{(m)}) - \Phi(\alpha_j^{(m-1)}); m = 1, ..., M. \tag{4.14}$$

Proof.
Since

$$\begin{aligned} \mathrm{E}[Y_{ijk}^{(m)}] &= \mathrm{E}[\mathrm{E}(Y_{ijk}^{(m)} \mid \boldsymbol{\zeta}_i)] \\ &= \mathrm{E}_{\boldsymbol{\zeta}_i}[g^{-1}(\delta_{ij}^{(m)}) - g^{-1}(\delta_{ij}^{(m-1)})] \\ &= \Phi(\alpha_j^{(m)}) - \Phi(\alpha_j^{(m-1)}), \end{aligned} \tag{4.15}$$

where

$$\delta_{ij}^{(m)}(\boldsymbol{\zeta}_i) = \gamma_m - \eta_{ij}.$$

and the last equality is obtained from Zeger, Liang, and Albert 1988, which is obtained from the direct calculation of the integral. ■

The diagonal variance components, $\boldsymbol{V}_j$ is directly obtained from $\pi_j^{(m)}$, and it is only needed to substitute $\Phi(\alpha_j^{(m)}) - \Phi(\alpha_j^{(m-1)})$ for $\pi_j^{(m)}$ in (4.6) and (4.7).

To calculate the covariance elements, we now calculate $\mathrm{E}[p_{ij}^{(m)}(\boldsymbol{\zeta}_i)p_{ij'}^{(m')}(\boldsymbol{\zeta}_i)]; \forall j, j', m, m'$ using equations (4.8) and (4.9) and state them in Lemma 4.1.2.

To obtain the second order moment of the response variable, we note that the answer can be obtained from the simpler model (2.21), which is a random intercept ordered regression model. Since it is more straightforward to obtain the results in a special case. Later in this chapter, we discuss on the achievement of the covariance elements in the more general case of the model.

Lemma 4.1.2. *Let* $q_{j0}^{(m)}(z) = \frac{(1+\sigma^2)^{\frac{1}{2}}}{(1+2\sigma^2)^{\frac{1}{2}}}(\gamma_m - \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} - \frac{\sigma^2 z}{(1+\sigma^2)^{\frac{1}{2}}})$*; then, in the random intercept ordered probit model for all* $j, j' = 1, ..., J, m, m' = 1, ..., M$*.*

$$\mathrm{E}[p_{ij}^{(m)}(\zeta_{i0})p_{ij'}^{(m')}(\zeta_{i0})] = \int_{-\infty}^{\alpha_{j0}^{(m)}} \phi(z)\Phi(q_{j'0}^{(m')}(z))\mathrm{z} \tag{4.16}$$

Proof.

The proof is the same as the proof in Lemma 3.1.2 for the computation of $\mathrm{E}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})]$ for binary mixed effects regression models, except for the substitution of $(\gamma_m - \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta})$ for $\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}$. ■

Due to the form of $\pi_j^{(m)}$ in (4.14), variance elements (4.6) and (4.7) can be calculated. The variance matrix can be constructed in terms of the new terms. As the marginal second order moment of the response variable in the random intercept ordered probit model is in terms of the one dimensional integral, we need to compute it with numerical Simpson's rule method. The components of matrix $\mathbf{D}$ are constructed in the subsequent lemma.

Lemma 4.1.3. *The components of matrix* $\mathbf{D}$ *in (2.46) for the ordinal mixed effects probit regression model, is formulated as follows:*

$$\mathbf{D}_j^{(m)} = -\boldsymbol{f}^\top(\boldsymbol{x}_j)(1+\sigma^2\boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{h}(\boldsymbol{x}_j))^{-\frac{1}{2}}[\phi(\alpha_j^{(m)}) - \phi(\alpha_j^{(m-1)})], \tag{4.17}$$

with dimension $1 \times p$.

Proof. The proof is obtained from $\frac{\partial}{\partial \boldsymbol{\beta}^\top}\pi_j^{(m)}$ in (4.14). ■

The dimension of matrix $\mathbf{D}$ is with dimension $Mn \times p$. As for the covariance matrix $\mathbf{V}$, also the matrix $\mathbf{D}$ has to be reduced accordingly by omitting the same class; then, it contains $(M-1)n \times p$ dimension.

According to Lemma 4.1.2, the second order moment of the response variable is dependent upon the unclosed integral. In the random intercept binary regression model (Chapter 3), this integral is approximated based on the modified Lin's method. However, this approach is not valid for the random intercept ordered regression model. Since using M levels, there is a need to approximate the second order moment of the response variable at each level; therefor applying the modified Lin's method leads to less precise approximation of the off-diagonal covariance elements. Because of this, it is decided to utilize two approximations of the covariance elements, which are stated subsequently:

1- **Numerical approximation.** The numerical computation of $\mathrm{E}[p_{ij}^{(m)}(\zeta_{i0})p_{ij'}^{(m')}(\zeta_{i0})]$ based on the Simpson's rule.

2- **First new approximation.** Firstly the covariance elements are obtained based on Lemma 4.1.2; then, the unclosed integral in (4.16) is computed based on the Simpson's rule.

Finally, by the substitution of the new approximations of covariance and formula (4.17) for the components of $\mathbf{D}$, the quasi Fisher information matrix $\mathbf{M}^Q(\xi,\boldsymbol{\beta})$ in (2.69) is formulated.

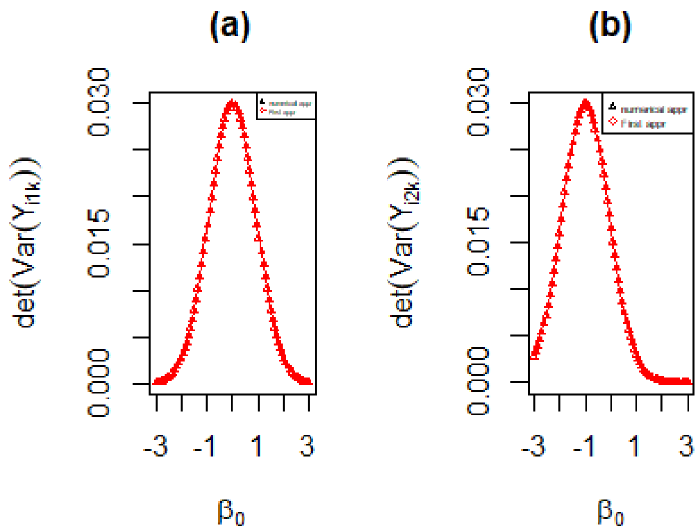

Figure 4.1: Variance matrix determinant of $\boldsymbol{Y}_{ijk}$, $\boldsymbol{V}_j$ (probit link) (a): $j = 1$, (b): $j = 2$

Assessment of new approximation

In this section, the components of variance matrix $\mathbf{V}$ and the quasi Fisher information matrix $\mathbf{M}^Q(\xi, \boldsymbol{\beta})$ are computed based on the numerical approximation and the first new approximation methods, in a specific case of ordinal mixed effects regression model. It is assumed that $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ and consider a special case, where x is the one dimensional straight line regression as the general experimental setting; then, $\boldsymbol{f}^\top(x) = (1, x)$, $\boldsymbol{\zeta}_i = \zeta_{i0}$ and $\boldsymbol{h}(x) = 1$. In addition, the dispersion parameter $\sigma^2 = 1$. We consider the two-point design, with fixed support points $x_1 = 0$ and $x_2 = 1$, and $n_1 = n_2 = 25$. Therefore, $n = 50$. Regarding the response variable, we presume $M = 3$, $\gamma_1 = -1$ and $\gamma_2 = 1$. Therefore, $Y_{ijk} = (Y_{ijk}^{(1)}, Y_{ijk}^{(2)}, Y_{ijk}^{(3)})^\top$. Also, we choose γ_0 as an absolute small value and γ_3 as an absolute large value. Each component $\boldsymbol{V}_j$, $\boldsymbol{C}_j$ and $\boldsymbol{C}_{jj'}$ in (2.24), (4.4) and (4.5), respectively contains 3×3 dimension and due to the singularity problem, the last level of the response variable is omitted. Figure 4.1 shows the determinant of $\boldsymbol{V}_j; j = 1, 2$ against β_0, when $\beta_1 = 1$. In order to search that variance matrix $\mathbf{V}$ in (4.12) is nonsingular, we obtain

$$\det(\mathbf{V}) = \det \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$$

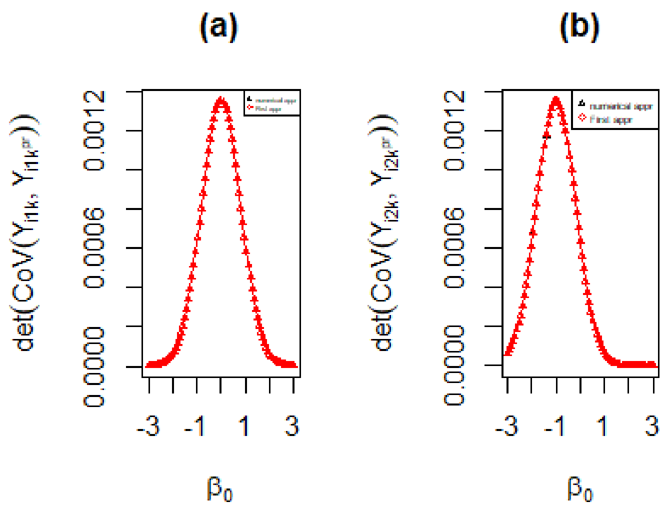

Figure 4.2: Covariance matrix determinant of $\boldsymbol{Y}_{ijk}$ and $\boldsymbol{Y}_{ijk'}$, $\boldsymbol{C}_j$ (probit) (a): $j = 1$, (b): $j = 2$

$$= \det(\mathbf{V}_{11}) \left(\det(\mathbf{V}_{22} - \mathbf{V}_{12}\mathbf{V}_{11}^{-1}\mathbf{V}_{21})\right)$$

In case $\det(\mathbf{V}) > 0$; then, the result is gained. Its maximum value is 0.030. For extreme values of β_0, which tends to -3 or 3, the variance tends zero.

Figure 4.2 illustrates the determinant of $\boldsymbol{C}_j; j = 1, 2, \forall k \neq k'$; against β_0, when $\beta_1 = 1$. Its highest value occurs approximately in the center of β_0 interval, and its lowest values are near to zero, which occurs at the most extreme values of $\beta_0 \in [-3, 3]$. Subsequently, the determinant of $\boldsymbol{C}_{jj'}; j = 1, 2, \forall j \neq j', k, k'$ is plotted in Figure 4.3. Note that $C_{21} = C_{12}^{\top}$. Hence, the determinants coincide, which means that $\det(C_{21}) = \det(C_{12})$. Its maximum value is considerably low and also, similar to the previous two graphs, its minimum values take place at the extreme points of $\beta_0 \in [-3, 3]$. Finally, the determinant of the general variance matrix $\mathbf{V}$ in power of $(\frac{1}{100})$ is sketched in Figure 4.4. The whole dimension of matrix $\mathbf{V}$ is 100×100. It takes its maximum value at around $\beta_0 = 0$ and its minimum values at the edge of the interval of β_0. Moreover, the determinant of the individual quasi Fisher information matrix against $\beta_0 \in [-3, 3]$, as $\beta_1 = 1$ is sketched in Figure 4.5. It takes its highest values on the interval $(-1, 0)$. In all of these figures the two methods of approximations overlap.

On the other side, the variance components, and the quasi Fisher information matrix against $\beta_1 \in [-3, 3]$, fixed at $\beta_0 = 1$ are plotted in Figures 4.6-4.10. Figure

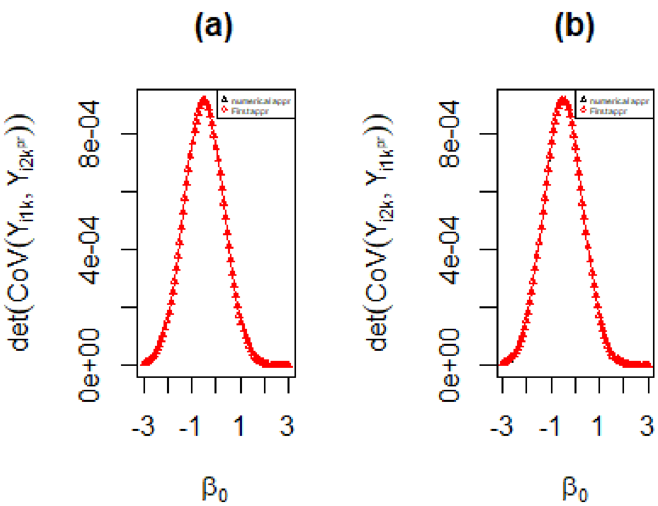

Figure 4.3: Covariance matrix determinant of $\boldsymbol{Y}_{ijk}$ and $\boldsymbol{Y}_{ij'k'}$, $\boldsymbol{C}_{jj'}$ $j \neq j'$ (probit) (a): $j = 1, j' = 2$, (b): $j = 2, j' = 1$

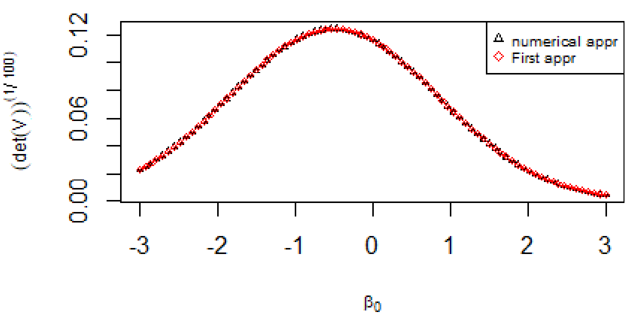

Figure 4.4: Determinant of variance matrix $\mathbf{V}$ in power of $\frac{1}{100}$ (probit)

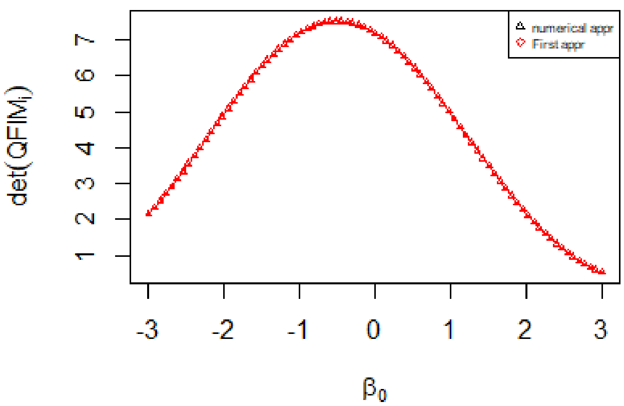

Figure 4.5: Quasi Fisher information matrix determinant (probit)

4.6 shows $\det(\boldsymbol{V}_j)$. Since $x_1 = 0$, $\det(\boldsymbol{V}_j)$ is constant against β_1; however, the component in plot (b) shows a constant curve over β_1. Figure 4.7 illustrates $\det(\boldsymbol{C}_j); j = 1, 2;\ k \neq k'$ against $\beta_1 \in [-3, 3]$. It is also constant at 6×10^{-4}, due to the similar reason for obtaining $\det(\boldsymbol{V}_j)$ in Figure 4.6. However, the considered component with $j = 2$ takes its maximum value at around $\beta_1 = -1$.

Furthermore, we observe in Figure 4.8 that $\det(\boldsymbol{C}_{12}) = \det(\boldsymbol{C}_{21})$. It takes its maximum value at around the middle of the interval of β_1. At last, the determinant of matrix $\mathbf{V}$ in power of $\frac{1}{100}$ and $\mathbf{M}^Q(\xi, \boldsymbol{\beta})$ are computed and they are plotted respectively in Figures 4.9 and 4.10.

In Figure 4.9, $(\det(\mathbf{V}))^{\frac{1}{100}}$ takes its maximum value at around $\beta_1 = -1$, and it drops to values close to zero fast as β_1 deviates from zero. Figure 4.10 indicates the values of $\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta}))$ against β_1. It takes its highest value at negative β_1 close to zero at around the value 7. As β_1 tends to 3, $\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta}))$ tends to zero.

In the next subsection, we obtain the approximation of the quasi Fisher information matrix in a special case of ordered mixed logit model.

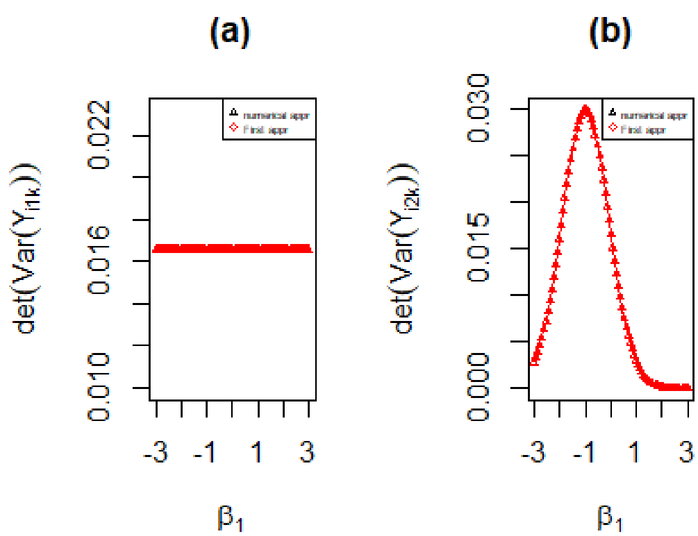

Figure 4.6: Variance matrix determinant of $\boldsymbol{Y}_{ijk}$ (probit) (a): $j = 1$, (b): $j = 2$

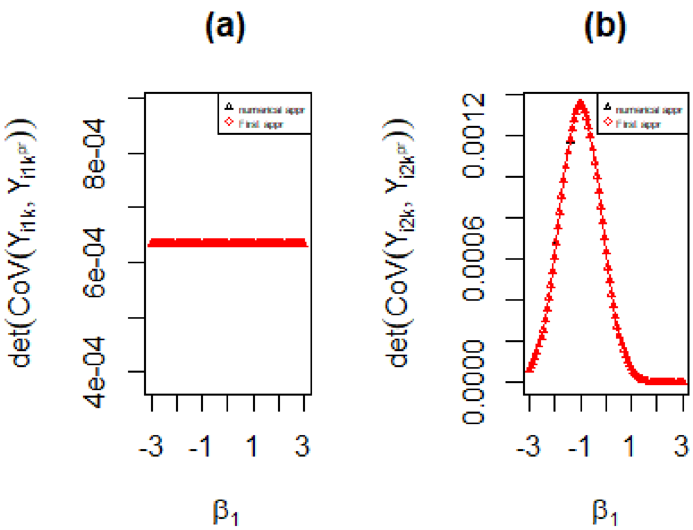

Figure 4.7: Covariance matrix determinant of $\boldsymbol{Y}_{ijk}$ and $\boldsymbol{Y}_{ijk'}$ (probit) (a): $j = 1$, (b): $j = 2$

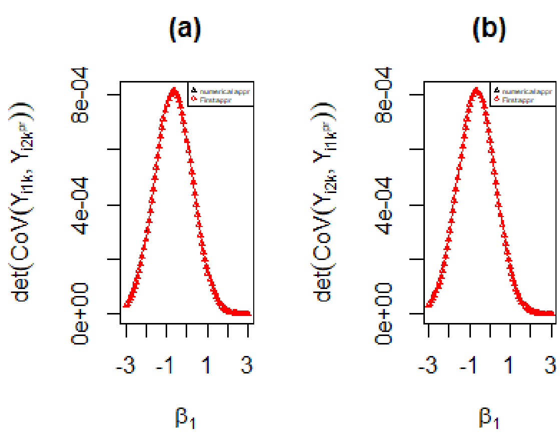

Figure 4.8: Covariance matrix determinant of $\boldsymbol{Y}_{ijk}$ and $\boldsymbol{Y}_{ij'k'}$, $j \neq j'$ (a): $j = 1, j' = 2$, (b): $j = 2, j' = 1$

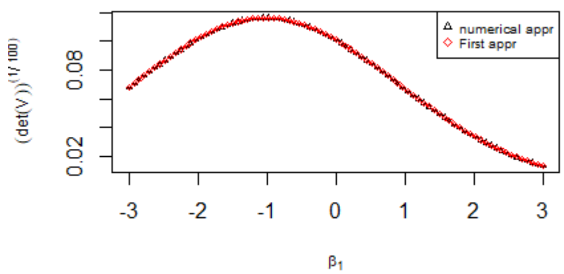

Figure 4.9: Determinant of variance matrix $\mathbf{V}$ in power of $(\frac{1}{100})$ in dependence on β_1 (probit)

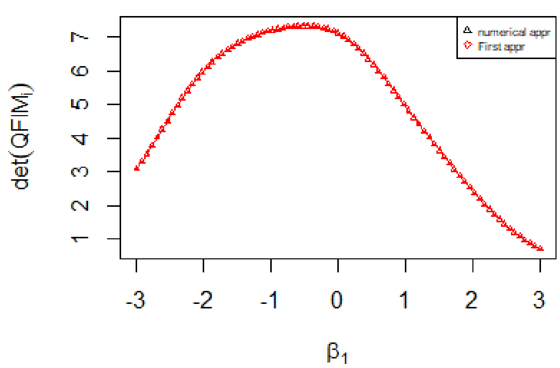

Figure 4.10: Quasi Fisher information matrix determinant in dependence on β_1 (probit)

Quasi Fisher information matrix in model with logit function

In this subsection, the approximation of the quasi Fisher information matrix for the random intercept ordinal regression model with the logit link function is established. The general procedure is related to the corresponding method in the random intercept binary model with the logit link function (See Section 3.1). Regarding the formulation of $\pi_j^{(m)}$ in (2.22),

$$\pi_j^{(m)} = \mathrm{E}[p_{ij}^{(m)}(\boldsymbol{\zeta}_i)] - \mathrm{E}[p_{ij}^{(m-1)}(\boldsymbol{\zeta}_i)]; \tag{4.18}$$

then,

$$\begin{aligned}\mathrm{E}[p_{ij}^{(m)}(\boldsymbol{\zeta}_i)] &\approx \int_{\mathbb{R}^q} \frac{\exp(\gamma_m - (\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} + \boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i}))}{1+\exp(\gamma_m - (\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} + \boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i}))} \phi_{\mathbf{0},\sigma^2\mathbf{I}_q}(\boldsymbol{\zeta}_i)\mathrm{d}\boldsymbol{\zeta}_i \\ &\approx \textstyle\int_{\mathbb{R}^q}\int_{-\infty}^{\gamma_{(m)}-(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i})} \phi_{0,\frac{1}{c^2}}(u)du\phi_{\mathbf{0},\sigma^2\mathbf{I}_q}(\boldsymbol{\zeta}_i)d\boldsymbol{\zeta}_i \\ &\approx \Phi(\alpha_{j(c)}).\end{aligned} \tag{4.19}$$

where $\alpha_{j(c)}^{(m)} = (1 + c^2\sigma^2\boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{h}(\boldsymbol{x}_j))^{-\frac{1}{2}}(\gamma_m - \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta})$, in which $c = \frac{16\sqrt{3}}{15\pi}$. The first approximate equality is gained from the approximation of the logistic cumulative distribution function $\frac{\exp(\gamma_m-(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i}))}{1+\exp(\gamma_m-(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}+\boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i}))}$ with the Normal cumulative distribution function, $\Phi_{0,\frac{1}{c^2}}(\gamma_m - (\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} + \boldsymbol{h}^\top(\boldsymbol{x}_j)\boldsymbol{\zeta_i}))$. The third approximate equality is gained from the direct calculation of the corresponding integral, in the second approximate equality (Zeger, Liang, and Albert 1988). Subsequently, for the computation of $\boldsymbol{V}_j$, we require to compute (4.6) and (4.7). They are a function of $\pi_j^{(m)}$; $\forall m$. Therefore, they are easily computed by (4.19). Additionally, in view of the approximation of $C_j^{(mm')}; k \neq k'$ and $C_{jj'}^{mm'}; j \neq j', \forall k, k'$, we apply the procedure in lemma 4.1.2 to obtain the second order moment of the response variable. The only difference is that σ^2 is transformed to $c^2\sigma^2$. Nevertheless, the relevant results are shown in Appendix A, Section A.6.

Also, to find the components of matrix $\mathbf{D}$, it is only needed to transform σ^2 to $c^2\sigma^2$, similar to lemma 4.1.3 (See Appendix A, Section A.6).

4.2 D-optimum design

In this section, we attempt to obtain the two point D-optimum design in the special case of the ordinal mixed effects model. In the following example, the determinant

Table 4.1: D-optimum weight, random intercept ordered probit model

$n \mid \beta_0$	-3	-2	-1	0	1	2	3
50	0.54	0.52	0.50	0.50	0.48	0.46	0.44
100	0.49	0.49	0.49	0.49	0.49	0.47	0.45

of the quasi Fisher information matrix against the w_1 for different values of model parameters is obtained.

Example.
In this example, a specific case of the model with $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, $(x_1, x_2) = (0, 1)$, $M = 3$, $\sigma^2 = 1$ and $\gamma_0 = -2000, \gamma_1 = -1, \gamma_2 = 1, \gamma_3 = 2000$ is considered. The D-optimum weight $w_1^\star$, when $\beta_0 \in \{-3, -2, -1, 0, 1, 2, 3\}$, $\beta_1 = 1$ and $n = 50, 100$ is calculated. The D-optimum weight $w_1^\star$ is obtained based on the number of replication at $x_1 = 0$. The result is shown in Table 4.1. It illustrates that as β_0 increases from -3 to 3, the D-optimum weight $w_1^\star$ decreases, which means that at $x_1 = 0$ less replication is required to take in comparison to $x_2 = 1$. Figure 4.1 shows the determinant of the quasi Fisher information matrix against weight w_1 for $n = 50$. This reveals that approximately in the middle of the interval $w_1 \in [0, 1]$, the determinant of the quasi Fisher information matrix is maximized. The individual two point D-optimum design can be extended to the population D-optimum design following Schmelter 2007, Theorem 1. The achievement of the D-optimum weight $w_1^\star$ is directly conducted by the computation of the determinant of the quasi Fisher information matrix at different values of w_1 and then choosing a weight which maximizes the corresponding determinant.

In the situation where $\beta_1 \in \{-3, -2, -1, 0, 1, 2, 3\}$, $\beta_0 = 1$ and $n = 50, 100$ the D-optimum weight $w_1^\star$ is obtained. The results are indicated in Table 4.2. Moreover, as n increases from 50 to 100, the D-optimum weight $w_1^\star$ remains constant or decreases so slightly. When n becomes large, it is time consuming to compute the D-optimum design. More calculation regarding the achievement of D-optimum design is postponed to the further research. In Figure 4.12 shows the determinant of the quasi Fisher information matrix for each $w_1 \in [0, 1]$ for different values of $\beta_1 \in \{-3, -2, -1, 0, 1, 2, 3\}$. Like in the previous plot, the maximum value of the determinant of the quasi Fisher information matrix occurs approximately in the middle of the interval of w_1. Overall, the computation can be complex and in our work, we hope to simplify our approach.

Finally, further research regarding the achievement of D-optimum design for the aim of creating the simplicity in the method is required. Accordingly the quasi Fisher information matrix could be viewed in more transparent way and more transparent ideas and theories considering the D-optimum design can be

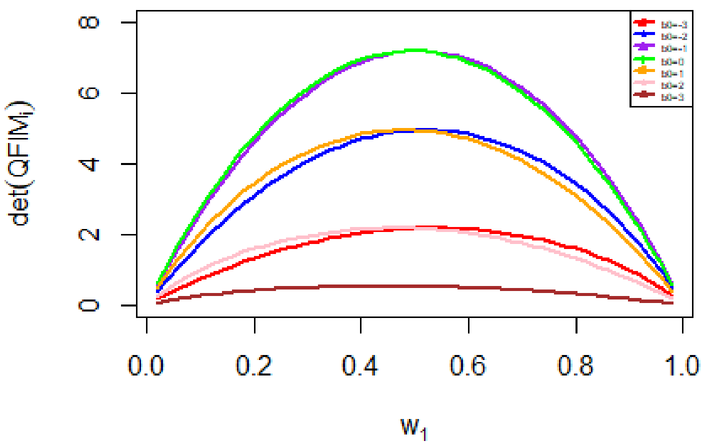

Figure 4.11: Determinant of the quasi Fisher information matrix in dependence on w_1 for different β_0, (probit link)

Table 4.2: D-optimum weight, $w_1^\star$, random intercept ordered probit model

n | β_1	-3	-2	-1	0	1	2	3
50	0.48	0.50	0.50	0.50	0.48	0.46	0.42
100	0.48	0.49	0.49	0.50	0.49	0.48	0.42

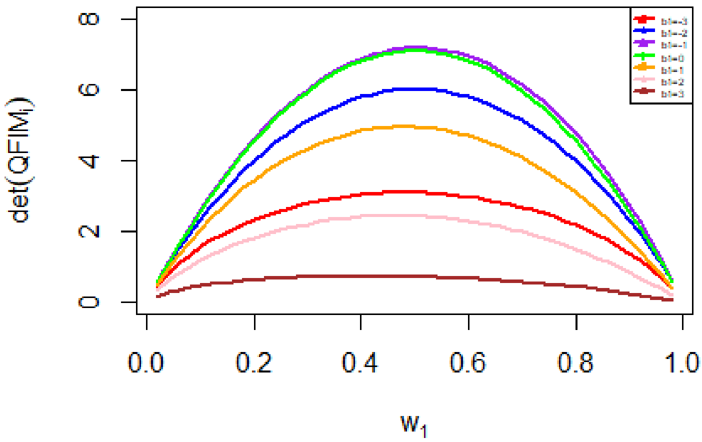

Figure 4.12: Determinant of the quasi Fisher information matrix in dependence on w_1 for different β_1, (probit link)

reformed. This perspective can be done in further research and investigation.

Chapter 5

Optimum Design in a Nonlinear Longitudinal Poisson Regression Model

The aim of this chapter is to obtain D-optimum design for the estimation of the mean ability parameter in a nonlinear longitudinal Poisson regression model. The reason for using the considered model has already been stated in Section 2.1. In this chapter, we draw the conclusion that due to the complex form of the marginal likelihood function, the maximum likelihood estimate of the model parameters are not explicitly achievable, and the Fisher information matrix is rather impossible to obtain and then we need to approximate the Fisher information matrix by the quasi Fisher information matrix (Wedderburn 1974). Additionally, the properties of this model and the relevant quasi Fisher information matrix are seeking to be obtained. Then, based on these properties, the equivalence theorem is constructed.

Finally, the D-optimum design for the quasi maximum likelihood estimate of the model parameters is calculated. Since the individuals take the same design, the D-optimum design is obtained for individual and it is extended to the population D-optimum design. This design depends on the unknown parameters, we assume some initial values for them. Therefore, the resulting D-optimum design is called a local D-optimum design. Moreover, the D-efficiency of the D-optimum design is computed. Since this criterion is close to one at the end, we conclude that it is not needed to take more sample to gain the same D-efficiency, if the true population parameter is not chosen.

5.1 Properties of the nonlinear longitudinal Poisson regression model

In this section, the nonlinear longitudinal Poisson regression model is represented briefly. The response variable Y_{ijk} conditioned on the random ability parameter $\Lambda_{ij} = \lambda_{ij}$ follows the Poisson distribution. The random ability parameter Λ_{ij} follows the multivariate Gamma distribution. The conditional mean response is dependent upon the random ability λ_{ij} and the exponential function of the mean ability parameter $\theta(\boldsymbol{t}, \boldsymbol{\beta})$, where t_j is the jth time point and $\boldsymbol{\beta}$ is the population parameter.

In this section, the properties of the model in order to form the quasi Fisher information matrix are obtained. The marginal response mean $\mu_j = \mathrm{E}(Y_{ijk})$, where $\mathrm{E}(Y_{ijk}) = \exp(\theta(t_j, \boldsymbol{\beta}))$ and marginal variance $\mathrm{Var}(Y_{ijk}) = \mu_j(\mu_j\tau + 1)$ are already obtained in formulas (2.34) and (2.35). Additionally, the marginal covariance elements of the response variable are gained as follows:

$$\begin{aligned}\mathrm{cov}(Y_{ijk}, Y_{ijk'}) &= \mathrm{cov}(\mathrm{E}(Y_{ijk} \mid \Lambda_{ij}), \mathrm{E}(Y_{ijk'} \mid \Lambda_{ij})) + \mathrm{E}(\mathrm{cov}(Y_{ijk}, Y_{ijk'} \mid \Lambda_{ij})) \\ &= \mu_j^2 \tau; k \neq k', \end{aligned} \tag{5.1}$$

since the second term is zero and the first term is

$$\begin{aligned}\mathrm{cov}(\mathrm{E}(Y_{ijk} \mid \Lambda_{ij}), \mathrm{E}(Y_{ijk'} \mid \Lambda_{ij})) &= \mathrm{cov}(\Lambda_{ij}\exp(\theta(t_j, \boldsymbol{\beta})), \Lambda_{ij}\exp(\theta(t_j, \boldsymbol{\beta}))) \\ &= \mu_j^2 \mathrm{var}(\Lambda_{ij}).\end{aligned}$$

For different experimental settings, t_j and $t_{j'}$,

$$\begin{aligned}\mathrm{cov}(Y_{ijk}, Y_{ij'k'}) &= \mathrm{cov}(\mathrm{E}(Y_{ijk} \mid \Lambda_{ij}), \mathrm{E}(Y_{ij'k'} \mid \Lambda_{ij'})) + \mathrm{E}(\mathrm{cov}(Y_{ijk}, Y_{ij'k'} \mid \Lambda_{ij}, \Lambda_{ij'})), \\ &= \rho\mu_j\mu_{j'}\tau; j \neq j', k, k', \end{aligned} \tag{5.2}$$

since the second term is equal to zero and the first term is written as:

$$\begin{aligned}\mathrm{cov}(\mathrm{E}(Y_{ijk} \mid \Lambda_{ij}), \mathrm{E}(Y_{ij'k'} \mid \Lambda_{ij'})) &= \mathrm{cov}(\Lambda_{ij}\exp(\theta(t_j, \boldsymbol{\beta})), \Lambda_{ij'}\exp(\theta(t_{j'}, \boldsymbol{\beta}))) \\ &= \mu_j\mu_{j'}\mathrm{cov}(\Lambda_{ij}, \Lambda_{ij'}),\end{aligned}$$

and according to (2.32) $\mathrm{cov}(\Lambda_{ij}, \Lambda_{ij'}) = \rho\tau$.

Let $\mathbf{1}_m$ be the vector of length m with all entries equal to 1. For the construction of covariance matrix for each subject, consider $\mathbf{V}_{jj} = \mathbf{C}_j + \rho\tau\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top$, with dimension $n_j \times n_j$ and $\mathbf{V}_{jj'} = \rho\tau\boldsymbol{\mu}_j\boldsymbol{\mu}_{j'}^\top$ with dimension $n_j \times n_{j'}$, where $\mathbf{C}_j = \boldsymbol{\mu}_j\mathbf{I}_{n_j} + (1-\rho)\tau\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top$, such that $\boldsymbol{\mu}_j = \mu_j\mathbf{1}_{n_j}$. Moreover,

$$\mathbf{C} = \text{diag}(\mathbf{C}_j); j = 1, ..., J,$$

then,

$$\mathbf{V} = \mathbf{C} + \rho\tau\boldsymbol{\mu}\boldsymbol{\mu}^\top, \tag{5.3}$$

with dimension $n \times n$. diag(.) represents the diagonal matrix in general. $\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1}^\top, ..., \boldsymbol{Y}_{iJ}^\top)^\top$, $\boldsymbol{Y} = (\boldsymbol{Y}_1^\top, ..., \boldsymbol{Y}_N^\top)^\top$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, ..., \boldsymbol{\mu}_J^\top)^\top$. $\text{E}(\boldsymbol{Y}_i) = \boldsymbol{\mu}$ and $\text{E}(\boldsymbol{Y}) = 1_N \otimes \boldsymbol{\mu}$ and since subjects are presumed to be independent for all observations, $\text{Cov}(\boldsymbol{Y}) = \mathbf{I}_N \otimes \mathbf{V}$.

Quasi Fisher information matrix

In this section, we construct the quasi-Fisher information matrix for model (2.33).

In order to build the general form of the individual quasi Fisher information matrix, the subsequent lemmas and theorem are stated.

Let $\psi_j(\xi) = \frac{n_j\mu_j}{1+(1-\rho)\tau n_j\mu_j}$ as weighting factor and

$$\boldsymbol{d}_j = (1, -\exp(-\beta_2 t_j), \beta_1 t_j \exp(-\beta_2 t_j))^\top \exp(\theta(t_j, \boldsymbol{\beta}))$$

be the gradient of μ_j with respect to $\boldsymbol{\beta}$ with dimension 3×1. Then, $\mathbf{D} = (\mathbf{D}_1^\top, ..., \mathbf{D}_J^\top)^\top$, where $\mathbf{D}_j^\top = \boldsymbol{d}_j \mathbf{1}_{n_j}^\top$.

Further, consider $\boldsymbol{d}_{0j} = \frac{1}{\mu_j}\boldsymbol{d}_j$, as the scale gradient,

$$\boldsymbol{d}_{0j} = (1, -\exp(-\beta_2 t_j), \beta_1 t_j \exp(-\beta_2 t_j))^\top,$$

Let $\mathbf{D}_0 = (\boldsymbol{d}_{01}^\top, ..., \boldsymbol{d}_{0J}^\top)^\top$ be the $J \times p$ essential scaled design matrix.

Lemma 5.1.1. *The inverse of the individual variance covariance matrix (5.3) is obtained as:*

$$\mathbf{V}^{-1} = \mathbf{C}^{-1} - \frac{\rho\tau}{1 + \rho\tau \sum_{j=1}^J \psi_j} \mathbf{C}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{C}^{-1}. \tag{5.4}$$

Lemma 5.1.2. *The individual quasi Fisher information matrix based on (2.69) for model (2.33) is written as follows:*

$$\mathbf{M}^Q(\xi) = (\sum_{j=1}^J \psi_j \boldsymbol{d}_{0j}\boldsymbol{d}_{0j}^\top) - \frac{\rho\tau}{1 + \rho\tau \sum_{j=1}^J \psi_j(\xi)} (\sum_{j=1}^J \psi_j(\xi)\boldsymbol{d}_{0j})(\sum_{j=1}^J \psi_j(\xi)\boldsymbol{d}_{0j}^\top) \tag{5.5}$$

Proof.

The proof is stated in Appendix A, Section A.9 ■

Lemma 5.1.2 is still valid when some of the n_j's are equal to zero. To process the following Lemmas and Theorem, let $\mathbf{e}_1$ be the first p-dimensional unit vector; such that $\mathbf{D}_0\mathbf{e}_1 = \mathbf{1}_J$. Also, $\boldsymbol{\Psi}(\xi) = \text{diag}(\psi_1(\xi), ..., \psi_J(\xi))$ and $\mathbf{M}(\xi) = \mathbf{D}_0^\top \boldsymbol{\Psi}(\xi)\mathbf{D}_0$.

Remark to Lemma 5.1.2.

This representation of the quasi information matrix can be used for generalization to approximate designs. With the notation introduced before, it can be written as

$$\mathbf{M}^Q(\xi) = \mathbf{D}_0^\top \boldsymbol{\Psi}(\xi)\mathbf{D}_0 - \frac{\rho\tau}{1+\rho\tau\mathbf{1}_J^\top\boldsymbol{\Psi}(\xi)\mathbf{1}_J}\mathbf{D}_0^\top\boldsymbol{\Psi}(\xi)\mathbf{1}_J\mathbf{1}_J^\top\boldsymbol{\Psi}(\xi)\mathbf{D}_0, \tag{5.6}$$

where only the weighting matrix $\boldsymbol{\Psi}$ depends on the design and the restriction to integer n_j can be relaxed.

Theorem 5.1.3. *In longitudinal Poisson regression model (2.33), the individual quasi Fisher information matrix is obtained as follows:*

$$\mathbf{M}^Q(\xi) = (\mathbf{M}(\xi)^{-1} + \rho\tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1} \tag{5.7}$$

Proof.

The proof is stated in Appendix A, Section A.9. ■

In case where there is no random effect in model (2.33), the quasi Fisher information matrix can be constructed in the following lemma.

Lemma 5.1.4. *Let $\boldsymbol{\mathcal{E}}(\xi) = \mathrm{diag}(\mu_1 n_1, ..., \mu_J n_J)$, the quasi Fisher information matrix for fixed effect Poisson model is:*

$$\mathbf{M}^Q(\xi) = \mathbf{D}_0^\top\boldsymbol{\mathcal{E}}(\xi)\mathbf{D}_0 \tag{5.8}$$

Proof.

The proof is stated in Appendix A, Section A.9. ■

The quasi Fisher information matrix in the fixed effects Poisson model, can be obtained in the corresponding mixed effects model, when τ tends to zero in (5.6).

Lemma 5.1.5. *Let $W(\xi) = \mathrm{diag}(1/(1+(1-\rho)\tau\mu_j n_j(\xi)))_{j=1,...,J}$. The D-optimality criterion (2.75) for model (2.33) based on quasi Fisher information matrix (5.7) is differentiable.*

Proof.

The proof is stated in Appendix A, Section A.9 ■

5.2 D-optimum design

The individual D-optimality criterion is revealed based on the individual quasi-Fisher information matrix, as

$$\mathfrak{D}^Q(\xi,\boldsymbol{\beta}) = \log(\det(\mathbf{M}^Q(\xi,\boldsymbol{\beta}))).$$

Now, we seek for some properties of the D-optimality criterion.

Lemma 5.2.1. ***Monotonicity property***. *If ξ_1 dominates ξ_2, i.e. $\xi_1 \succ \xi_2$ if $\mathbf{M}^Q(\xi_1) \geq \mathbf{M}^Q(\xi_2)$; then, the following inequality holds for the individual D-optimality criterion:*

$$\mathfrak{D}^Q(\xi_1, \boldsymbol{\beta}) \geq \mathfrak{D}^Q(\xi_2, \boldsymbol{\beta}).$$

Lemma 5.2.2. ***Concavity property***. *The quasi Fisher information matrix $\mathbf{M}^Q(\xi)$ is concave, i.e. $\mathbf{M}^Q((1-\epsilon)\xi_1+\epsilon\xi_2) \geq (1-\epsilon)\mathbf{M}^Q(\xi_1)+\epsilon\mathbf{M}^Q(\xi_2)$ for every $\epsilon \in (0,1)$ and every ξ_1, ξ_2.*

Proof.
The proof is stated in Appendix A, Section A.9 ■

In the following, we state the equivalence theorem for the corresponding D-optimality criterion for model (2.33).

Theorem 5.2.3. *(Equivalence Theorem)*

Let $\mathbf{D}_0\mathbf{e}_1 = \mathbf{1}_J$ and denote by $\mathbf{W}(\xi) = \operatorname{diag}\left(\frac{1}{1+(1-\rho)\tau n_j(\xi)\mu_j}\right)_{j=1,\ldots,J}$. Then ξ^* is D-optimum if and only if

$$\frac{n\mu_j}{(1+(1-\rho)\tau n_j(\xi^*)\mu_j)^2}\boldsymbol{d}_{0j}^\top \mathbf{M}^{-1}(\xi^*)\mathbf{M}^Q(\xi^*)\mathbf{M}^{-1}(\xi^*)\boldsymbol{d}_{0j}$$

$$\leq \operatorname{tr}\left(\mathbf{W}(\xi^*)\boldsymbol{\Psi}(\xi^*)\mathbf{D}_0\mathbf{M}^{-1}(\xi^*)\mathbf{M}^Q(\xi^*)\mathbf{M}^{-1}(\xi^*)\mathbf{D}_0^\top\right) \tag{5.9}$$

for all $j = 1, \ldots J$.

Proof.
The proof is stated in Appendix A, Section A.9. ■

Corollary 5.2.3.1. *The equality in (5.9) holds, at any j for which $n_j > 0$, i.e.*

$$F_{\mathfrak{D}^Q}(\xi^*, \xi_j) = 0, \tag{5.10}$$

where ξ_j is the one-point design which assigns all n observations to time point t_j.

We can obtain the D-optimality criterion in the subsequent lemma for the purpose of optimization with statistical software R (R Core Team 2021) with respect to design. Using the D-optimality criterion in the subsequent Lemma may ease the computation in the satatistical software R.

Lemma 5.2.4. *The D-optimality criterion as a function of quasi Fisher information matrix (5.7) for model (2.33) is as follows:*

$$\mathfrak{D}^Q(\xi) = \log\left(\frac{\det(\mathbf{M}(\xi))}{1+\rho\tau\mathbf{1}_J^\top\boldsymbol{\Psi}(\xi)\mathbf{1}_J}\right) \tag{5.11}$$

Table 5.1: Locally D-optimum designs for $J = 3$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $\rho = 0.9, \tau = 1$, $n = 120$

β_2	1.0	1.5	2.0	2.5	3.0	3.5
w_1	0.511	0.541	0.556	0.564	0.569	0.571
w_2	0.272	0.249	0.235	0.226	0.221	0.217
w_3	0.216	0.210	0.209	0.210	0.211	0.211

Proof.

The proof is stated in Appendix A, Section A.9. ■

In this section, we aim to obtain D-optimum design in special cases of model (2.33). In fact, as $\xi^\star$ may change in terms of parameter $\boldsymbol{\beta}$, some specific values of $\boldsymbol{\beta}$ are considered as:

$\beta_0 = 3, 6, \beta_1 = 2, \beta_2 = \{1, 1.5, 2, 2.5, 3, 3.5\}$

The weight $w_j = \frac{n_j}{n}$ is allocated to the j^{th} time point. Furthermore, the D-optimum weight $w_j^\star$ depends on the total number n. Actually, the weight distribution of the D-optimum design for the model is asserted. We seek to find the D-optimum design in terms of maximizing D-optimality criterion $\mathfrak{D}^Q(\xi)$ (5.11) with respect to w_j; $\sum_{j=1}^{J} w_j = 1$; $w_j > 0$ with fixed $t_j = j - 1$.

For this aim we use the statistical software R (R Core Team 2021) with the class of indirect solvers and implement the augmented Lagrange multiplier method with an SQP interior algorithm (Ye 1987). This method is used in the package Rsolnp. For the aim of the computation, the total number of items is needed to be specified. It is here fixed with $n = 120$. Further, $J \geq p$ is required for estimability of the ability parameter. At the end, the true choice of the weights determines n_j in the rest. Moreover, the assumed values of hyper parameters are set to be $\rho = 0.9$, since $\rho = c_0\tau$; $c_0 = 0.9, c = 0.1$ and $\tau = 1$.

The locally D-optimum designs for $\beta_0 = 3, 6, \beta_1 = 2$ are shown in Tables 5.1 and 5.2. For all β_2 ranging from 1 to 3.5, it shows a decreasing trend from w_1 to w_3, which means for each β_2 we need smaller number of replications at higher indices of time points. Also, in higher values of β_2, the more value of w_1 is required, in other words, as β_2 rises up w_1 is also increasing. All in all, these tables reveal an ascending trend through weights when β_2 changes for fixed β_0 and β_1. Also, Figure 5.1 reveals the corresponding consequences. It illustrates that w_3 interacts slightly with w_2 for $\beta_2 = 2.5, 3.0, 3.5$.

Tables A.1 and A.2 (Appendix A, Section A.11) show the D-optimum design with $J = 4$, with the same parametrization as the above situation. In both tables w_1 takes the highest values. w_2 takes the second highest value among w_j's, $j = 1, ..., 4$. Then, w_4 and w_3, respectively. The values in both tables are quite

Table 5.2: Locally D-optimum designs for $J = 3$ periods, $\beta_0 = 6, \beta_1 = 2$, $\rho = 0.9, \tau = 1$, $n = 120$

β_2	1.0	1.5	2.0	2.5	3.0	3.5
w_1	0.512	0.542	0.557	0.565	0.570	0.572
w_2	0.272	0.249	0.235	0.226	0.220	0.217
w_3	0.216	0.209	0.209	0.209	0.210	0.211

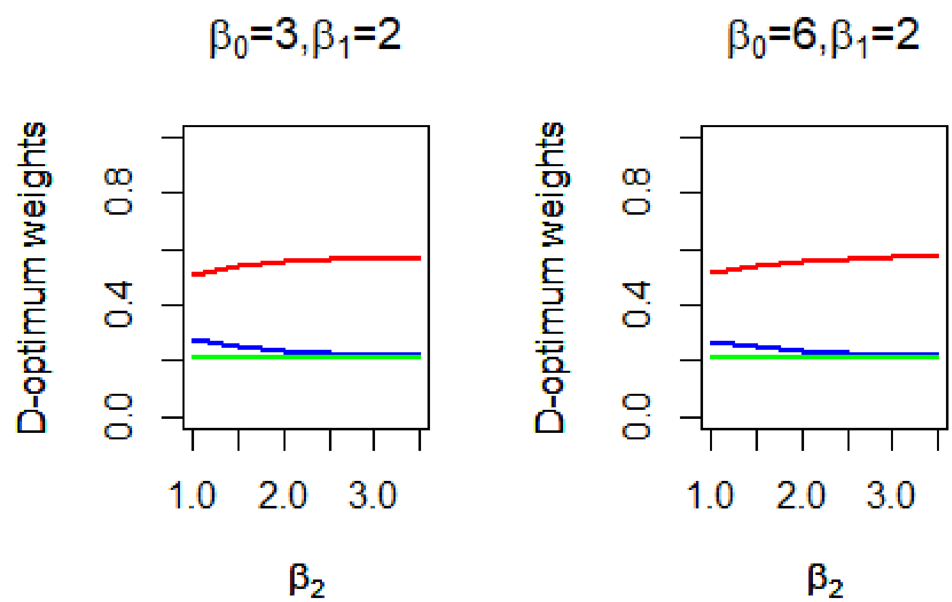

Figure 5.1: Local D-optimum designs, $J = 3$ periods, $\rho = 0.9$, $\tau = 1$, red line: $w_1^\star$, blue line: $w_2^\star$, green line: $w_3^\star$

analogous, which means that the D-optimum design is not sensitive with respect to increasing β_0 from 3 to 6. Figure A.1 illustrates the corresponding changes and trends.

Tables A.3 and A.4 (Appendix A, Section A.11) show the D-optimum design for $J = 7$ for the same model parametrization. w_1 in both situations for $\beta_0 = 3, 6$ takes the highest value and it rises up as β_2 increases, and w_2 takes the second highest value. The rest of the weights take on a ration of less than 0.1. The weight values are quite the same between two tables similar to the four point D-optimum design. Figure A.2 illustrates the corresponding consequences.

We then use (2.86) and compute the D-efficiencies of $\xi^{\star}_{\boldsymbol{\beta}}$ for various values of $\boldsymbol{\beta}^0$. This means that after we provide the true population parameter $\boldsymbol{\beta}^0$, we can assess how $\xi^{\star}_{\boldsymbol{\beta}}$ deviates from $\xi^{\star}_{\boldsymbol{\beta}_0}$ in terms of the determinant of the quasi Fisher information matrix. What is shown is that when assuming $\boldsymbol{\beta}^0 = (\beta_0, \beta_1, \beta_{2m'})$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_{2m})$, as β_0 and β_1, takes values from the sets $\{3, 6\}$, $\{2\}$ respectively and $\beta_{2m}, \beta_{2m'}$ take value from $\{1.0, 1.5, 2.0, 2.5, 3.0, 3.5\}$; $m \neq m' = 1, ..., 6$, then the D-efficiency criterion obtains values more than 0.98, which means as the true population parameter is not the value we are considering and obtain the D-optimum design based on the assumed $\boldsymbol{\beta}^0$, then we will be able to neglect taking more sample based on $\xi^{\star}_{\boldsymbol{\beta}}$. According to this, there does not exist substantial sensitivity of the D-optimum design with respect to changes of parameter $\boldsymbol{\beta}$.

5.2.1 Special cases

Considering model (2.33), and the corresponding intraclass correlation $\rho = \frac{c_0}{c_0+c}$, we consider special cases where:

1) $c_0 = 0$, then $\rho = 0$ and $\psi_{j0} = \frac{n_j\mu_j}{1+\tau n_j\mu_j}$, which means that there is no correlation of the response variable among time points.

2) $c = 0$, then $\rho = 1$ and $\psi_{j1} = n_j\mu_j$, which means that there is the block effect consistent regarding each subject.

n is the total number of observations in the ith individual. μ_j is the marginal response mean for the jth time point, n_j is the number of replication at the jth time point. c and c_0 are shape parameters for the Gamma distribution of the random ability parameters. Each situation is probed in two different subsections and the results regarding the achievement of D-optimum design and the equivalence theorem are obtained.

1. Special case $\rho = 0$

Corollary 5.2.4.1. *The quasi-Fisher information matrix for model (2.33) is written as:*

$$\mathbf{M}^Q(\xi) = \mathbf{M}(\xi) \tag{5.12}$$

where $\mathbf{M}(\xi) = \mathbf{D}_0^\top \mathbf{\Psi}_0(\xi)\mathbf{D}_0$ *with* $\mathbf{\Psi}_0(\xi) = \mathrm{diag}(\psi_{10}, ..., \psi_{J0})$.

Corollary 5.2.4.2. *(Equivalence Theorem)*

Let $\mathbf{D}_0\mathbf{e}_1 = \mathbf{1}_J$ *and denote by* $\mathbf{W}(\xi) = \mathrm{diag}\left(\frac{1}{1+\tau n_j(\xi)\mu_j}\right)_{j=1,\dots,J}$*; Then,* $\xi^\star$ *is D-optimum if and only if*

$$\frac{n\mu_j}{(1+\tau n_j(\xi^*)\mu_j)^2}\boldsymbol{d}_{0j}^\top \mathbf{M}^{-1}(\xi^*)\mathbf{M}^Q(\xi^*)\mathbf{M}^{-1}(\xi^*)\boldsymbol{d}_{0j} \tag{5.13}$$

$$\leq \mathrm{tr}\left(\mathbf{W}(\xi^*)\mathbf{\Psi}_0(\xi^*)\mathbf{D}_0\mathbf{M}^{-1}(\xi^*)\mathbf{M}^Q(\xi^*)\mathbf{M}^{-1}(\xi^*)\mathbf{D}_0^\top\right) \tag{5.14}$$

for all $j = 1, \dots J$.

Corollary 5.2.4.3. *The D-optimality criterion as a function of quasi Fisher information matrix (5.7) for model (2.33) is illustrated as follows:*

$$\mathfrak{D}_0^Q(\xi, \boldsymbol{\beta}) = \log(\det(\mathbf{M}(\xi))). \tag{5.15}$$

2. Special case $\rho = 1$

Corollary 5.2.4.4. *The quasi-Fisher information matrix for model (2.33) is written as:*

$$\mathbf{M}^Q(\xi) = (\mathbf{M}^{-1}(\xi) + \tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1} \tag{5.16}$$

where $\mathbf{M}(\xi) = \mathbf{D}_0^\top \mathbf{\Psi}_1(\xi)\mathbf{D}_0$ *with* $\mathbf{\Psi}_1(\xi) = \mathrm{diag}(\psi_{11}, ..., \psi_{J1})$.

Corollary 5.2.4.5. *(Equivalence Theorem)*

Let $\mathbf{D}_0\mathbf{e}_1 = \mathbf{1}_J$*. Then,* $\xi^\star$ *is D-optimum if and only if*

$$n\mu_j\boldsymbol{d}_{0j}^\top \mathbf{M}^{-1}(\xi^*)\mathbf{M}^Q(\xi^*)\mathbf{M}^{-1}(\xi^*)\boldsymbol{d}_{0j} \tag{5.17}$$

$$\leq \mathrm{tr}\left(\mathbf{\Psi}_1(\xi^*)\mathbf{D}_0\mathbf{M}^{-1}(\xi^*)\mathbf{M}^Q(\xi^*)\mathbf{M}^{-1}(\xi^*)\mathbf{D}_0^\top\right) \tag{5.18}$$

for all $j = 1, \dots J$.

Remark. In above Corollary $\mathbf{W}(\xi) = \mathbf{I}_J$.

Corollary 5.2.4.6. *The D-optimality criterion as a function of quasi Fisher information matrix (5.7) for model (2.33) is:*

$$\mathfrak{D}_1^Q(\xi, \boldsymbol{\beta}) = \log\left(\frac{\det(\mathbf{M}(\xi))}{1+\tau\mathbf{1}_J^\top\mathbf{\Psi}_1(\xi)\mathbf{1}_J}\right). \tag{5.19}$$

Table 5.3: Locally D-optimum designs for $J = 3$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $\rho = 0, \tau = 10$, $n = 120$

β_2	1.0	1.5	2.0	2.5	3.0	3.5
w_1	0.536	0.565	0.580	0.583	0.592	0.594
w_2	0.243	0.220	0.204	0.199	0.189	0.185
w_3	0.221	0.216	0.216	0.218	0.219	0.220

Table 5.4: Locally D-optimum designs for $J = 3$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $\rho = 1, \tau = 10$, $n = 120$

β_2	1.0	1.5	2.0	2.5	3.0	3.5
w_1	0.471	0.493	0.505	0.511	0.515	0.517
w_2	0.290	0.272	0.260	0.252	0.248	0.245
w_3	0.238	0.235	0.235	0.237	0.238	0.239

D-optimum design

In the case of $\rho = 0$, which means that there is no intraclass correlation among the random effects of each time point, the D-optimum designs are calculated. As it is shown in Table 5.3, the D-optimum design is obtained with $\beta_0 = 3$, $\beta_1 = 2$, $\beta_2 = \{1, ..., 3.5\}$. As β_2 changes from 1 to 3.5, w_1 increases from .536 to .594. For $\beta_2 = 1, 1.5$, w_2 takes the second highest value, and w_2 w_3 interact at $\beta_2 = 2$. Tables A.5 and A.6 (Appendix A, Section A.11) illustrate the four-point and seven-point D-optimum designs, in turn for the same case. There is the significant decrease from w_1 to w_2 for all considered parameter values. w_1 rises up as β_2 increases from 1 to 3.5 in both cases, and the second highest values are allocated to w_2 in all considered parameter values. For seven-point D-optimum design, w_3 to w_7 take less than 0.1. correspondingly, Figure A.3 illustrates the trends for the three scenarios.

For the case of block effect, i.e. $\rho = 1$, D-optimum designs are computed and they are indicated in Tables 5.4, A.7 and A.8 (Appendix A, Section A.11) for $J = 3, 4, 7$, respectively. In Table 5.4 w_j decreases as j increases for all β_2. In Table A.7 w_3 takes nearly zero value for all β_2's. In fact the four point D-optimum design is transformed to three point D-optimum design. In Table A.8, w_1 takes the highest value. Then, w_2 takes the second highest value. The other time points take the weights less or equal to 0.1, mostly. Figure A.4 shows the corresponding trend and changes. The D-efficiency criterion is calculated. Each time we assume that one specific value of β_2 is true and the D-efficiency of the D-optimum design is computed if the true value of β_2 is not correctly specified. Similar to the previous situation, the D-efficiency criterion take the values greater than or equal to 0.98. This means that it is not required to take more samples to obtain the same D-efficiency if the true parameter of β_2 is not chosen.

Chapter 6

Discussions and Conclusions

In this dissertation, three types of models are considered as:
1) The binary mixed effects regression model
2) The ordinal mixed effects regression model
3) The nonlinear longitudinal Poisson regression model,

which are used to model longitudinal data, when they follow Bernoulli, Multinomial and Poisson distribution in each subject, respectively. The binary and ordinal mixed effects models belong to the class of generalized linear mixed effects models. They have been specified. Their properties have been investigated. We drew a conclusion that the Fisher information matrix lacks the analytical closed form. As the likelihood function of these models do not have any closed form expression, no closed form of the Fisher information matrix existed. Therefore, the quasi Fisher information matrix was decided to be used for further analysis (Wedderburn 1974). Since this matrix is based on the marginal first and second order moments of the response variable, we need to obtain them. This function relies on the marginal first and second order moments of the response variable. These moments in most situations do not have explicit closed forms. Therefore, firstly special cases of the models, as the random intercept binary and ordinal mixed model and the binary model with two random effects were considered and then the considered moments were approximated. The new approximations were then evaluated by comparing them with numerical results.

The quasi Fisher information matrix which is based on the quasi log-likelihood function, is obtained from the new approximation of the first and second order moments of the response variable. Accordingly, the two point D-optimum design with treatment control support points for different values of the model parameters is calculated numerically and its sensitivity with respect to parameter changes is investigated.

The third type of model which is the Nonlinear longitudinal Poisson regression model was specified in Chapter 2. The properties of this model were discussed and the form of the quasi Fisher information matrix was built up in chapter 5. The quasi Fisher information matrix had the analytical closed form and its properties could be investigated analytically. The general equivalence theorem in order to evaluate the D-optimality of D-optimum design was built up and D-efficiency of D-optimum design was calculated. The D-efficiency of D-optimum design showed that as the model population parameter is not correctly specified, and we obtain the D-optimum design, it is not required to take more sample in order to gain the same D-efficiency as the population parameter was correctly specified. The likelihood for this model had an explicit form, even for small J as the number of time points. Hence we could not obtain the maximum likelihood estimate of the parameters explicitly. As a result the Fisher information matrix lacked the analytical explicit form. For this reason the quasi Fisher information matrix was chosen for further analysis.

Finally, the sensitivity of the D-optimum design is investigated by its calculation based on different values of the model parameters.

One question which arise here is that how the quasi maximum likelihood estimate of the models perform and which properties do they carry? Is it worth using the quasi Fisher information matrix? For this reason, some general methods of estimation as the maximum likelihood estimation, generalized least square estimation and the quasi maximum-likelihood estimation have been compared and their properties were investigated.

In principle, we conclude that the quasi maximum-likelihood estimate of parameters follow some well behaved asymptotic properties. However, it was not adequate due to the lack of sufficient analytical proof in the random effects regression model.

In this dissertation, we expect that the quasi maximum-likelihood estimate of the parameters, follow well behaved asymptotic properties. In Chapter 2, the asymptotic properties of these estimates in a special case of the binary mixed effects model was investigated by numerical computation, and it showed satisfying properties. For additional research, we can investigate the same properties numerically in some specific situations of ordinal mixed model and Nonlinear longitudinal Poisson model.

Moreover, regarding the achievement of the approximate quasi Fisher information matrix in the binary mixed effects model with two random effects, it is possible to compute D-optimum design and then to observe the changes in D-optimum design when it was gained in the random intercept binary regression model.

It is also suggested to form the approximate quasi Fisher information matrix in the ordinal mixed effects model with two random effects. Then, it is recommended to compute the respective D-optimum design, to observe whether there are significant changes in the form of D-optimum design in the random intercept ordered regression model.

Considering the binary and ordinal mixed models with the logit function, all the procedures to achieve the D-optimum design is the same as the corresponding models with the probit link function, except that the dispersion parameter σ^2 is transformed to $c^2\sigma^2$, where $c = \frac{16\sqrt{3}}{15\pi}$.

All in all, the other suggestion is to find A, D_A and D_s optimum designs in the models of consideration. These criteria are represented in Silvey 1980. These designs and their theory can be compared and their relations and properties can be investigated.

Appendix A

Appendix

A.1 Individual D-optimum design to population D-optimum design

Assumption 1. 1) $\mathfrak{D}^Q$ is a real-valued function on the whole set $\mathcal{M}$ of symmetric non-negative definite matrices of dimension p: $\mathfrak{D}^Q : \mathcal{M} \to (-\infty, +\infty]$.

2) $\mathfrak{D}^Q$ is monotone with respect to Loewner ordering on $\mathcal{M}$ in the sense that $\mathbf{M}^Q(\xi_1) > \mathbf{M}^Q(\xi_2)$; then, $\mathfrak{D}^Q(\mathbf{M}^Q(\xi_1)) \geq \mathfrak{D}^Q(\mathbf{M}^Q(\xi_2))$.

Theorem A.1.1. *(Schmelter 2007) Let*

$$\xi^\star = \begin{pmatrix} \xi_i^\star \\ 1 \end{pmatrix} \in \Xi_i$$

be a D-optimum design in the class of single group designs Ξ_i, where $\mathfrak{D}^Q$ is a criterion satisfying Assumption 1. Then, $\xi^\star$ is also D-optimum in the larger class of group designs $\Xi^{(p)}$.

A.2 Second approximation of covariance (binary mixed probit model)

For the achievement of the subsequent rules, consider the following terms:

Let

$u_{j0} = \sqrt{a_{j0}^2 + 2\alpha}$,

$a_{j0} = -\frac{(2\sigma^2+1)^{\frac{1}{2}}}{\sigma^2}$,

$b_{j0} = \frac{(\sigma^2+1)^{\frac{1}{2}}}{\sigma^2}\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}$,

$s_{j0}(\pm\tau) = \frac{a_{j0}b_{j0}\pm\tau}{u_{j0}^2}, r_{j0}(\pm\tau) = \exp\left(\frac{1}{2}\frac{(a_{j0}b_{j0}\pm\tau)^2}{u_{j0}^2} - \frac{1}{2}b_{j0}^2\right)$,

and

$$I_{1,j,j'} = \Phi(b_{j'0}) + r_{j'0}(+\tau)\frac{a_{j'0}}{2u_{j'0}}(1 - \Phi(s_{j'0}(+\tau)))$$

$$I_{2,j,j'} = \Phi(0) + r_{j'0}(+\tau)\frac{a_{j'0}}{2u_{j'0}}[1 - \Phi(u_{j'0}(-\frac{b_{j'0}}{a_{j'0}} + s_{j'0}(+\tau)))]$$

$$I_{3,j,j'} = -a_{j'0}\frac{r_{j'0}(-\tau)}{2u_{j'0}}[\Phi(\frac{a_{j'0}b_{j'0}-\tau}{u_{j'0}}) - \Phi(u_{j'0}(-\frac{b_{j'0}}{a_{j'0}} + s_{j'0}(-\tau)))]$$

$$I_{4,j,j'} = -a_{j'0}\frac{r_{j'0}(-\tau)}{2u_{j'0}}[\Phi(\frac{a_{j'0}b_{j'0}-\tau}{u_{j'0}}) - \Phi(u_{j'0}(\frac{\alpha_{j0}}{a_{j'0}} + s_{j'0}(-\tau)))]$$

$$I_{5,j,j'} = -a_{j'0}\frac{r_{j'0}(-\tau)}{2u_{j'0}}[\Phi(\frac{a_{j'0}b_{j'0}-\tau}{u_{j'0}}) - \Phi(u_{j'0}(\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} + s_{j'0}(-\tau)))]$$

$$I_{6,j,j'} = \Phi(\alpha_{j0}) + r_{j'0}(+\tau)\frac{a_{j'0}}{2u_{j'0}}[1 - \Phi(u_{j'0}(\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} + s_{j'0}(+\tau)))]$$

$$I_{7,j,j'} = \Phi(\alpha_{j0} + b_{j'0}) + r_{j'0}(+\tau)\frac{a_{j'0}}{2u_{j'0}}[1 - \Phi(u_{j'0}(\frac{\alpha_{j0}}{a_{j'0}} + s_{j'0}(+\tau)))]$$

Rule 1. Let $\mathbf{1}_{\{.\}}$ be an indicator function; and

$$\begin{cases} A_1 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} = 0 \Rightarrow \alpha_{j0} = 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} = 0 \Rightarrow, b_{j'0} = 0 \end{cases} & \Rightarrow \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in \{\} \end{cases} \\ A_2 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} = 0 \Rightarrow \alpha_{j0} = 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} > 0 \Rightarrow, b_{j'0} > 0 \end{cases} & \Rightarrow \begin{cases} t \in [0,\infty), t \in [\frac{-b_{j'0}}{a_{j'0}}, \infty) \xrightarrow{\frac{-b_{j'0}}{a_{j'0}}>0} t \in [\frac{-b_{j'0}}{a_{j'0}}, \infty) \\ t \in (-\infty,0), t \in [\frac{-b_{j'0}}{a_{j'0}}, \infty) \xrightarrow{\frac{-b_{j'0}}{a_{j'0}}>0} t \in \{\} \end{cases} \\ A_3 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} = 0 \Rightarrow \alpha_{j0} = 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} < 0 \Rightarrow, b_{j'0} < 0 \end{cases} & \Rightarrow \begin{cases} t \in [0,\infty), t \in [\frac{-b_{j'0}}{a_{j'0}}, \infty) \xrightarrow{\frac{-b_{j'0}}{a_{j'0}}<0} t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{-b_{j'0}}{a_{j'0}}, \infty) \xrightarrow{\frac{-b_{j'0}}{a_{j'0}}<0} t \in [\frac{-b_{j'0}}{a_{j'0}}, 0) \end{cases} \end{cases}$$

Assume $c_{j,j'} = [(\sigma^2+1)^{-\frac{1}{2}}\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} - \frac{(\sigma^2+1)^{\frac{1}{2}}}{\sigma^2}\boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta}]$, $\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} = \frac{c_{j,j'}}{a_{j'0}}$

$$\begin{cases} A_4 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} > 0 \Rightarrow \alpha_{j0} > 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} = 0 \Rightarrow, b_{j'0} = 0 \end{cases} \xrightarrow{\frac{\alpha_{j0}}{a_{j'0}}<0} \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}}{a_{j'0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0], t \in [\frac{\alpha_{j0}}{a_{j'0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}}{a_{j'0}},0] \end{cases} \\ A_5 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} > 0 \Rightarrow \alpha_{j0} > 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} > 0 \Rightarrow, b_{j'0} > 0 \end{cases} \Rightarrow \begin{cases} A_{51} : c_{j,j'} = 0 \Rightarrow \frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} = 0 \\ A_{52} : c_{j,j'} > 0 \Rightarrow \frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} < 0 \\ A_{53} : c_{j,j'} < 0 \Rightarrow \frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} > 0 \end{cases} \\ A_6 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} > 0 \Rightarrow \alpha_{j0} > 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} < 0 \Rightarrow, b_{j'0} < 0 \end{cases} \xrightarrow{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}<0} \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},0] \end{cases} \end{cases}$$

Also,

$$\begin{cases} A_{51} : \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in \{\} \end{cases} \\ A_{52} : \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},0) \end{cases} \\ A_{53} : \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in \{\} \end{cases} \end{cases}$$

$$\begin{cases} A_7 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} < 0 \Rightarrow \alpha_{j0} < 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} = 0 \Rightarrow, b_{j'0} = 0 \end{cases} \xrightarrow{\frac{\alpha_{j0}}{a_{j'0}}>0} \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}}{a_{j'0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}}{a_{j'0}},\infty) \\ t \in (-\infty,0], t \in [\frac{\alpha_{j0}}{a_{j'0}},\infty) \Rightarrow t \in \{\} \end{cases} \\ A_8 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} < 0 \Rightarrow \alpha_{j0} < 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} > 0 \Rightarrow, b_{j'0} > 0 \end{cases} \xrightarrow{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}>0} \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}},\infty) \Rightarrow t \in \{\} \end{cases} \\ A_9 : \begin{cases} \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} < 0 \Rightarrow \alpha_{j0} < 0 \\ \boldsymbol{f}^\top(\boldsymbol{x}_{j'})\boldsymbol{\beta} < 0 \Rightarrow, b_{j'0} < 0 \end{cases} \Rightarrow \begin{cases} A_{91} : c_{j,j'} = 0 \Rightarrow \frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} = 0 \\ A_{92} : c_{j,j'} > 0 \Rightarrow \frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} < 0 \\ A_{93} : c_{j,j'} < 0 \Rightarrow \frac{\alpha_{j0}-b_{j'0}}{a_{j'0}} > 0 \end{cases} \end{cases}$$

Moreover,

$$\begin{cases} A_{91}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in \{\} \end{cases} \\ A_{92}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in \{\} \end{cases} \\ A_{93}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, \infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}, 0] \end{cases} \end{cases}$$

Then, in the binary mixed probit regression model

$$\text{cov}(Y_{ijk}, Y_{ij'k'}) \approx \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] - \Phi(\alpha_{j0})\Phi(\alpha_{j'0}); j \neq j', k, k' \tag{A.1}$$

where

$$\begin{aligned} \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx\ & I_{1,j,j'}\mathbf{1}_{\{A_1\}} + I_{2,j,j'}\mathbf{1}_{\{A_2\}} + (I_{1,j,j'} + I_{3,j,j'})\mathbf{1}_{\{A_3\}} \\ & +(I_{1,j,j'} + I_{4,j,j'})\mathbf{1}_{\{A_4\}} + I_{1,j,j'}\mathbf{1}_{\{A_{51}\}} + (I_{1,j,j'} + I_{5,j,j'})\mathbf{1}_{\{A_{52}\}} \\ & +I_{6,j,j'}\mathbf{1}_{\{A_{53}\}} + (I_{1,j,j'} + I_{5,j,j'})\mathbf{1}_{\{A_6\}} + I_{7,j,j'}\mathbf{1}_{\{A_7\}} + I_{6,j,j'}\mathbf{1}_{\{A_8\}} \\ & +I_{1,j,j'}\mathbf{1}_{\{A_{91}\}} + I_{6,j,j'}\mathbf{1}_{\{A_{92}\}} + (I_{1,j,j'} + I_{5,j,j'})\mathbf{1}_{\{A_{93}\}}. \end{aligned} \tag{A.2}$$

Proof.

In order to gain $\mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})]$ in (3.4) for the computation of $\text{cov}(Y_{ijk}, Y_{ij'k'})$, and according to (3.13) it is needed to solve

$$\int_{-\infty}^{\alpha_{j0}} \phi(z)\Phi(q_{j'0}(z))\mathrm{d}z. \tag{A.3}$$

By changing variable principle,

$$\int_{-\infty}^{\alpha_{j0}} \phi(z)\Phi(q_{j'0}(z))\mathrm{d}z \approx -\int_{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}}^{+\infty} a_{j'0}\phi(a_{j'0}t + b_{j'0})\Phi^{\star}(t)\mathrm{d}t. \tag{A.4}$$

According to the approximation of the cumulative distribution function of the standard Normal distribution in (3.21), for positive and negative signs of t, and the integral bound in the right hand side of (A.4), there exist nine different conditions $A_1, ..., A_9$. Therefore, there are different integral calculations for the achievement of $\mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})]$ based on the corresponding sets. They are illustrated as follows:

$$A_1: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$A_2: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_{\frac{-b_{j'0}}{a_{j'0}}}^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$A_3: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$-\int_{\frac{-b_{j'0}}{a_{j'0}}}^0 a_{j'0}\phi(a_{j'0}t + b_{j'0})(\tfrac{1}{2}\exp(-\tau(-t) - \alpha(-t)^2))dt$$

$$A_4: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$-\int_{\frac{\alpha_{j0}}{a_{j'0}}}^0 a_{j'0}\phi(a_{j'0}t + b_{j'0})(\tfrac{1}{2}\exp(-\tau(-t) - \alpha(-t)^2))dt$$

$$A_{5.1}: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$A_{52}: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$-\int_{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}}^0 a_{j'0}\phi(a_{j'0}t + b_{j'0})(\tfrac{1}{2}\exp(-\tau(-t) - \alpha(-t)^2))dt$$

$$A_{53}: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}}^\infty a_{j'0}\phi(a_{j'0}t+b_{j'0})(1-\tfrac{1}{2}\exp(-\tau t-\alpha t^2))dt$$

$$A_6: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$-\int_{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}}^0 a_{j'0}\phi(a_{j'0}t + b_{j'0})(\tfrac{1}{2}\exp(-\tau(-t) - \alpha(-t)^2))dt$$

$$A_7: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_{\frac{\alpha_{j0}}{a_{j'0}}}^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$A_8: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}}^\infty a_{j'0}\phi(a_{j'0}t+b_{j'0})(1-\tfrac{1}{2}\exp(-\tau t-\alpha t^2))dt$$

$$A_{9.1}: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$A_{9.2}: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}}^\infty a_{j'0}\phi(a_{j'0}t+b_{j'0})(1-\tfrac{1}{2}\exp(-\tau t-\alpha t^2))dt$$

$$A_{9.3}: \quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx -\int_0^\infty a_{j'0}\phi(a_{j'0}t + b_{j'0})(1 - \tfrac{1}{2}\exp(-\tau t - \alpha t^2))dt$$

$$-\int_{\frac{\alpha_{j0}-b_{j'0}}{a_{j'0}}}^{0} a_{j'0}\phi(a_{j'0}t+b_{j'0})(\tfrac{1}{2}\exp(-\tau(-t)-\alpha(-t)^2))dt$$

where $\phi(.)$ is the density function of the standard Normal distribution. Then, by calculating the integrals in each condition above, the following terms are obtained for the achievement of $\mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})]$.

Under A_1 : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'}$

Under A_2 : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{2,j,j'}$

Under A_3 : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'} + I_{3,j,j'}$

Under A_4 : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'} + I_{4,j,j'}$

Under A_{51} : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'}$

Under A_{52} : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'} + I_{5,j,j'}$

Under A_{53} : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{6,j,j'}$

Under A_6 : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'} + I_{5,j,j'}$

Under A_7 : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{7,j,j'}$

Under A_8 : $\quad \mathrm{E}_{(p)}[p_{ij}^{(\zeta_{i0})}p_{ij'}^{(\zeta_{i0})}] \approx I_{6,j,j'}$

Under A_{91} : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'}$

Under A_{92} : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{6,j,j'}$

Under A_{93} : $\quad \mathrm{E}_{(p)}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})] \approx I_{1,j,j'} + I_{5,j,j'}$

Finally, $\mathrm{cov}(Y_{ijk}, Y_{ij'k'})$ is approximated. ■

Rule 2. Let $\mathbf{1}_{\{.\}}$ be an indicator function; and

$$\begin{cases} B_1 : \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} = 0 \Rightarrow \alpha_{j0} = 0, b_{j0} = 0. \Rightarrow \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}}, \infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty, 0), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}}, \infty) \Rightarrow t \in \{\} \end{cases} \\ B_2 : \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} > 0 \Rightarrow \alpha_{j0} > 0, b_{j0} > 0 \Rightarrow \begin{cases} B_{21} : c_{j,j} = 0 \Rightarrow \frac{\alpha_{j0}-b_{j0}}{a_{j0}} = 0 \\ B_{22} : c_{j,j} > 0 \Rightarrow \frac{\alpha_{j0}-b_{j0}}{a_{j0}} < 0 \\ B_{23} : c_{j,j} < 0 \Rightarrow \frac{\alpha_{j0}-b_{j0}}{a_{j0}} > 0 \end{cases} \\ B_3 : \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} < 0 \Rightarrow \alpha_{j0} < 0, b_{j0} < 0 \Rightarrow \begin{cases} B_{31} : c_{j,j} = 0 \Rightarrow \frac{\alpha_{j0}-b_{j0}}{a_{j0}} = 0 \\ B_{32} : c_{j,j} > 0 \Rightarrow \frac{\alpha_{j0}-b_{j0}}{a_{j0}} < 0 \\ B_{33} : c_{j,j} < 0 \Rightarrow \frac{\alpha_{j0}-b_{j0}}{a_{j0}} > 0 \end{cases} \end{cases}$$

Moreover,

$$\begin{cases} B_{21}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in \{\} \end{cases} \\ B_{22}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},0] \end{cases} \\ B_{23}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in \{\} \end{cases} \end{cases}$$

and

$$\begin{cases} B_{31}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in \{\} \end{cases} \\ B_{32}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [0,\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},0] \end{cases} \\ B_{33}: \begin{cases} t \in [0,\infty), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \\ t \in (-\infty,0), t \in [\frac{\alpha_{j0}-b_{j0}}{a_{j0}},\infty) \Rightarrow t \in \{\} \end{cases} \end{cases}$$

then, in the binary mixed probit regression model

$$\mathrm{cov}(Y_{ijk}, Y_{ijk'}) \approx \mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] - [\Phi(\alpha_{j0})]^2; k \neq k', \tag{A.5}$$

where

$$\begin{aligned} \mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] &\approx I_{1,j,j}\mathbf{1}_{\{B_1\}} + I_{1,j,j}\mathbf{1}_{\{B_{21}\}} + (I_{1,j,j} + I_{5,j,j})\mathbf{1}_{\{B_{22}\}} \\ &+ I_{6,j,j}\mathbf{1}_{\{B_{23}\}} + I_{1,j,j}\mathbf{1}_{\{B_{31}\}} + (I_{1,j,j} + I_{5,j,j})\mathbf{1}_{\{B_{32}\}} + I_{6,j,j}\mathbf{1}_{\{B_{33}\}}. \end{aligned} \tag{A.6}$$

Proof.

For the proof of this Corollary, it is merely needed to substitute j' for j in **Rule 1**. Therefore, the following results under the sets $B_1, ..., B_{33}$ are obtained.

Under B_1 : $\mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] \approx I_{1,j,j}$

Under B_{21} : $\mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] \approx I_{1,j,j}$

Under B_{22} : $\mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] \approx I_{1,j,j} + I_{5,j,j}$

Under B_{23} : $\mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] \approx I_{6,j,j}$

Under B_{31} : $\mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] \approx I_{1,j,j}$

Under B_{32} : $\mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] \approx I_{1,j,j} + I_{5,j,j}$

Under B_{33} : $\mathrm{E}_{(p)}[p_{ij}^2(\zeta_{i0})] \approx I_{6,j,j}$

Finally, $\mathrm{cov}(Y_{ijk}, Y_{ijk'})$ is approximated.

A.3 First approximation of covariance (binary mixed logit model)

Lemma A.3.1. *Let* $\alpha_{j0(c)} = (1 + c^2\sigma^2)^{-\frac{1}{2}} \boldsymbol{f}^\top(\boldsymbol{x}_j\boldsymbol{\beta})$ *and*

$$q_{j0(c)}(z) = \frac{(c^2\sigma^2+1)^{\frac{1}{2}}}{(2c^2\sigma^2+1)^{\frac{1}{2}}}(\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} - \frac{c^2\sigma^2 z}{(c^2\sigma^2+1)^{\frac{1}{2}}}); \tag{A.7}$$

then,

$$\mathrm{Var}(Y_{ijk}) \approx [\Phi(\alpha_{j0(c)})][1 - \Phi(\alpha_{j0(c)})], \tag{A.8}$$

$$\mathrm{cov}(Y_{ijk}, Y_{ijk'}) \approx \int_{-\infty}^{\alpha_{j0(c)}} \phi(z)\Phi(q_{j0(c)}(z))dz - \{\Phi(\alpha_{j0(c)})\}^2; k \neq k' \tag{A.9}$$

and

$$\mathrm{cov}(Y_{ijk}, Y_{ij'k'}) \approx \int_{-\infty}^{\alpha_{j0(c)}} \phi(z)\Phi(q_{j'(c)}(z))dz - \{\Phi(\alpha_{j0(c)})\}\{\Phi(\alpha_{j'(c)})\}; j \neq j', k, k' \tag{A.10}$$

The proof is the same as for the proof in Lemma 3.1.2 except for the approximation of the cumulative distribution function of the logistic distribution with cumulative distribution function of the Normal distribution, $N(0, \frac{1}{c^2})$ in the integrand of $\mathrm{E}[p_{ij}^2(\zeta_{i0})]$, and $\mathrm{E}[p_{ij}(\zeta_{i0})p_{ij'}(\zeta_{i0})]$.

The second approximation of covariance elements (3.3) and (3.4) for the binary mixed logistic regression model is obtained in the same way as the results in **Rule 1** for the random intercept probit regression model, where $\alpha_{j0} = \alpha_{j0(c)}$, $a_{j0} = a_{j0(c)}$, $u_{j0} = u_{j0(c)}$, $b_{j0} = b_{j0(c)}$, $r_{j0} = r_{j0(c)}(\pm\tau)$ and $s_{j0}(\pm\tau) = s_{j(c)}(\pm\tau)$, $u_{j0(c)} = \sqrt{a_{j0(c)}^2 + 2\alpha}$, in which

$a_{j0(c)} = -\frac{(2c^2\sigma^2+1)^{\frac{1}{2}}}{c^2\sigma^2}$,

$b_{j0(c)} = \frac{(c^2\sigma^2+1)^{\frac{1}{2}}}{c^2\sigma^2}\boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta}$,

$r_{j0(c)}(\pm\tau) = \frac{a_{j0(c)}}{2u_{j0(c)}} \exp\left(\frac{1}{2}\frac{(a_{j0(c)}b_{j0(c)}\pm\tau)^2}{a_{j0(c)}^2+2\alpha} - \frac{1}{2}b_{j0(c)}^2\right)$,

$s_{j(c)}(\pm\tau) = \frac{a_{j0(c)}b_{j0(c)}\pm\tau}{a_{j0(c)}^2+2\alpha}$.

Theorem A.3.2. *The components of matrix* $\mathbf{D}$ *in (2.46) for the binary logit mixed regression model is formed as follows:*

$$\mathbf{D}_j^\top \approx (c^2\sigma^2+1)^{-\frac{1}{2}}\boldsymbol{f}^\top(\boldsymbol{x}_j)\phi(\alpha_{j(c)}) \tag{A.11}$$

Proof. By the definition of $\mathbf{D}$ in (2.46) and $\mathbf{D}_j$, the following equation leads to the result in (A.11).

$$\frac{\partial}{\partial\boldsymbol{\beta}^\top}\pi_j \approx \frac{\partial}{\partial\boldsymbol{\beta}^\top}\Phi(\alpha_{j(c)}), \tag{A.12}$$

since $\mathrm{E}(Y_{ijk}) \approx \Phi(\alpha_{j0(c)})$. ■

A.4 Proof of Lemma 3.1.4.

$$\mathbf{V} = \begin{pmatrix} v_1\mathbf{1}_{n_1}\mathbf{1}_{n_1}^\top & c_{12}\mathbf{1}_{n_1}\mathbf{1}_{n_2}^\top \\ c_{21}\mathbf{1}_{n_2}\mathbf{1}_{n_1}^\top & v_2\mathbf{1}_{n_2}\mathbf{1}_{n_2}^\top \end{pmatrix} + \begin{pmatrix} (\pi_1(1-\pi_1)-v_1)\mathbf{I}_{n_1} & 0 \\ 0 & (\pi_2(1-\pi_2)-v_2)\mathbf{I}_{n_2} \end{pmatrix}. \tag{A.13}$$

The eigenvalues of the matrix are stated as follows:

$e_{j1} = \frac{1}{2}\{(\pi_1(1-\pi_1)+\pi_2(1-\pi_2)+(n_1-1)v_1+(n_2-1)v_2)\pm$

$([(\pi_1(1-\pi_1)-\pi_2(1-\pi_2))+((n_1-1)v_1-(n_2-1)v_2)]^2+4n_1n_2c_{12}c_{21})^{\frac{1}{2}}\}; j=1,2,$

$e_{jk} = \pi_j(1-\pi_j)-v_j; j=1,2, k=2,...,n_{ij}.$

In order to evaluate the positive definiteness of the approximate variance matrix, it is sufficient to check the positiveness of the eigenvalues for $j=1,2$, $k=1,2,...,n_{ij}$; therefore,

1- $e_{11}>0$, $e_{21}>0 \Leftrightarrow \det(\Upsilon)>0$, since

$$\begin{aligned} e_{11}>0,\ e_{21}>0 \Leftrightarrow & (\pi_1(1-\pi_1)+\pi_2(1-\pi_2)+(n_1-1)v_1+(n_2-1)v_2)^2 \geq \\ & (\pi_1(1-\pi_1)-\pi_2(1-\pi_2)+(n_1-1)v_1-(n_2-1)v_2)^2+4n_1n_2c_{12}c_{21} \\ \Leftrightarrow \quad & 4\pi_1(1-\pi_1)\pi_2(1-\pi_2)+4\pi_1(1-\pi_1)(n_2-1)v_2+4\pi_2(1-\pi_2)(n_1-1)v_1 \\ & +4(n_1-1)v_1(n_2-1)v_2-4n_1n_2c_{12}c_{21}>0 \\ \Leftrightarrow \quad & \det(\Upsilon)>0 \end{aligned}$$

.

2- $e_{jk}>0 \Leftrightarrow [\pi_j(1-\pi_j)-v_j]>0$; $j=1,2, k=2,...,n_j$. ■

A.5 Optimum design in binary mixed model

Proof of Theorem 3.3.1.

In accordance with the ith quasi Fisher information matrix in (2.69), matrix $\mathbf{D}$ and $\mathbf{V}$ are constructed. The inverse of matrix $\mathbf{V}$, which is a block matrix with reference to (Schott 1997, Corollary 1.7.2 in Appendix A, Section A.10 and Puntanen 2007, Section 3.4), is indicated as follows:

$$\mathbf{V}^{-1} = \begin{pmatrix} \frac{1}{\pi_1(1-\pi_1)-v_1}[\mathbf{I}_{n_1} - \frac{1}{n_1}\frac{\det \Upsilon^{**}}{\det \Upsilon}\mathbf{1}_{n_1}\mathbf{1}_{n_1}^\top] & -\frac{c_{12}}{\det \Upsilon}\mathbf{1}_{n_1}\mathbf{1}_{n_2}^\top \\ -\frac{c_{21}}{\det \Upsilon}\mathbf{1}_{n_2}\mathbf{1}_{n_1}^\top & \frac{1}{\pi_2(1-\pi_2)-v_2}[\mathbf{I}_{n_2} - \frac{1}{n_2}\frac{\det \Upsilon^{*}}{\det \Upsilon}\mathbf{1}_{n_2}\mathbf{1}_{n_2}^\top] \end{pmatrix},$$

where

$$\Upsilon^{**} = \begin{pmatrix} n_1 v_1 & n_1 c_{12} \\ n_2 c_{21} & \pi_2(1-\pi_2) + (n_2 - 1)v_2 \end{pmatrix},$$

$$\Upsilon^{*} = \begin{pmatrix} \pi_1(1-\pi_1) + (n_1 - 1)v_1 & n_1 c_{12} \\ n_2 c_{21} & n_2 v_2 \end{pmatrix}.$$

Substitution of $\mathbf{V}^{-1}$ and $\mathbf{D}$ into (2.69) leads to:

$$\mathbf{M}^Q(\xi, \boldsymbol{\beta}) = \frac{1}{\det \Upsilon}\mathbf{D}^{(0)\top}\boldsymbol{Q}\mathbf{D}^{(0)}. \tag{A.14}$$

Furthermore, given the design points (x_1, x_2) and model parameters, it is aimed to maximize individual D-optimality criterion (2.76) with respect to w_1.

Since $\det(\mathbf{D}^{(0)\top}) = \det(\mathbf{D}^{(0)})$,

$$\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta})) = \left(\frac{1}{\det \Upsilon}\right)^2 [\det(\mathbf{D}^{(0)})]^2 \det(\boldsymbol{Q}). \tag{A.15}$$

We set $\frac{d}{dw_1}\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta})) = 0$, which is analogous to $\frac{d}{dw_1}\frac{n^2 w_1(1-w_1)}{\det(\Upsilon)} = 0$.

By algebraic calculation, it was obtained that

$$\frac{n^3\lambda_d + 2n^2\lambda_{n1}w_1 - n^2\lambda_{n1}}{[\det(\Upsilon)]^2} = 0. \tag{A.16}$$

As $\lambda_d \neq 0$, the root w_1^* from the latter equation causes the expression in (3.27). As $\lambda_d = 0$; then, $w_1^\star = \frac{1}{2}$. Also, it was checked that $w_1^\star \in [0, 1]$.

Finally, the optimality of $\det(\mathbf{M}^Q(\xi, \boldsymbol{\beta}))$ is checked by the second derivative of the corresponding function at $w_1^\star$ and it is shown that for all $w_1 \in [0, 1]$ the second derivative is negative. ■

A.6 Covariance approximation in ordinal mixed logit model

Lemma A.6.1. *Let* $q_{j0(c)}^{(m)}(z) = \frac{(1+c^2\sigma^2)^{\frac{1}{2}}}{(1+c^2 2\sigma^2)^{\frac{1}{2}}}\left(\gamma_m - \boldsymbol{f}^\top(\boldsymbol{x}_j)\boldsymbol{\beta} - \frac{c^2\sigma^2 z}{(1+c^2\sigma^2)^{\frac{1}{2}}}\right)$*; then, in the ordinal mixed effects regression model with the logit link function, for all* $j, j' = 1, ..., J, m, m' = 1, ..., M$,

$$\mathrm{E}[p_{ij}^{(m)}(\zeta_{i0})p_{ij'}^{(m')}(\zeta_{i0})] \approx \int_{-\infty}^{\alpha_{j0(c)}^{(m)}} \phi(z)\Phi(q_{j'0(c)}^{(m')}(z))\mathrm{d}z \tag{A.17}$$

A.7 Density function of random effect in Non-linear longitudinal Poisson model

The multivariate density function of $\boldsymbol{\Lambda}_i$ using the approach of Mathai and Moschopoulos 1991 is given as follows:

$$f_{\boldsymbol{\Lambda}_i}(\boldsymbol{\lambda}_i;\rho,\tau) = (J-1)\sum_{m=1}^{J}\mathbf{f}_m(\lambda_{i1},...,\lambda_{iJ};\rho,\tau)\mathbf{1}_{\{\lambda_m=\min\{\lambda_{i1},...,\lambda_{iJ}\}\}}, \tag{A.18}$$

$\mathbf{f}_m(\lambda_{i1},...,\lambda_{iJ}) =$

$C\{\prod_{j=1}^{J}\lambda_{ij}\}^{((\frac{1-\rho}{\tau})-1)}\lambda_{im}^{\frac{\rho}{\tau}}\sum_{r_0,...,r_J=0}^{\infty}[\frac{(-\lambda_{im})^{r_0}}{r_0!}.\frac{(-\lambda_{im})^{r_m}}{r_m!}.\ \mathrm{E}(\prod_{j\neq m}\frac{(-\lambda_{ij})^{r_{j0}}}{r_{j0}!}).$

$Beta(\frac{\rho}{\tau}+r_0,(\frac{1-\rho}{\tau})+r_m).F_D(\frac{\rho}{\tau}+r_0,(\frac{1-\rho}{\tau})+r_1,...,(\frac{1-\rho}{\tau})+r_{m-1},(\frac{1-\rho}{\tau})+r_{m+1},...$

$$...,(\frac{1-\rho}{\tau})+r_J;\frac{\rho}{\tau}+r_0+(\frac{1-\rho}{\tau})+r_m;\frac{\lambda_{im}}{\lambda_{i1}},...,\frac{\lambda_{im}}{\lambda_{i(m-1)}},\frac{\lambda_{im}}{\lambda_{i(m+1)}},...,\frac{\lambda_{im}}{\lambda_{iJ}})], \tag{A.19}$$

where

$$C = (\tau)^{J((\frac{1-\rho}{\tau})-1)}\{\Gamma(\frac{\rho}{\tau})(\tau\rho\tau\mathrm{e}_1\mathrm{e}_1^\top)^{J(\frac{1-\rho}{\tau})}\Gamma(\frac{1-\rho}{\tau})^J\}^{-1},$$

and F_D is the Lauricella function (A. M. Mathai and Saxena 1978) having a convergent series representation for $\mid \frac{\lambda_{im}}{\lambda_{ij}} \mid < 1; \lambda_{ij} > 0; j = 1, ..., J;\ j \neq m$ and positive domain of $\frac{\rho}{\tau}$ and $(\frac{1-\rho}{\tau})$. λ_{im} is unique with probability one.

A.8 Marginal density function of response in Nonlinear longitudinal Poisson model

The marginal density function of model (2.33) is calculated from equations (2.37) and (2.36)

For finding the marginal density $L(\boldsymbol{\beta}, \boldsymbol{y}_i)$, the conditional density of $\boldsymbol{Y}_i$, is firstly calculated:

$$f_{\mathbf{Y}_i|\boldsymbol{\Lambda}_i=\boldsymbol{\lambda}_i,\boldsymbol{\beta}}(\boldsymbol{y}_i) = \prod_{j=1}^{J}\prod_{k=1}^{n_j} \frac{\mu_j(\lambda_{ij})^{y_{ijk}}}{y_{ijk}!} \exp(-\mu_j(\lambda_{ij})), \tag{A.20}$$

where $\mu_j(\lambda_{ij}) = \lambda_{ij}\exp(\theta_{j0} + \sigma_{ij})$.

Lemma A.8.1. *The marginal density function of the repeated Poisson model (2.33) in terms of the i^{th} subject with $J = 3$ is obtained as:*

$f_{\mathbf{Y_i};\boldsymbol{\beta}}(\boldsymbol{y}_i) =$

$$2\{\textstyle\prod_{j=1}^{3}\exp(\theta_j n_j \bar{y}_{j.})\}\{\prod_{j=1}^{3}\prod_{k=1}^{n_j}\frac{1}{Y_{ijk}!}\}C\sum_{r_0,\ldots,r_3=0}^{\infty} \displaystyle\prod_{n_1,n_2,n_3}\sum_{n_j\neq n_m}\frac{(-1)^{\sum_{j=0}^{3} r_{j0}}}{\prod_{j=0}^{3} r_{j0}!}$$

$$Beta(\tfrac{\rho}{\tau} + r_0, (\tfrac{1-\rho}{\tau}) + r_m).\frac{\displaystyle\prod_{j\neq m}(-((\frac{1-\rho}{\tau}) + r_{j0} - 1))_{n_j}(\frac{\rho}{\tau} + r_0)_{\sum_{j\neq m} n_j}}{(\frac{\rho}{\tau}+r_0+(\frac{1-\rho}{\tau})+r_m)_{\sum_{j\neq m} n_j}\displaystyle\prod_{j\neq m} n_j!}$$

$$\Gamma(n_m\bar{y}_{m.} + \tfrac{\rho}{\tau} + r_0 + r_m + (\tfrac{1-\rho}{\tau}) + \sum_{j\neq m} n_j)(n_m\exp(\theta_m))^{-(n_m\bar{y}_{m.}+\frac{\rho}{\tau}+r_0+r_m+(\frac{1-\rho}{\tau})+\sum_{j\neq m} n_j)}$$

$$\prod_{j\neq m}\Gamma(n_j\bar{y}_{j.}+(\frac{1-\rho}{\tau})+r_{j0}-m_j)(n_j\mu_{j0})^{-(n_j\bar{y}_{j.}+(\frac{1-\rho}{\tau})+r_{j0}-m_j)}. \tag{A.21}$$

Proof.

The marginal density function of the response variable in (2.36) for model (2.33) is obtained from the following calculation:

$$f_{\mathbf{Y}_i;\boldsymbol{\beta}}(\boldsymbol{y}_i) = \int_0^\infty \int_0^\infty \int_0^\infty f_{\mathbf{Y}_i|\boldsymbol{\Lambda}_i,\boldsymbol{\beta}}(\boldsymbol{y}_i;\boldsymbol{\beta} \mid \boldsymbol{\lambda}).[2\sum_{m=1}^{3} \mathbf{f}_m(\lambda_{i1},\lambda_{i2},\lambda_{i3})]\mathrm{d}\lambda_{i1}\mathrm{d}\lambda_{i2}\mathrm{d}\lambda_{i3}, \tag{A.22}$$

the multivariate density function of $\boldsymbol{\lambda}_i$ is defined in (A.18) and (A.19). Substituting this term and the conditional density of the response variable (A.20) above into (A.22), and solving the integral using the application of the Gamma function leads to (A.21). ■

A.9 Proofs of Lemmas and Theorem in Chapter 5

Proof of Lemma 5.1.1.

We will make repeated use of the inversion formula from corollary 1.7.2 (Schott 1997), Appendix A, Section A.10.

For positive definite $\mathbf{A} = \mu_j \mathbf{I}_{n_j}$ and $\mathbf{b} = \mathbf{c} = \sqrt{(1-\rho)\tau \mu_j} = \sqrt{(1-\rho)\tau}\mu_j \mathbf{1}_{n_j}$, we obtain

$$\mathbf{C}_j^{-1} = \frac{1}{\mu_j}\mathbf{I}_{n_j} - \frac{(1-\rho)\tau}{1+(1-\rho)\tau\mu_j n_j}\mathbf{1}_{n_j}\mathbf{1}_{n_j}^\top$$

and hence, $\mathbf{C}^{-1} = \mathrm{diag}(\mathbf{C}_1^{-1},\ldots,\mathbf{C}_J^{-1})$.

Next, we note that $\mathbf{C}_j^{-1}\boldsymbol{\mu}_j = \frac{1}{1+(1-\rho)\tau\mu_j n_j}\mathbf{1}_{n_j}$ and $\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j = \frac{\mu_j n_j}{1+(1-\rho)\tau\mu_j n_j} = \psi_j$.

Now, with $\mathbf{A} = \mathbf{C}$ and $\mathbf{b} = \mathbf{c} = \sqrt{\rho\tau}\boldsymbol{\mu}$, we obtain by the inversion formula that

$$\mathbf{V}^{-1} = \mathbf{C}^{-1} - \frac{\rho\tau}{1+\rho\tau\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j}\mathbf{C}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{C}^{-1}$$

which proves the Lemma.

■

Proof of Lemma 5.1.2.

First of all, we obtain $\mathbf{C}^{-1} = \mathrm{diag}(\mathbf{C}_1^{-1},...,\mathbf{C}_J^{-1})$, in which $\mathbf{C}_j^{-1} = \frac{1}{\mu_j}[\mathbf{I}_{n_j} - \iota_{j0}\mathbf{1}_{n_j}\mathbf{1}_{n_j}^\top]$, in which $\iota_{j0} = \frac{(1-\rho)\tau\mu_j}{1+(1-\rho)\tau\mu_j n_j}$. $\mathbf{C}_j{}^{-1}$ is obtained from corollary 1.7.2 Schott 1997, Appendix A, Section A.10.

Then, by the substitution of $\mathbf{V}^{-1}$ in (5.4) in the individual quasi-Fisher information matrix, we result in the following form:

$$\mathbf{M}^Q = \mathbf{D}^\top\mathbf{C}^{-1}\mathbf{D} - \mathbf{D}^\top\frac{\rho\tau}{1+\rho\tau\sum_{j=1}^{J}\psi_j}\mathbf{C}^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^\top\mathbf{C}^{-1}\mathbf{D}. \tag{A.23}$$

Note that $\mathbf{D}_j^\top \mathbf{C}_j^{-1} = \frac{1}{\mu_j(1+(1-\rho)\tau n_j \mu_j)} \boldsymbol{d}_j \mathbf{1}_{n_j}^\top$, since

$$\begin{aligned}
\mathbf{D}_j^\top \mathbf{C}_j^{-1} &= \mu_j \mathbf{D}_{0j}^\top \tfrac{1}{\mu_j}[\mathbf{I}_{n_j} - \iota_{j0} \mathbf{1}_{n_j} \mathbf{1}_{n_j}^\top] \\
&= \boldsymbol{d}_{0j} \mathbf{1}_{n_j}^\top [\mathbf{I}_{n_j} - \iota_{j0} \mathbf{1}_{n_j} \mathbf{1}_{n_j}^\top] \\
&= \boldsymbol{d}_{0j} [\mathbf{1}_{n_j}^\top - \iota_{j0} n_j \mathbf{1}_{n_j}^\top] \\
&= \boldsymbol{d}_{0j} [1 - \iota_{j0} n_j] \mathbf{1}_{n_j}^\top \\
&= \tfrac{1}{1+(1-\rho)\tau \mu_j n_j} \boldsymbol{d}_{0j} \mathbf{1}_{n_j}^\top \\
&= \tfrac{1}{(1+(1-\rho)\tau \mu_j n_j)\mu_j} \boldsymbol{d}_j \mathbf{1}_{n_j}^\top .
\end{aligned}$$

Then,

$$\begin{aligned}
\mathbf{D}_j^\top \mathbf{C}_j^{-1} \mathbf{D}_j &= \tfrac{1}{(1+(1-\rho)\tau \mu_j n_j)\mu_j} \boldsymbol{d}_j \mathbf{1}_{n_j}^\top \mathbf{1}_{n_j} \boldsymbol{d}_j^\top \\
&= \tfrac{1}{1+(1-\rho)\tau \mu_j n_j} \boldsymbol{d}_{0j} \mathbf{1}_{n_j}^\top \mathbf{1}_{n_j} \mu_j \boldsymbol{d}_{0j}^\top \\
&= \tfrac{n_j \mu_j}{1+(1-\rho)\tau \mu_j n_j} \boldsymbol{d}_{0j} \boldsymbol{d}_{0j}^\top
\end{aligned}$$

and therefore, $\mathbf{D}_j^\top \mathbf{C}_j^{-1} \mathbf{D}_j = \frac{\psi_j}{\mu_j^2} \boldsymbol{d}_j \boldsymbol{d}_j^\top$. Now, $\mathbf{D}^\top \mathbf{C}^{-1} = (\mathbf{D}_1^\top \mathbf{C}_1^{-1}, \ldots, \mathbf{D}_J^\top \mathbf{C}_J^{-1})$, and, hence, $\mathbf{D}^\top \mathbf{C}^{-1} \mathbf{D} = \sum_{j=1}^J \frac{\psi_j}{\mu_j^2} \boldsymbol{d}_j \boldsymbol{d}_j^\top$.

Further $\mathbf{D}_j^\top \mathbf{C}_j^{-1} \boldsymbol{\mu}_j = \frac{\psi_j}{\mu_j} \boldsymbol{d}_j$ and, hence, $\mathbf{D}^\top \mathbf{C}^{-1} \boldsymbol{\mu} = \sum_{j=1}^J \frac{\psi_j}{\mu_j} \boldsymbol{d}_j$

Combining these results we obtain

$$\mathbf{M}^Q = \sum_{j=1}^J \frac{\psi_j}{\mu_j^2} \boldsymbol{d}_j \boldsymbol{d}_j^\top - \frac{\rho\tau}{1 + \rho\tau \sum_{j=1}^J \psi_j} \left(\sum_{j=1}^J \frac{\psi_j}{\mu_j} \boldsymbol{d}_j \right) \left(\sum_{j=1}^J \frac{\psi_j}{\mu_j} \boldsymbol{d}_j \right)^\top$$

which proves the Lemma because of $\boldsymbol{d}_{0j} = \frac{1}{\mu_j} \boldsymbol{d}_j$. ■

Proof of Theorem 5.1.3.

We proof the following version of the representation of the inverse quasi-information matrix

$$\mathbf{M}^{Q^{-1}} = \mathbf{M}^{-1} + \rho\tau \mathbf{e}_1 \mathbf{e}_1^\top \tag{A.24}$$

which is also valid for approximate designs.

With $\mathbf{A} = \mathbf{M}^{-1}$ and $\mathbf{b} = \sqrt{\rho\tau}\mathbf{e}_1$ we obtain by the inversion formula for the inverse of the right hand side

$$\left(\mathbf{M}^{-1} + \rho\tau\mathbf{e}_1\mathbf{e}_1^\top\right)^{-1} = \mathbf{M} - \frac{\rho\tau}{1 + \rho\tau\mathbf{e}_1^\top\mathbf{M}\mathbf{e}_1}\mathbf{M}\mathbf{e}_1\mathbf{e}_1^\top\mathbf{M}.$$

Remember that $\mathbf{M} = \mathbf{D}_0^\top\mathbf{\Psi}\mathbf{D}_0$. Now $\mathbf{D}_0\mathbf{e}_1 = \mathbf{1}_J$ for the scaled essential design matrix $\mathbf{D}_0$. Hence, $\mathbf{M}\mathbf{e}_1 = \mathbf{D}_0^\top\mathbf{\Psi}\mathbf{1}_J$ and $\mathbf{e}_1^\top\mathbf{M}\mathbf{e}_1 = \mathbf{1}_J^\top\mathbf{\Psi}\mathbf{1}_J$.

Thus, the inverse of the right hand side of (A.24) is equal to the quasi-information matrix $\mathbf{M}^Q$ which proved the Theorem.

■

Proof of Lemma 5.1.4.

In the form of individual quasi Fisher information matrix (2.69), it is required to obtain matrices $\mathbf{D}$ and $\mathbf{V}$. As $\mu_j = \exp(\theta_j)$, matrix $\mathbf{D}$ is constructed as (2.46). Furthermore, based on matrix $\mathbf{V}$ and due to the independence of response variables between subjects and repetitions, $\mathbf{V}$ is gained as follows:

$$\mathbf{V} = \mathrm{diag}(\exp(\theta_1)\mathbf{I}_{n_1}, ..., \exp(\theta_J)\mathbf{I}_{n_J}). \tag{A.25}$$

Therefore, the inverse of the diagonal block matrix is:

$$\mathbf{V}^{-1} = \mathrm{diag}(\exp(-\boldsymbol{\theta}_1)\mathbf{I}_{n_1}, ..., \exp(-\boldsymbol{\theta}_J)\mathbf{I}_{n_J}). \tag{A.26}$$

Substitution of matrices $\mathbf{D}$ and $\mathbf{V}^{-1}$ into (2.69) leads to

$$\mathbf{M}^Q = \sum_{j=1}^{J} \boldsymbol{d}_j\boldsymbol{d}_j^\top \exp(-\boldsymbol{\theta}_j)n_j,$$

$$= \sum_{j=1}^{J} \boldsymbol{d}_{0j}\boldsymbol{d}_{0j}^\top \exp(\boldsymbol{\theta}_j)n_j, \tag{A.27}$$

Since $\boldsymbol{d}_j = \boldsymbol{d}_{0j}\boldsymbol{\mu}_j$, the last equality can be written in the form of (5.8). ■

Proof of Lemma 5.1.5.

$$F_{\mathfrak{D}^Q}(\xi, \eta) = \lim_{\epsilon \to 0}\tfrac{1}{\epsilon}[\log(\det(\mathbf{M}^{-1}((1-\epsilon)\xi + \epsilon\eta) + \rho\tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1})$$

$$- \log(\det((\mathbf{M}(\xi))^{-1} + \rho\tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1})]$$

$$= \tfrac{\mathrm{d}}{\mathrm{d}\epsilon}\log(\det(\mathbf{M}^{-1}((1-\epsilon)\xi + \epsilon\eta) + \rho\tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1})\,|_{\epsilon=0^+}$$

Since for any matrix $\boldsymbol{A}$ which is a differentiable map from real numbers to $p \times p$ matrices,

$$\frac{\mathrm{d}}{\mathrm{d}t}\det(\boldsymbol{A}(t)) = \det(\boldsymbol{A}(t)).\mathrm{tr}(\boldsymbol{A}^{-1}(t)\frac{\mathrm{d}\boldsymbol{A}(t)}{\mathrm{d}t}), \tag{A.28}$$

and

$A(\epsilon) = \mathbf{M}^{Q^{-1}}((1-\epsilon)\xi + \epsilon\eta) = \mathbf{M}^{-1}((1-\epsilon)\xi + \epsilon\eta) + \rho\tau e_1 e_1^\top,$

the following equality is

$$F_{\mathfrak{D}^Q}(\xi, \eta) = -\mathrm{tr}(\mathbf{M}^Q((1-\epsilon)\xi + \epsilon\eta)$$

$$\tfrac{\mathrm{d}}{\mathrm{d}\epsilon}[(\mathbf{M}^{-1}((1-\epsilon)\xi + \epsilon\eta) + \rho\tau\mathbf{e}_1\mathbf{e}_1^\top)] \mid_{\epsilon=0^+}$$

$$= \mathrm{tr}\left(\mathbf{M}^Q(\xi)\mathbf{M}^{-1}(\xi)\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbf{M}(\xi + \epsilon(\eta-\xi))\mathbf{M}^{-1}(\xi)\right)$$

Because $\mathbf{M}(\xi) = \mathbf{D}_0^\top\mathbf{\Psi}(\xi)\mathbf{D}_0$ and $\mathbf{\Psi}(\xi)$ is a diagonal matrix, we have

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbf{M}(\xi + \epsilon(\eta-\xi)) = \mathbf{D}_0^\top\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbf{\Psi}(\xi + \epsilon(\eta-\xi))\mathbf{D}_0$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbf{\Psi}(\xi + \epsilon(\eta-\xi)) = \mathrm{diag}\left(\frac{\mathrm{d}}{\mathrm{d}\epsilon}\psi_j(\xi + \epsilon(\eta-\xi))\right)_{j=1,\ldots,J}.$$

Finally, as $\psi_j(\xi) = \frac{n_j(\xi)\mu_j}{1+(1-\rho)\tau n_j(\xi)\mu_j}$ and

$$n_j(\xi + \epsilon(\eta-\xi)) = n_j(\xi) + \epsilon(n_j(\eta) - n_j(\xi)),$$

we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\epsilon}\psi_j(\xi + \epsilon(\eta-\xi)) &= \frac{(n_j(\eta) - n_j(\xi))\mu_j}{(1+(1-\rho)\tau n_j(\xi)\mu_j)^2} \\ &= \frac{n_j(\eta)\mu_j}{(1+(1-\rho)\tau n_j(\xi)\mu_j)^2} - \frac{1}{1+(1-\rho)\tau n_j(\xi)\mu_j}\psi_j(\xi)\end{aligned}$$

Hence,

$F_{\mathfrak{D}^Q}(\xi, \eta) =$

$$\mathrm{tr}(\mathbf{M}^Q(\xi)\mathbf{M}^{-1}(\xi)\mathbf{D}_0^\top W(\xi)\mathrm{diag}(n_1(\eta)\mu_1, ..., n_J(\eta)\mu_J)\mathbf{W}(\xi)\mathbf{D}_0\mathbf{M}^{-1}(\xi)) -$$

$$\mathrm{tr}(\mathbf{M}^Q(\xi)\mathbf{M}^{-1}(\xi)\mathbf{D}_0^\top W(\xi)\mathbf{\Psi}(\xi)\mathbf{D}_0\mathbf{M}^{-1}(\xi)),$$

which is linear in η. ■

Proof of Lemma 5.2.2

Assume $b_j = (1-\rho)\tau\mu_j$, we can rewrite $\psi_j; j = 1, ..., J$ as a function of ξ as follows:

$$\psi_j(\xi) = \tfrac{\mu_j}{b_j}(1 - \tfrac{b_j^{-1}}{b_j^{-1}+n_j}).$$

Since

$$n_j = (1-\epsilon)n_{1j} + \epsilon n_{2j}, \tag{A.29}$$

where n_{1j} is the number of replication at the jth experimental setting for design ξ_1 and n_{2j} is the number of replication at the jth experimental setting for design ξ_2.

$$b_j^{-1} + n_j = (1-\epsilon)(b_j^{-1} + n_{1j}) + \epsilon(b_j^{-1} + n_{2j}) \tag{A.30}$$

$$\frac{1}{b_j^{-1} + n_j} \leq (1-\epsilon)\frac{1}{b_j^{-1} + n_{1j}} + \epsilon\frac{1}{b_j^{-1} + n_{2j}}. \tag{A.31}$$

Therefore,

$$\psi_j(\xi) \geq (1-\epsilon)\psi_j(\xi_1) + \epsilon\psi_j(\xi_2), \tag{A.32}$$

which means $\psi_j(\xi)$ is concave in ξ. It shows that

$$\sum_{j=1}^{J} \boldsymbol{d}_{0j}\boldsymbol{d}_{0j}{}^{\top}\psi_j(\xi) \geq (1-\epsilon)\sum_{j=1}^{J} \boldsymbol{d}_{0j}\boldsymbol{d}_{0j}{}^{\top}\psi_j(\xi_1) + \epsilon\sum_{j=1}^{J} \boldsymbol{d}_{0j}\boldsymbol{d}_{0j}{}^{\top}\psi_j(\xi_2). \tag{A.33}$$

In other words,

$$\mathbf{D}_0^{\top}\boldsymbol{\Psi}(\xi)\mathbf{D}_0 \geq (1-\epsilon)\mathbf{D}_0^{\top}\boldsymbol{\Psi}(\xi_1)\mathbf{D}_0 + \epsilon\mathbf{D}_0^{\top}\boldsymbol{\Psi}(\xi_2)\mathbf{D}_0, \tag{A.34}$$

where $\boldsymbol{\Psi}(\xi) = \text{diag}(\psi_1(\xi), ..., \psi_J(\xi))$. Now, let $\mathbf{M}(\xi) = \mathbf{D}_0^{\top}\boldsymbol{\Psi}(\xi)\mathbf{D}_0$. It follows that

$$\mathbf{M}(\xi) \geq (1-\epsilon)\mathbf{M}(\xi_1) + \epsilon\mathbf{M}(\xi_2),$$

$$\mathbf{M}^{-1}(\xi) \leq ((1-\epsilon)\mathbf{M}(\xi_1) + \epsilon\mathbf{M}(\xi_2))^{-1}$$

In order to show that $\mathbf{M}^Q$ is concave, we form $\mathbf{M}^Q$ from each side of inequality.

$$\mathbf{M}^{-1}(\xi) + \rho\tau\mathbf{e}_1\mathbf{e}_1^{\top} \leq ((1-\epsilon)\mathbf{M}(\xi_1) + \epsilon\mathbf{M}(\xi_2))^{-1} + \rho\tau\mathbf{e}_1\mathbf{e}_1^{\top}$$

$$(\mathbf{M}^{-1}(\xi) + \rho\tau\mathbf{e}_1\mathbf{e}_1^{\top})^{-1} \geq (((1-\epsilon)\mathbf{M}(\xi_1) + \epsilon\mathbf{M}(\xi_2))^{-1} + \rho\tau\mathbf{e}_1\mathbf{e}_1^{\top})^{-1}$$

Matrix $\rho\tau\mathbf{e}_1\mathbf{e}_1^\top$ is not invertible; hence, we use some regularization in the right hand side of the above inequality.

$$\lim_{\nu\to 0^+}[((1-\epsilon)\mathbf{M}(\xi_1)+\epsilon\mathbf{M}(\xi_2))^{-1}+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)]^{-1} \tag{A.35}$$

$$= \lim_{\nu\to 0^+}[(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}-(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}$$

$$((1-\epsilon)\mathbf{M}(\xi_1)+\epsilon\mathbf{M}(\xi_2)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1})^{-1}$$

$$(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}]$$

$$= \lim_{\nu\to 0^+}[(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}-(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}$$

$$((1-\epsilon)(\mathbf{M}(\xi_1)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1})+$$

$$\epsilon(\mathbf{M}(\xi_2)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}))^{-1}(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}]$$

$$\geq \lim_{\nu\to 0^+}[(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}-(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}$$

$$((1-\epsilon)(\mathbf{M}(\xi_1)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1})^{-1}+$$

$$\epsilon(\mathbf{M}(\xi_2)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1})^{-1}(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}].$$

$$= \lim_{\nu\to 0^+}(1-\epsilon)(\mathbf{M}(\xi_1)^{-1}+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p))^{-1}+$$

$$\epsilon(\mathbf{M}(\xi_2)^{-1}+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p))^{-1}$$

The first and the last equality are obtained from Corollary 1.7.1 in Schott 1997, which is stated in Lemma A.9.1.

Lemma A.9.1. *(Schott 1997) Suppose that A, B and $A+B$ are all $m\times m$ non-singular matrices. Then,*

$$(A+B)^{-1} = A^{-1} - A^{-1}(B^{-1}+A^{-1})^{-1}A^{-1}$$

The third inequality is obtained from the following:

$$((1-\epsilon)(\mathbf{M}(\xi_1)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1})+\epsilon(\mathbf{M}(\xi_2)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1}))^{-1} \leq$$

$$(1-\epsilon)(\mathbf{M}(\xi_1)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1})^{-1}+\epsilon(\mathbf{M}(\xi_2)+(\rho\tau\mathbf{e}_1\mathbf{e}_1^\top+\nu\mathbf{I}_p)^{-1})^{-1},$$

which is gained from $((1-\epsilon)A+\epsilon B)^{-1} \leq (1-\epsilon)A^{-1}+\epsilon B^{-1}$, for nonsingular matrices A and B. (Fedorov and Hackl 1997, p. 107). Thus,

$$(\mathbf{M}^{-1}(\xi)+\rho\tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1} \geq (1-\epsilon)(\mathbf{M}^{-1}(\xi_1)+\rho\tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1}+\epsilon(\mathbf{M}^{-1}(\xi_2)+\rho\tau\mathbf{e}_1\mathbf{e}_1^\top)^{-1} \tag{A.36}$$

and the D-optimality criterion is concave.

■

Proof of Theorem 5.2.3.

According to Lemma 5.1.5, the directional derivative is linear in ξ, and we can restrict in the equivalence theorem to one-point designs $\eta = \xi_j$ which assign all n observations to the time point t_j, i.e. $n_j(\xi_j) = n$ and $n_{j'}(\xi_j) = 0$ for $j \neq j'$.

For these one-point designs, we observe that $\frac{n_{j'}(\xi_j)\mu_{j'}}{(1+(1-\rho)\tau n_j'(\xi)\mu_{j'})^2}$ is non-zero only for $j = j'$. Hence, the directional derivative of $\mathbf{M}(\xi)$ in the direction of ξ_j can be written as

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbf{M}(\xi+\epsilon(\xi_j-\xi)) = \frac{n\mu_j}{(1+(1-\rho)\tau n_j(\xi)\mu_j)^2}\boldsymbol{d}_{0j}\boldsymbol{d}_{0j}^\top - \mathbf{D}_0^\top\mathbf{W}(\xi)\boldsymbol{\Psi}(\xi)\mathbf{D}_0.$$

In view of the general equivalence theorem (see Silvey 1980), the design ξ^* is optimal if and only if the directional derivative

$$F_{\mathfrak{D}^Q}(\xi^*,\xi_j) =$$

$$\operatorname{tr}\left(\mathbf{M}^Q(\xi^*)\mathbf{M}^{-1}(\xi^*)\left(\frac{n\mu_j}{(1+(1-\rho)\tau n_j(\xi^*)\mu_j)^2}\boldsymbol{d}_{0j}\boldsymbol{d}_{0j}^\top - \mathbf{D}_0^\top\mathbf{W}(\xi^*)\boldsymbol{\Psi}(\xi^*)\mathbf{D}_0\right)\mathbf{M}^{-1}(\xi^*)\right) \tag{A.37}$$

is less or equal to zero for all $j = 1,\dots,J$. Rewriting this condition completes the proof of the Theorem.

■

Proof of Lemma 5.2.4.

By taking the determinant of $\mathbf{M}^Q(\xi)$ in (5.7), and using matrix determinant lemma (Harville 1997), we can obtain

$$\det(\mathbf{M}^Q(\xi)) = \det(\mathbf{M}(\xi))(1+\rho\tau\mathbf{e}_1^\top\mathbf{M}(\xi)\mathbf{e}_1)^{-1}.$$

Since

$$\mathbf{e}_1^\top\mathbf{M}(\xi)\mathbf{e}_1 = \mathbf{1}_J^\top\boldsymbol{\Psi}(\xi)\mathbf{1}_J,$$

in the case that $\mathbf{D}_0\mathbf{e}_1 = \mathbf{1}_n$, (5.11) will be obtained. ■

A.10 Corollary 1.7.2

Schott 1997. Let $\mathbf{A}$ be an $m \times m$ nonsingular matrix. If $\mathbf{b}$ and $\mathbf{c}$ are both $m \times 1$ vectors and $(\mathbf{A} + \mathbf{b}\mathbf{c}^\top)$ is nonsingular; then,

$$(\mathbf{A} + \mathbf{b}\mathbf{c}^\top)^{-1} = \mathbf{A}^{-1} - \frac{1}{1 + \mathbf{c}^\top \mathbf{A}^{-1}\mathbf{b}} \mathbf{A}^{-1}\mathbf{b}\mathbf{c}^\top \mathbf{A}^{-1}.$$

A.11 D-optimum design in Nonlinear longitudinal Poisson model

In this section, the D-optimum design with four and seven support points for model (2.33) are obtained. The results are shown in frame of tables and Figures.

Table A.1: Locally D-optimum designs for $J = 4$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $\rho = 0.9$, $\tau = 1$, $n = 120$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.478	0.501	0.513	0.520	0.523	0.526
w_2	0.234	0.223	0.214	0.207	0.203	0.200
w_3	0.123	0.118	0.122	0.127	0.130	0.133
w_4	0.166	0.157	0.151	0.146	0.143	0.141

Table A.2: Locally D-optimum designs for $J = 4$ periods, $\beta_0 = 6$,$\beta_1 = 2$, $\rho = 0.9$, $\tau = 1$, $n = 120$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.478	0.501	0.513	0.520	0.524	0.526
w_2	0.235	0.224	0.215	0.208	0.203	0.200
w_3	0.123	0.119	0.122	0.127	0.130	0.133
w_4	0.165	0.156	0.150	0.146	0.143	0.141

Table A.3: Locally D-optimum designs for $J = 7$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $\rho = 0.9$, $\tau = 1$, $n = 120$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.410	0.431	0.442	0.448	0.452	0.453
w_2	0.179	0.185	0.182	0.178	0.175	0.173
w_3	0.097	0.077	0.069	0.068	0.070	0.071
w_4	0.068	0.070	0.074	0.075	0.075	0.075
w_5	0.074	0.077	0.077	0.077	0.076	0.076
w_6	0.083	0.080	0.078	0.077	0.076	0.076
w_7	0.088	0.081	0.078	0.077	0.076	0.076

Table A.4: Locally D-optimum designs for $J = 7$ periods, $\beta_0 = 6$,$\beta_1 = 2$, $\rho = 0.9$, $\tau = 1$, $n = 120$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.411	0.431	0.443	0.448	0.451	0.453
w_2	0.179	0.184	0.182	0.178	0.175	0.173
w_3	0.097	0.077	0.069	0.069	0.070	0.072
w_4	0.068	0.070	0.074	0.075	0.075	0.075
w_5	0.074	0.077	0.077	0.076	0.076	0.075
w_6	0.083	0.080	0.078	0.077	0.076	0.076
w_7	0.088	0.081	0.078	0.077	0.076	0.076

Table A.5: Locally D-optimum designs for $J = 4$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $n = 120$, $\rho = 0$, $\tau = 10$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.479	0.503	0.511	0.517	0.521	0.523
w_2	0.232	0.221	0.204	0.197	0.193	0.191
w_3	0.123	0.119	0.139	0.145	0.147	0.148
w_4	0.165	0.157	0.146	0.142	0.140	0.138

Table A.6: Locally D-optimum designs for $J = 7$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $n = 120$, $\rho = 0$, $\tau = 10$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.416	0.439	0.450	0.456	0.459	0.461
w_2	0.172	0.176	0.174	0.171	0.168	0.166
w_3	0.097	0.076	0.067	0.066	0.068	0.070
w_4	0.067	0.068	0.073	0.074	0.075	0.074
w_5	0.074	0.077	0.077	0.077	0.076	0.075
w_6	0.082	0.079	0.077	0.076	0.075	0.075
w_7	0.091	0.085	0.082	0.080	0.080	0.079

Table A.7: Locally D-optimum designs for $J = 4$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $n = 120$, $\rho = 1$,$\tau = 10/9$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.327	0.343	0.352	0.358	0.361	0.363
w_2	0.174	0.158	0.149	0.143	0.140	0.138
w_3	0.048	0.000	0.000	0.000	0.000	0.000
w_4	0.498	0.499	0.499	0.499	0.499	0.499

Table A.8: Locally D-optimum designs for $J = 7$ periods, $\beta_0 = 3$,$\beta_1 = 2$, $n = 120$, $\rho = 1$, $\tau = 10/9$

β_2	1	1.5	2	2.5	3	3.5
w_1	0.396	0.417	0.426	0.429	0.431	0.432
w_2	0.199	0.215	0.209	0.202	0.198	0.195
w_3	0.094	0.048	0.042	0.050	0.058	0.064
w_4	0.047	0.060	0.072	0.076	0.077	0.077
w_5	0.066	0.080	0.082	0.081	0.079	0.077
w_6	0.091	0.088	0.084	0.081	0.079	0.077
w_7	0.106	0.091	0.085	0.081	0.079	0.077

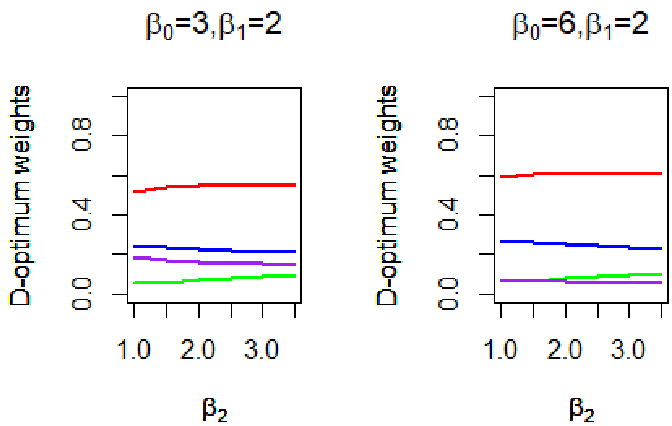

Figure A.1: Local D-optimum designs for $J = 4$ periods, $\rho = 0.9$, $\tau = 1$, $t_j = j - 1$, red line: $w_1^\star$, blue line: $w_2^\star$, green line: $w_3^\star$, purple line: $w_4^\star$

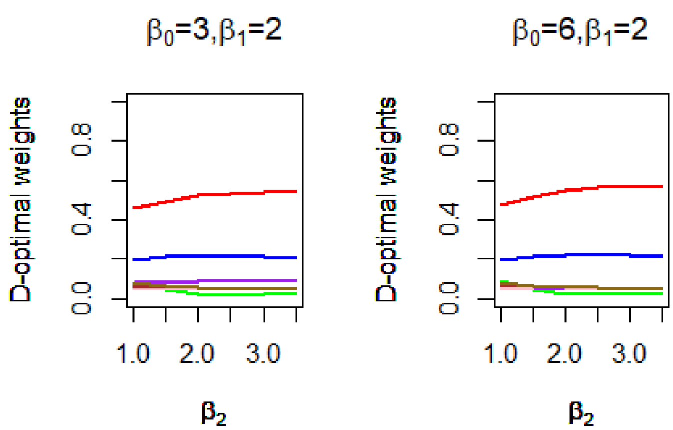

Figure A.2: Local D-optimum designs for $J = 7$ periods, $\rho = 0.9$, $\tau = 1$, $t_j = j - 1$, red line: $w_1^\star$, blue line: $w_2^\star$, green line: $w_3^\star$, purple line: $w_4^\star$, pink line: $w_5^\star$, brown line: $w_6^\star$, gold line: $w_7^\star$

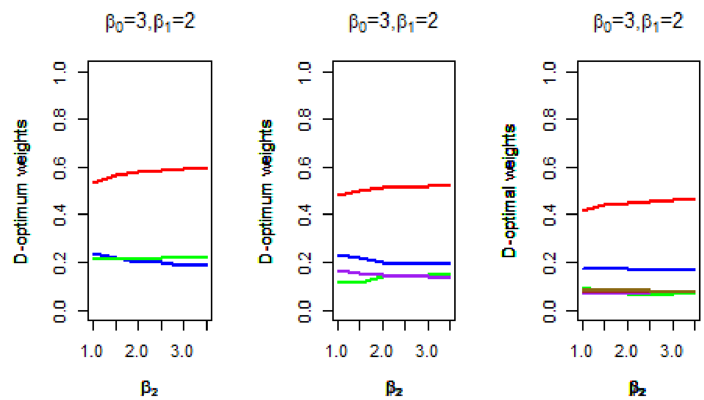

Figure A.3: Local D-optimum designs for $J = 3, 4, 7$ periods, $\rho = 0$, $\tau = 10$, $t_j = j - 1$, red line: $w_1^\star$, blue line: $w_2^\star$, green line: $w_3^\star$, purple line: $w_4^\star$, pink line: $w_5^\star$, brown line: $w_6^\star$, gold line: $w_7^\star$

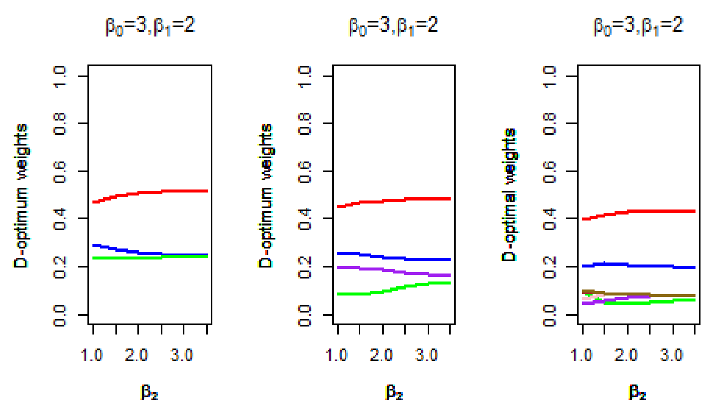

Figure A.4: Locally D-optimum designs for $J = 3, 4, 7$ periods, $\rho = 1$, $\tau = 10/9$, $t_j = j - 1$, red line: $w_1^\star$, blue line: $w_2^\star$, green line: $w_3^\star$, purple line: $w_4^\star$, pink line: $w_5^\star$, brown line: $w_6^\star$, gold line: $w_7^\star$

List of Symbols

$\boldsymbol{\alpha}$	Parameter vector characterizing $\boldsymbol{R}(\boldsymbol{\alpha})$, 30
$\boldsymbol{\beta}$	Population parameter with p dimension, 16
$\boldsymbol{\Lambda}_i$	Vector of random ability parameter for individual i, 25
$\mathbf{D}_i$	Matrix of the derivative of the ith response mean, 30
$\mathbf{M}^Q(\boldsymbol{\beta})$	Population quasi Fisher information matrix, 37
$\mathbf{M}_i^Q(\boldsymbol{\beta})$	ith individual quasi Fisher information matrix, 37
$\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\alpha})$	Population score function, 30
$\mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$	ith individual score function, 31
$\mathbf{V}_G$	Variance of $\hat{\beta}_G$, 32
$\mathbf{V}_I$	Variance of $\hat{\beta}_I$, 30
$\mathbf{V}_i$	Marginal variance of the ith individual response variable, 30
$\mathbf{V}_{jj'}$	Marginal variance matrix of $\boldsymbol{Y}_i$ at $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{ij'}$, 43
$\boldsymbol{\mu}_i$	Mean vector $\mathrm{E}(\boldsymbol{Y}_i)$ of $\boldsymbol{Y}_i$, 31
$\boldsymbol{\pi}_{ij}$	Marginal expectation of the response variable in binary and ordinal mixed effects model, 24
$\boldsymbol{\Sigma}$	Dispersion parameter of the ith random effect, 17
$\boldsymbol{\theta}(\boldsymbol{t}, \boldsymbol{\beta})$	Vector of ability parameter, 24
$\boldsymbol{\zeta}_i$	Random effect with q dimension for individual i, 16

$\boldsymbol{f}(\boldsymbol{x})$	Vector of regression functions of experimental setting $\boldsymbol{x}$ regarding population parameter $\boldsymbol{f}(\boldsymbol{x}) = (1, f_1(\boldsymbol{x}), ..., f_{p-1}(\boldsymbol{x}))^\top$, 16
$\boldsymbol{h}(\boldsymbol{x})$	Vector of regression functions of experimental setting $\boldsymbol{x}$ regarding random effect $\boldsymbol{h}(\boldsymbol{x}) = (1, h_1(\boldsymbol{x}), ..., h_{q-1}(\boldsymbol{x}))^\top$, 16
$\boldsymbol{R}(\boldsymbol{\alpha})$	Correlation matrix, 31
$\boldsymbol{x}$	Experimental setting in general, 16
$\boldsymbol{x}_{ij}$	jth experimental setting at the ith individual, $j = 1, ..., J$, $i = 1, ..., N$, 16
$\boldsymbol{Y}$	Vector of response variable of all observations with n dimension, 26
$\boldsymbol{y}$	Observational vector of response variable with n dimension, 20
$\boldsymbol{Y}_{ijk}$	Vector of response variable of the kth replication for individual i at setting x_{ij} (ordinal case), 21
$\boldsymbol{Y}_{ij}$	Vector of response variables for individual i at setting x_{ij}, 29
$\boldsymbol{y}_{ij}$	Observational vector of response variables for individual i at setting x_{ij}, 17
$\boldsymbol{Y}_i$	ith vector of response variable, 26
$\boldsymbol{y}_i$	Observational ith vector of response variable with n_i dimension, 20
$\delta_{ij}^{(m)}(\boldsymbol{\zeta}_i)$	Individual standardized thresholds for individual i at setting $\boldsymbol{x}_{ij}$ (ordinal case), 22
ϵ_{ijk}	Random error of latent variable, 17
η_{ij}	Individual linear component for individual i at setting x_{ij}, 16
γ_m	Threshold regarding the mth level of the response variable, 21
$\hat{\boldsymbol{\beta}}_G$	Generalized least square estimate of population parameter, 32
$\hat{\boldsymbol{\beta}}_I$	Least square estimate of the population parameter, 29
$\hat{\boldsymbol{\beta}}_{QL}$	Quasi maximum-likelihood estimate of population parameter, 34

Λ_{ij}	Random ability parameter for individual i at setting t_j, 25
λ_{ij}	realization of random ability parameter for individual i at setting t_j, 27
λ_i	Realization of the random ability regarding ith individual, 25
$\mathbf{e}_1$	p-dimensional unit vector with first entry equal to 1, 85
$\mathbf{i}_{\boldsymbol{\beta}}$	Fisher information matrix of the linear model, 35
$\mathfrak{B}$	Parameter space of all $\boldsymbol{\beta}$, 16
$\mathfrak{D}^Q(\xi^{(d)}, \boldsymbol{\beta})$	Population D-optimality criterion, 38
$\mathfrak{D}^Q(\xi_i^{(d)}, \boldsymbol{\beta})$	Individual D-optimality criterion, 38
$\mathfrak{P}_i$	Weight regarding ith individual approximate design ξ_i, 39
$\text{eff}_D(\xi)$	D-efficiency criterion, 41
$\text{f}(\boldsymbol{\zeta}_i)$	Multivariate density function of the random effect, 20
$\text{f}(Y_{ijk} \mid \boldsymbol{\zeta}_i)$	Conditional density of response variable conditioned on random effect, 20
μ	Any potential mean value of Y_{ijk}, 17
μ_{ij}	Mean vector of the response variable $\boldsymbol{Y}_{ij}$, 31
μ_j	Marginal response mean of the response variable regarding jth time point, 26
$\Phi(.)$	Cumulative distribution function of the standard Normal distribution, 17
ϕ	Additional parameter, 29
$\Phi^\star(.)$	Modified Lin's approximate cumulative standard Normal distribution, 46
$\Phi^{Lin}(.)$	Lin's approximate cumulative standard Normal distribution, 45
$\pi_{ij}^{(m)}$	Marginal expectation of the response variable at the mth level, 24

ρ	Intra-class correlation, 26
σ^2	Variance of one element of the ith random effect, 17
σ_ϵ^2	Variance of error ϵ_{ijk}, 21
τ	Scale parameter, 26
$\theta(t_j, \boldsymbol{\beta})$	Ability parameter at the jth time point, 25
ξ	Approximate population design, 39
$\xi^{(d)}$	Discrete population design design, 38
$\xi^{\star(d)}$	Discrete D-optimum design, 38
$\xi_i^{(d)}$	ith individual discrete design, 37
b_{ij}	Suitably chosen function of individual linear predictor, 29
c_0, c	Shape parameters, 26
$F_{\mathfrak{D}^Q}(\xi_1, \xi_2)$	Directional derivative from ξ_1 in the direction of ξ_2, 40
$g(.)$	Link function, 17
i	Index for the individual, 16
J	The number of experimental settings or time points, 16
j	Index for experimental setting, 16
k	Index for the replication, 16
$L(\boldsymbol{\beta} \mid \boldsymbol{y})$	Population marginal likelihood function, 20
$L(\boldsymbol{\beta} \mid \boldsymbol{y}_i)$	Individual marginal likelihood function, 20
M	The total number of the level of response variable, 21
m	index for the thresholds of the response variable, 21
N	The total number of individuals, 16
n	The total number of observations, 38
n_{ij}	The number of replication within each experimental setting or time point, 16

n_i	The number of observations within each individual, 16
p	Dimension of the fixed effects parameter $\boldsymbol{\beta}$, 15
$p_{ij}(\boldsymbol{\zeta}_i)$	Probability of success in binary mixed effects model conditioned on the random effect, 17
$p_{ij}^{(m)}(\boldsymbol{\zeta}_i)$	Probability of success of the mth level of the response variable conditioned on the random effect, 21
q	Dimension of the random effect $\boldsymbol{\zeta}_i$, 16
$ql(.,.)$	Quasi likelihood function, 33
S_{i0}	Individual block effect being constant over time (Nonlinear longitudinal Poisson regression model), 25
S_{ij}	Random effect within the ith individual related to the jth time point (Nonlinear longitudinal Poisson regression model), 25
t_{ij}	jth time point at the ith individual, 16
U_{ijk}	Latent variable, 21
w_{ij}	Weight regarding $\boldsymbol{x}_{ij}$, 39
Y_{ijk}	Response variable for the ith individual at experimental setting $\boldsymbol{x}_{ij}$ (or time point t_{ij}) for the kth replication, 16
Y_{ijk}	Response variable of the kth replication for individual i at setting x_{ij}, 16
y_{ijk}	Observations of the kth replication for individual i at setting x_{ij}, 17
$Y_{ijk}^{(m)}$	Response variable at the mth level (ordinal case), 21
$Y_{ijk}^{(m)}$	Response variable regarding the mth level, 21

Bibliography

Abebe, Haftom T et al. (2014). “Robustness of Bayesian D-optimal design for the logistic mixed model against misspecification of autocorrelation”. In: *Computational Statistics* 29.6, pp. 1667–1690.

Agresti, Alan (2003). *Categorical data analysis*. Vol. 482. John Wiley & Sons.

Agresti, Alan and Joseph B Lang (1993). “A proportional odds model with subject-specific effects for repeated ordered categorical responses”. In: *Biometrika* 80.3, pp. 527–534.

Andrews, Donald WK (1987). “Consistency in nonlinear econometric models: A generic uniform law of large numbers”. In: *Econometrica: Journal of the Econometric Society* 55.6, pp. 1465–1471.

Atkinson, Anthony et al. (2007). *Optimum experimental designs, with SAS*. Vol. 34. Oxford University Press.

Bono, Roser, Rafael Alarcón, and Maria J Blanca (2021). “Report Quality of Generalized Linear Mixed Models in Psychology: A Systematic Review”. In: *Frontiers in psychology* 12, p. 1345.

Booth, James G and James P Hobert (1999). “Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.1, pp. 265–285.

Bradley (1973). “The equivalence of maximum likelihood and weighted least squares estimates in the exponential family”. In: *Journal of the American Statistical Association* 68.341, pp. 199–200.

Bradley and John J Gart (1962). “The asymptotic properties of ML estimators when sampling from associated populations”. In: *Biometrika* 49.1/2, pp. 205–214.

Breslow, Norman E and David G Clayton (1993). “Approximate inference in generalized linear mixed models”. In: *Journal of the American statistical Association* 88.421, pp. 9–25.

Butler, Steven M and Thomas A Louis (1997). "Consistency of maximum likelihood estimators in general random effects models for binary data". In: *The Annals of Statistics* 25.1, pp. 351–377.

Cramér, Harald (1946). *Mathematical Methods of Statistics*. Vol. 9. Princeton university press.

Fedorov and Peter Hackl (1997). *Model-oriented design of experiments*. Vol. 125. Springer Science & Business Media.

Fedorov, Valerii Vadimovich (1972). *Theory of optimal experiments*. Academic Press, New York.

Goldstein, Harvey (1991). "Nonlinear multilevel models, with an application to discrete response data". In: *Biometrika* 78.1, pp. 45–51.

Green, Peter J (1987). "Penalized likelihood for general semi-parametric regression models". In: *International Statistical Review/Revue Internationale de Statistique*, pp. 245–259.

Harville, David A. (1997). *Matrix Algebra From a Statistician Perspective*. Springer, New York.

Hoadley, Bruce (1971). "Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case". In: *The Annals of mathematical statistics* 42.6, pp. 1977–1991.

Jefferson, Hakeem (2019). "Policing Norms: Punishment and the Politics of Respectability among Black Americans". PhD thesis. University of Michigan.

Jiang, Jiming (1998). "Consistent estimators in generalized linear mixed models". In: *Journal of the American Statistical Association* 93.442, pp. 720–729.

Jiang, Jiming and Weihong Zhang (2001). "Robust estimation in generalised linear mixed models". In: *Biometrika* 88.3, pp. 753–765.

Jørgensen, Bent (1983). "Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models". In: *Biometrika* 70.1, pp. 19–28.

Liang, Kung-Yee and Scott L Zeger (1986). "Longitudinal data analysis using generalized linear models". In: *Biometrika* 73.1, pp. 13–22.

Lin, Jinn-Tyan (1989). "Approximating the normal tail probability and its inverse for use on a pocket calculator". In: *Applied Statistics* 38.1, pp. 69–70.

Mathai, Arakaparampil M and Rajendra Kumar Saxena (1978). *The H function with applications in statistics and other disciplines*. Wiley Eastern, New Delhi and Wiley Halsted, New York.

Mathai and Moschopoulos (1991). “On a multivariate gamma”. In: *Journal of Multivariate Analysis* 39.1, pp. 135–153.

McCullagh (1983). “Quasi-likelihood functions”. In: *The Annals of Statistics* 11.1, pp. 59–67.

McCullagh, P and J. A. Nelder (1983). *Generalized Linear Models*. Chapman and Hall.

Mielke, Tobias and Rainer Schwabe (2010). “Some considerations on the Fisher information in nonlinear mixed effects models”. In: *mODa* 9, pp. 129–136.

Niaparast, Mehrdad (2010). “Optimal designs for mixed effects poisson regression models”. PhD thesis. Magdeburg, Univ., Diss., 2010.

Niaparast, Mehrdad and Rainer Schwabe (2013). “Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients”. In: *Journal of Statistical Planning and Inference* 143.2, pp. 296–306.

Nie, Lei (2006). “Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models”. In: *Metrika* 63.2, pp. 123–143.

Puntanen, Simo (2007). “Matrix Algebra: Theory, Computations, and Applications in Statistics by James E. Gentle”. In: *International Statistical Review* 75.3, pp. 435–435.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Robinson, George K (1991). “That BLUP is a good thing: the estimation of random effects”. In: *Statistical science* 6.1, pp. 15–32.

Schmelter, Thomas (2007). “The optimality of single-group designs for certain mixed models”. In: *Metrika* 65.2, pp. 183–193.

Schott, James R (1997). *Matrix analysis for statistics*. Wiley, New York.

Seurat, Jérémy, Thu Thuy Nguyen, and France Mentré (2020). “Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes”. In: *Statistical methods in medical research* 29.3, pp. 934–952.

Sheiner, Lewis B and Stuart L Beal (1983). “Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine clinical pharmacokinetic data”. In: *Journal of pharmacokinetics and biopharmaceutics* 11.3, pp. 303–319.

Silvey, Samuel (1980). *Optimal design: an introduction to the theory for parameter estimation*. Vol. 1. Springer Science & Business Media.

Sutradhar, Brajendra C (2004). “On exact quasilikelihood inference in generalized linear mixed models”. In: *Sankhyā: The Indian Journal of Statistics* 66.2, pp. 263–291.

Tekle, Fetene B, Frans ES Tan, and Martijn PF Berger (2008). “Maximin D-optimal designs for binary longitudinal responses”. In: *Computational Statistics & Data Analysis* 52.12, pp. 5253–5262.

Thode, Henry C (2002). *Testing for normality.* CRC press.

Tommasi, Chiara, Juan M Rodriguez-Diaz, and M Teresa Santos-Martin (2014). “Integral approximations for computing optimum designs in random effects logistic regression models”. In: *Computational Statistics & Data Analysis* 71, pp. 1208–1220.

Tutz, Gerhard and Wolfgang Hennevogl (1996). “Random effects in ordinal regression models”. In: *Computational Statistics & Data Analysis* 22.5, pp. 537–557.

Wald, Abraham (1949). “Note on the consistency of the maximum likelihood estimate”. In: *The Annals of Mathematical Statistics* 20.4, pp. 595–601.

Wedderburn, Robert WM (1974). “Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method”. In: *Biometrika* 61.3, pp. 439–447.

Winship, Christopher and Robert D Mare (1984). “Regression models with ordinal variables”. In: *American sociological review* 49, pp. 512–525.

Ye, Yinyu (1987). “Interior algorithms for linear, quadratic, and linearly constrained non-linear programming”. PhD thesis. Department of ESS, Stanford University.

Zeger, Scott L and M Rezaul Karim (1991). “Generalized linear models with random effects; a Gibbs sampling approach”. In: *Journal of the American statistical association* 86.413, pp. 79–86.

Zeger, Scott L, Kung-Yee Liang, and Paul S Albert (1988). “Models for longitudinal data: a generalized estimating equation approach”. In: *Biometrics* 44.4, pp. 1049–1060.