



HAL
open science

On optimal design for a poisson regression model with random intercept

Mehrdad Niaparast

► **To cite this version:**

Mehrdad Niaparast. On optimal design for a poisson regression model with random intercept. *Statistics and Probability Letters*, 2009, 79 (6), pp.741. 10.1016/j.spl.2008.10.035 . hal-00511174

HAL Id: hal-00511174

<https://hal.science/hal-00511174>

Submitted on 24 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

On optimal design for a poisson regression model with random intercept

Mehrdad Niaparast

PII: S0167-7152(08)00508-7
DOI: [10.1016/j.spl.2008.10.035](https://doi.org/10.1016/j.spl.2008.10.035)
Reference: STAPRO 5260

To appear in: *Statistics and Probability Letters*

Received date: 24 April 2008
Revised date: 28 September 2008
Accepted date: 17 October 2008

Please cite this article as: Niaparast, M., On optimal design for a poisson regression model with random intercept. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.10.035

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



ON OPTIMAL DESIGN FOR A POISSON REGRESSION MODEL
WITH RANDOM INTERCEPT

MEHRDAD NIAPARAST

Institute for Mathematical Stochastics, Otto-von-Guericke University,
PSF 4120, 39016, Magdeburg, Germany

ABSTRACT. Most of the research on optimal designs concentrates on linear and non-linear models with fixed effects. In this paper we discuss optimal designs for a Poisson regression model with random intercept. It is shown that the optimal designs are identical across the individuals, but depend on the variance.

keywords: Poisson Regression Models; Optimal Designs; Quasi-likelihood ; Information Matrix

1. INTRODUCTION

Design optimality has received growing attention by statisticians over the past few decades. The main goal is to find the best experimental settings x_i , which maximize the information matrix of parameters as the inverse of variance-covariance matrix. Both binary and count data have been extensively studied in the literature as cases of Generalized Linear Models. Optimal designs for binary data models, especially for logistic models have been investigated by Abdelbasit and Plackett (1983), Minkin (1987) and many others in the fixed effects cases. Recently, Mairo et al. (2004) have extended the work on optimal designs for logistic models with random intercept. Also, Minkin (1993) and Yanping et al. (2006) have done some extensive work to find optimal designs especially D-optimal designs in fixed effects Poisson regression models.

Despite wide theoretical work on the analysis in Generalized Linear Mixed Models (GLMM) (see e.g. McCulloch and Searle (2001)), few results exist on this topic in optimal designs.

In this paper, we consider as a simple case of GLMM, a simple Poisson Regression with Random Intercept. The aim is to derive the information matrix for design optimal criteria and to obtain optimal designs.

Due to the random effect in this model, a closed form of the likelihood function to estimate the fixed parameters of this model is quite intractable. So a Quasi-Likelihood approach is used to determine the information matrix for these parameters (McCullagh and Nelder (1997), chapter 9). The only assumptions on the data are those concerning the first two moments.

As we will see later, the information matrix depends on the unknown parameters. So, it poses a two fold problem: to find the optimal designs we must know the parameters, and to know the parameters we need to design first. A simple approach to this problem is to look for **Locally optimal** designs which are based on an initial guess of the parameters and then find optimal designs which are optimal with respect to this initial guess.

The outline of the paper is as follows: In the next section, we introduce the model and the variance-covariance structure. In section 3, we apply the quasi-likelihood

method to obtain the information matrix. And finally, in section 4, we consider a simple example and determine the D-optimal designs.

2. THE MODEL

We consider a Poisson Regression Model with random intercept which can be written as:

$$Y_{ijk} | b_i \stackrel{ind}{\sim} P(\mu_{ij}(b_i)) \begin{cases} i = 1, \dots, s \\ j = 1, \dots, t_i \\ k = 1, \dots, m_{ij} \end{cases} \begin{cases} \sum_{j=1}^{t_i} m_{ij} = m_i \\ n = \sum_{i=1}^s m_i \end{cases}$$

where $\mu_{ij}(b_i) = \exp(b_i + \mathbf{f}^T(x_{ij})\boldsymbol{\beta})$ is specified by the canonical linked function.

Here, Y_{ijk} stands for the k th replication for the individual i at the experimental setting x_{ij} from the experimental region \mathfrak{X} . Also we suppose that m_{ij} denotes the number of replications of individual i at the j th level of x . The vector of known regression functions $\mathbf{f} = (1, f_1, \dots, f_{p-1})^T$ is the same for all individuals. The p -dimensional vector of parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^T$ associated with the mean response curve is unknown. The random variable b_i is the individual deviation from the overall population intercept β_0 , i.e. $\beta_0 + b_i$ is the random intercept which varies and depends on the different individuals. The deviation b_i is assumed to be normal distributed with mean 0 and known variance σ^2 . The random intercepts are uncorrelated for different individuals, i.e., $cov(b_i, b_{i'}) = 0$ for all $i \neq i'$. Note that $Var(Y_{ijk} | b_i) = E(Y_{ijk} | b_i) = \mu_{ij}(b_i)$. For further simplification, without loss of generality, we assume $t_i = t \forall i$.

With regard to the properties of b_i , observations from the same individuals have the following variance-covariance structure:

$$\begin{aligned} Var(Y_{ijk}) &= Var(E(Y_{ijk} | b_i)) + E(Var(Y_{ijk} | b_i)) \\ &= e^{2\mathbf{f}^T(x_{ij})\boldsymbol{\beta} + \sigma^2} (e^{\sigma^2} - 1) + e^{\mathbf{f}^T(x_{ij})\boldsymbol{\beta} + \frac{1}{2}\sigma^2} \\ &= \mu_{ij}^2 (e^{\sigma^2} - 1) + \mu_{ij}, \end{aligned}$$

where $\mu_{ij} = E(Y_{ijk}) = E(E(Y_{ijk} | b_i)) = e^{\mathbf{f}^T(x_{ij})\boldsymbol{\beta} + \frac{1}{2}\sigma^2}$.

$$\begin{aligned} cov(Y_{ijk}, Y_{ij'k'}) &= cov(E(Y_{ijk} | b_i), E(Y_{ij'k'} | b_i)) + E(cov(Y_{ijk}, Y_{ij'k'} | b_i)) \\ &= \mu_{ij}\mu_{ij'} (e^{\sigma^2} - 1) \text{ for all } (j, k) \neq (j', k') \end{aligned}$$

Also, $cov(Y_{ijk}, Y_{i'j'k'}) = 0$ for all $i \neq i'$ and all j, j', k, k' .

Let $Y_i = (Y_{i1}^T, \dots, Y_{it_i}^T)^T$ be the $m_i \times 1$ vector of measurements on the i th individual, where $Y_{ij} = (Y_{ij1}, \dots, Y_{ijm_{ij}})^T$ ($j = 1, \dots, t_i$) is the $m_{ij} \times 1$ vector of replications of individual i at the j th level of x . Then by applying some matrix algebra, we have

$$\begin{aligned} V_i = var(Y_i) &= \begin{pmatrix} \mu_{i1}I_{m_{i1}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mu_{i2}I_{m_{i2}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mu_{it_i}I_{m_{it_i}} \end{pmatrix} \\ &+ (e^{\sigma^2} - 1) \begin{pmatrix} \mu_{i1}\mathbf{1}_{m_{i1}} \\ \mu_{i2}\mathbf{1}_{m_{i2}} \\ \vdots \\ \mu_{it_i}\mathbf{1}_{m_{it_i}} \end{pmatrix} \begin{pmatrix} \mu_{i1}\mathbf{1}_{m_{i1}}^T & \mu_{i2}\mathbf{1}_{m_{i2}}^T & \cdots & \mu_{it_i}\mathbf{1}_{m_{it_i}}^T \end{pmatrix} \\ &= \dot{A}_i + \dot{a}_i\dot{a}_i^T \end{aligned}$$

where, $\dot{a}_i^T = \sqrt{e^{\sigma^2} - 1} \begin{pmatrix} \mu_{i1}\mathbf{1}_{m_{i1}}^T & \cdots & \mu_{it_i}\mathbf{1}_{m_{it_i}}^T \end{pmatrix}$ and $\dot{A}_i = diag(\mu_{i1}I_{m_{i1}}, \dots, \mu_{it_i}I_{m_{it_i}})$. Here and through out I_v denotes the $v \times v$ identity matrix and $\mathbf{1}_v$ is an $v \times 1$ vector with all entries equal to 1.

We suppose that $Y^T = (Y_1^T, \dots, Y_s^T)$ is the vector of the whole observation. Independence of different individuals leads to,

$$V = var(Y) = \begin{pmatrix} V_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & V_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & V_s \end{pmatrix}$$

3. INFORMATION MATRIX

The role played by the information matrix is very clear in the Optimal Design studies. This role seems apparent comes from the relation between information matrix and variance-covariance matrix of the estimator of parameters based on the likelihood principle.

In our case due to the complicated form of the likelihood function, it is not easy to use this relation.

As an alternative we first outline the quasi-likelihood estimator as an estimator based on the quasi-score function.

3.1. Quasi-Likelihood Estimator. Suppose the $n \times 1$ random variable Y has mean $\boldsymbol{\mu}(\boldsymbol{\beta})$ and variance-covariance matrix $\phi^2 V(\boldsymbol{\beta}) = \phi^2 V(\boldsymbol{\mu}(\boldsymbol{\beta}))$ where $V(\boldsymbol{\mu})$ is called the variance function and shows the relation between the mean and the variance of \mathbf{Y} (McCulloch and Searle 2001). Both are known functions of the p -dimensional parameter vector $\boldsymbol{\beta}$ and $V(\boldsymbol{\beta})$ is a positive definite matrix.

The quasi-likelihood for the regression parameters $\boldsymbol{\beta}$ is defined by the quasi-score function

$$U(\boldsymbol{\beta}, \mathbf{y}) = \phi^2 D^T (V(\boldsymbol{\beta}))^{-1} (\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}))$$

In this expression the entries of the matrix D , of order $n \times p$, are $D_{jr} = \frac{\partial \mu_j}{\partial \beta_r}$, the partial derivatives of the components of $\boldsymbol{\mu}(\boldsymbol{\beta})$, with respect to the parameters.

The asymptotic variance-covariance of the quasi-score function $U(\boldsymbol{\beta}, \mathbf{Y})$, which equals the negative of the expectation of $\partial U(\boldsymbol{\beta}, \mathbf{Y}) / \partial \boldsymbol{\beta}$, is

$$M(\boldsymbol{\beta}) = D^T (V(\boldsymbol{\beta}))^{-1} D$$

This matrix plays the role of the Fisher information exactly in the same way as in fully parametric inference, and under the usual regularity conditions, the asymptotic variance-covariance matrix of the quasi-likelihood estimator of $\boldsymbol{\beta}$ equals $M(\boldsymbol{\beta})^{-1}$ (McCullagh and Nelder 1998, Sec.9.3). For the sake of clarity we call $M(\boldsymbol{\beta})$ the (quasi-)information matrix.

3.2. Information Matrix of an Individual Design. With regard to the definition of \mathbf{Y}_i and the quasi-score function, we can write the individual information matrix associated with the observations Y_i of a single individual i for our model as,

$$M_i(\boldsymbol{\beta}) = D_i^T V_i^{-1} D_i$$

where $D_i = \dot{A}_i \dot{F}_i$ with $\dot{F}_i = \begin{pmatrix} \mathbf{f}^T(x_{i1}) \\ \vdots \\ \mathbf{f}^T(x_{i1}) \\ \vdots \\ \mathbf{f}^T(x_{it_i}) \\ \vdots \\ \mathbf{f}^T(x_{it_i}) \end{pmatrix}_{m_i \times p}$ is the individual design matrix, and $V_i(\boldsymbol{\beta}) = V_i$. Then,

$$(3.1) \quad M_i(\boldsymbol{\beta}) = D_i^T (\dot{A}_i + \dot{a}_i \dot{a}_i^T)^{-1} D_i$$

Note that the information matrix depends on the number of measurement m_i taken from individual i in a non-linear relation. So we can not find a normalized version depending only on the proportion of the observations at the different settings, as one might obtain in fixed effect models. Also, from (3.1), the information matrix is strongly dependent on the parameters.

Lemma 3.1. *The individual information matrix (3.1) can be represented as*

$$(3.2) \quad M_i(\boldsymbol{\beta}) = \dot{F}_i^T (\dot{A}_i^{-1} + (e^{\sigma^2} - 1) \mathbf{1}_{m_i} \mathbf{1}_{m_i}^T)^{-1} \dot{F}_i$$

Proof.

$$M_i(\boldsymbol{\beta}) = D_i^T (\dot{A}_i + \dot{a}_i \dot{a}_i^T)^{-1} D_i = D_i^T \dot{A}_i^{-1} (\dot{A}_i^{-1} + \dot{A}_i^{-1} \dot{a}_i \dot{a}_i^T \dot{A}_i^{-1})^{-1} \dot{A}_i^{-1} D_i$$

Since $D_i^T \dot{A}_i^{-1} = \dot{F}_i^T$ and $\dot{A}_i^{-1} \dot{a}_i = \sqrt{e^{\sigma^2} - 1} \mathbf{1}_{m_i}$, the result follows. \square

Define $F_i = \begin{pmatrix} \mathbf{f}^T(x_{i1}) \\ \vdots \\ \mathbf{f}^T(x_{it_i}) \end{pmatrix}$ the row individual design matrix neglecting the number of replications. Then the information matrix simplifies.

Lemma 3.2. *The individuals information matrix can be represented as*

$$(3.3) \quad M_i(\boldsymbol{\beta}) = F_i^T (A_i^{-1} + (e^{\sigma^2} - 1) \mathbf{1}_{t_i} \mathbf{1}_{t_i}^T)^{-1} F_i = F_i^T \left(A_i - \frac{(e^{\sigma^2} - 1) A_i \mathbf{1}_{t_i} \mathbf{1}_{t_i}^T A_i}{1 + (e^{\sigma^2} - 1) \mathbf{1}_{t_i}^T A_i \mathbf{1}_{t_i}} \right) F_i$$

$$\text{with } A_i = \begin{pmatrix} m_{i1}\mu_{i1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & m_{it_i}\mu_{it_i} \end{pmatrix}$$

Proof. Because of (Schott 1997, Corollary 1.7.2)

$$M_i(\boldsymbol{\beta}) = \dot{F}_i^T (\dot{A}_i^{-1} + (e^{\sigma^2} - 1) \mathbf{1}_{m_i} \mathbf{1}_{m_i}^T)^{-1} \dot{F}_i = \dot{F}_i^T \left(\dot{A}_i - \frac{(e^{\sigma^2} - 1) \dot{A}_i \mathbf{1}_{m_i} \mathbf{1}_{m_i}^T \dot{A}_i}{1 + (e^{\sigma^2} - 1) \mathbf{1}_{m_i}^T \dot{A}_i \mathbf{1}_{m_i}} \right) \dot{F}_i$$

Since $\dot{F}_i^T \dot{A}_i \dot{F}_i = F_i^T A_i F_i$, $\dot{F}_i^T \dot{A}_i \mathbf{1}_{m_i} = F_i^T A_i \mathbf{1}_{t_i}$ and $\mathbf{1}_{m_i}^T \dot{A}_i \mathbf{1}_{m_i} = \sum_{j=1}^{t_i} m_{ij} \mu_{ij} = \mathbf{1}_{t_i}^T A_i \mathbf{1}_{t_i}$, we obtain

$$M_i(\boldsymbol{\beta}) = F_i^T A_i F_i - \frac{(e^{\sigma^2} - 1) (F_i^T A_i \mathbf{1}_{t_i}) (\mathbf{1}_{t_i}^T A_i F_i)}{1 + (e^{\sigma^2} - 1) \mathbf{1}_{t_i}^T A_i \mathbf{1}_{t_i}}$$

and the representation follows \square

On the basis of the last representation an approximate individual design ξ_i for individual i can be defined as

$$\xi_i = \left\{ \begin{array}{l} x_{i1}, \dots, x_{it_i} \\ p_{i1}, \dots, p_{it_i} \end{array} \right\},$$

where $x_{i1}, \dots, x_{it_i} \in \mathfrak{X}$ are experimental settings and p_{ij} ($j = 1, \dots, t_i$) is the proportion of individual i that will be observed at the j th level of x , i.e., x_{ij} , so $\sum_{j=1}^{t_i} p_{ij} = 1$ ($i = 1, \dots, s$). In other words, $p_{ij} m_i = m_{ij}$, where m_i is the total number of observations taken from individual i and m_{ij} is the number of observations taken at x_{ij} . We allow that m_{ij} might be not an integer, but an exact design can be obtained by rounding.

In view of (3.3) the corresponding information matrix can be defined as

$$(3.4) \quad \mathbf{m}(\xi_i) = F_i (A_i^{-1} + (e^{\sigma^2} - 1) \mathbf{1}_{t_i} \mathbf{1}_{t_i}^T)^{-1} F_i,$$

$$\text{where now } A_i = A(\xi_i) = m_i \begin{pmatrix} p_{i1}\mu_{i1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & p_{it_i}\mu_{it_i} \end{pmatrix}$$

Next, we consider the vector Y of all observations. The population information matrix is

$$M(\boldsymbol{\beta}) = D^T V^{-1} D \text{ where, } D = \begin{pmatrix} D_1 \\ \vdots \\ D_s \end{pmatrix}.$$

$$\text{Then } M(\boldsymbol{\beta}) = \sum_{i=1}^s D_i^T V_i^{-1} D_i = \sum_{i=1}^s F_i (A_i^{-1} + (e^{\sigma^2} - 1) \mathbf{1}_t \mathbf{1}_t^T)^{-1} F_i.$$

We define a population design as

$$\zeta = \left\{ \begin{array}{c} \xi_1, \dots, \xi_r \\ q_1, \dots, q_s \end{array} \right\}$$

where, $q_i (i = 1, \dots, s)$ stands for the proportion of the individuals that have been observed under the individual design ξ_i .

According to the population design, a representation of $M(\boldsymbol{\beta})$, based on the population information matrix for ζ becomes $\mathbf{m}(\zeta) = \sum_{i=1}^s q_i \mathbf{m}(\xi_i)$.

Theorem 3.1. *Under the assumptions*

i) $\Phi : \mathfrak{M} \rightarrow (-\infty, \infty]$ is a convex function on \mathfrak{M} , where \mathfrak{M} is the set of symmetric non-negative definite matrices of dimension p ,

ii) Φ is monotone in the sense of the Loewner ordering on \mathfrak{M} , i.e., if $M_1 \geq M_2$, then $\Phi(M_1) \leq \Phi(M_2)$,

iii) $\sum_{i=1}^s q_i m_i = m$,

an optimal design can be obtained among those which are uniform across the individuals, i.e., if ξ^ is Φ -optimal for individual design, then $\zeta^* = \left\{ \begin{array}{c} \xi^* \\ 1 \end{array} \right\}$ is a Φ -optimal population design.*

The proof is the same as the proof of Theorem 1 in Schmelter (2007), and has hence been omitted here. This theorem applies, in particular, to D-optimality.

Using this theorem, we can ignore the index i in the experimental settings.

4. APPLICATION: SIMPLE POISSON REGRESSION WITH RANDOM INTERCEPT

In a recent manuscript Yanping et al. (2006) obtain the locally D-optimal design in a Poisson regression model without random effects. In this section we consider the generalization to a Poisson regression model with random intercept, where we

have $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and $\mathbf{f}(x_j) = (1, x_j)^T$, to see the effect of block and, hence, intra-individual correlation increases. So our model reduces to

$$(4.1) \quad Y_{ijk} \mid b_i \sim P(\mu_j(b_i)) \quad \text{where} \quad \mu_j(b_i) = e^{b_i + \beta_0 + \beta_1 x_j}$$

We want to find the locally D-optimal design to estimate $\boldsymbol{\beta}$, which maximizes the determinant of the information matrix.

In most applications of this model, like e.g., bioscience, pharmacokinetics, etc. the design region is the non-negative real line or a subset of that. In other word $\mathfrak{X} = [h, \infty)$ is considered as an unbounded subset and $\mathfrak{X} = [h, g]$ is considered as a bounded subset, where $h \geq 0$ and $g > 0$ determine the bounds for the design region, which has to be defined by experimenter. The expectation $\mu(x_j) = e^{\frac{1}{2}\sigma^2 + \beta_0 + \beta_1 x_j}$ of Y_{ijk} , is a monotone function of x_j . We consider the special case, where $\mu(x_j)$ is considered as a decreasing function of x_j , i.e., $\beta_1 < 0$. Therefore the maximum and the minimum of the mean response are attained at the lower and upper bound of the design region respectively. Let $\mu_h = e^{\beta_0 + \beta_1 h + \frac{1}{2}\sigma^2}$ be the expectation of Y_{ijk} at h , the canonical standardized mean $\tilde{\mu}_j = \tilde{\mu}(x_j) = \frac{\mu_j}{\mu_h} = e^{\beta_1(x_j - h)}$, will always lie in $(0, 1]$ and $[\tilde{\mu}_g, 1]$ respectively corresponding to the design regions $[h, \infty)$ and $[h, g]$. We outline the following lemma which allows us to restrict ourselves to designs with only two different settings x_1 and x_2 .

Lemma 4.1. *For the model (4.1), the D-optimal design $\zeta^* = \left\{ \begin{array}{c} \xi^* \\ 1 \end{array} \right\}$ has exactly two different support points x_1^* and x_2^* , i.e., $\xi^* = \left\{ \begin{array}{cc} x_1^* & x_2^* \\ p_1^* & p_2^* \end{array} \right\}$ where $p_1^*, p_2^* > 0$.*

The proof can be found in appendix A.

Using lemma (4.1), we can restrict ourselves to the case $t = 2$.

Theorem 4.1. *Consider the model (4.1). In terms of the canonical standardized mean, the D-criterion to estimate $\boldsymbol{\beta}$ depends on the parameters only through $\gamma(m, \beta_0(h), \sigma^2) = m e^{\beta_0(h) + \frac{1}{2}\sigma^2} (e^{\sigma^2} - 1)$ as follows.*

$$(4.2) \quad \det(\mathbf{m}(\xi)) \propto \frac{p_1(1-p_1)\tilde{\mu}_1\tilde{\mu}_2(\ln(\tilde{\mu}_1) - \ln(\tilde{\mu}_2))^2}{1 + \gamma(m, \beta_0(h), \sigma^2)(p_1\tilde{\mu}_1 + p_2\tilde{\mu}_2)}$$

where $\beta_0(h) = \beta_0 + \beta_1 h$ is the "intercept" at the lower bound $x = h$ of the design region.

The proof of this theorem is given in appendix B.

Corollary 4.1. Let $\mathfrak{X} = [h, \infty)$, D-optimal design is given by $\xi^* = \begin{Bmatrix} x_1^* & x_2^* \\ p_1^* & p_2^* \end{Bmatrix}$, where $x_j^* = \frac{1}{\beta_1} \log \tilde{\mu}_j^* + h$, and $\tilde{\mu}_1^*, \tilde{\mu}_2^*, p_1^*, p_2^* = 1 - p_1^*$ maximize (4.2).

According to Theorem (4.1), numerical methods can be used to maximize this criterion in order to find D-optimal designs. The D-optimal design for some representative values of $\gamma(m, \beta_0(h), \sigma^2)$ are listed in Table 1 .

Table 1: D-optimal designs for model (4.1)

$\gamma(m, \beta_0(h), \sigma^2)$	p_1^*	$\tilde{\mu}_1^*$	$\tilde{\mu}_2^*$	$\gamma(m, \beta_0(h), \sigma^2)$	p_1^*	$\tilde{\mu}_1^*$	$\tilde{\mu}_2^*$
0	0.500	0.1353	1	30	0.746	0.0864	1
0.5	0.543	0.1279	1	60	0.762	0.0825	1
2	0.609	0.1155	1	100	0.769	0.0807	1
5	0.665	0.1044	1	1000	0.781	0.0779	1
10	0.703	0.0962	1	10000	0.782	0.0776	1

Note that $\tilde{\mu}_2^* = 1$ corresponds to $x_2^* = h$ which means that one point in the optimal setting will be at the lower bound of design region. For the Fixed Effect Simple Poisson Regression models, i.e. $\sigma^2 = 0$, the constant $\gamma(m, \beta_0(h), \sigma^2)$ will be zero for all m and $\beta_0(h)$. In this case we found that the D-optimal design is the design with 50% experimental run at $x_1^* = \frac{1}{\beta_1} \ln(0.1353) + h$ for all $\beta_1 < 0$ and the remaining at $x_2^* = h$, in accordance with the results in Yanping et al. (2006).

It can be easily seen that the intra-individual correlation, $corr_{\sigma^2}(Y_{ijk}, Y_{ij'k'})$, is an increasing function of σ^2 . From Table 1 it can be seen that a larger value of σ^2 and, hence, of the intra-individual correlation decreases the proportion of observations at $x_2^* = h$. When σ^2 tends to infinity, 78% of experiments should be run at $x_1^* = \frac{1}{\beta_1} \ln(0.0776) + h$ and 22% at $x_2^* = h$. These results coincide with the

results to find the D-optimal design for estimation of slope in the corresponding model with fixed block effects (Minkin (1993)).

In the case of a restricted region $\mathfrak{X} = [h, g]$, the D-optimal design is the same as the D-optimal in the unrestricted case, if $\tilde{\mu}_g \leq \tilde{\mu}_1^*$. Otherwise, $\tilde{\mu}_g$ and μ_2^* , i.e., $x_1^* = g$ and $x_2^* = h$ will be optimal values but with different weights p_1^* and p_2^* .

ACKNOWLEDGEMENT

I am grateful to Professor Rainer Schwabe for valuable helps and discussions throughout this work.

APPENDIX A

Proof of Lemma (4.1). For simplicity we ignore the index i in $\mathbf{m}(\xi)$ in (3.4). Since $(\sqrt{e^{\sigma^2} - 1})\mathbf{1}_t = Fc$, we have

$$\mathbf{m}(\xi) = F^T(A^{-1} + (Fc)(Fc)^T)^{-1}F = F^T(A^{-1} + e^{-\frac{1}{2}\sigma^2}FKF^T)^{-1}F$$

$$\text{Where } K = e^{\frac{1}{2}\sigma^2}cc^T \text{ with } c^T = \sqrt{e^{\sigma^2} - 1} \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$$

With regard to the Lemma 1 in Schmelter(2007), we can represent $\mathbf{m}(\xi)$ as follow

$$\mathbf{m}(\xi) = e^{\frac{1}{2}\sigma^2}(\mathbf{m}_0^{-1}(\xi) + K)^{-1}$$

where $\mathbf{m}_0(\xi) = e^{-\frac{1}{2}\sigma^2}(F^TAF)$ is the information matrix for the model without random intercept. Thus

$$\xi^* = \arg \min_{\xi} \Phi[(\mathbf{m}_0^{-1}(\xi) + K)^{-1}] = \arg \min_{\xi} \ln(\det(\mathbf{m}_0^{-1}(\xi) + K))$$

and a necessary and sufficient condition for ξ^* to be D-optimal for the estimation of (β_0, β_1) is that

$$\mu(x)f^T(x)\mathbf{m}_0^{-1}(\xi^*)(\mathbf{m}_0^{-1}(\xi^*) + K)^{-1}\mathbf{m}_0^{-1}(\xi^*)f(x) \leq \text{tr}(\mathbf{m}_0^{-1}(\xi^*) + K)^{-1}\mathbf{m}_0^{-1}(\xi^*) \quad (A.1)$$

(see, e.g. page 76-79 in Fedorov and Hackl(1997)). We represent the inequality (A.1) in the following form

$$\mu(x) \begin{pmatrix} 1 & x \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} \leq d \quad (A.2)$$

with appropriate constant a, b, c, d . Therefore the inequality (A.2) is equivalent to

$$\mu(x)(cx^2 + 2bx + a) \leq d \iff ax^2 + bx + c \leq \frac{d}{\mu(x)}$$

Let $k(x) = \frac{1}{\mu(x)} = e^{-\beta_0 - \beta_1 x - \frac{1}{2}\sigma^2}$. Suppose that the D-optimal design has at least three support points $z_1 < z_2 < z_3$, i.e. $h(z_i) = k(z_i)$ ($i = 1, 2, 3$) where $h(x) = \frac{1}{d}(cx^2 + 2bx + a)$. Applying Rolle's theorem, there are z'_1, z'_2 such that $z_1 < z'_1 < z_2 < z'_2 < z_3$ and $h'(z'_i) = k'(z'_i)$ ($i = 1, 2$).

Because $h(z) \leq k(z)$ for all z , the points z_i ($i = 1, 2, 3$) are local extrema, then $h'(z_2) = k'(z_2)$. By Rolle's theorem again for the function $h'(z) - k'(z)$, we receive points z''_1, z''_2 , such that $z'_1 < z''_1 < z_2 < z''_2 < z'_2$ and $h''(z''_i) = k''(z''_i)$ ($i = 1, 2$). Because $h''(z) = 2\frac{c}{d}$ for all z , we have $k''(z''_i) = 2\frac{c}{d}$ ($i = 1, 2$), which contradicts that $k''(x) = r$ has at most one root. Similar results can be found in Biedermann et al. (2006).

APPENDIX B

Proof of Theorem (4.1). From (3.3) the information matrix for model (4.1) is

$$\begin{aligned} \mathbf{m}(\xi) &= \begin{pmatrix} 1 & 1 \\ x_1 & x_2 \end{pmatrix} \left(\begin{pmatrix} m_1\mu_1 & 0 \\ 0 & m_2\mu_2 \end{pmatrix}^{-1} + (e^{\sigma^2} - 1) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^2 m_j\mu_j & \sum_{j=1}^2 m_j\mu_j x_j \\ \sum_{j=1}^2 m_j\mu_j x_j & \sum_{j=1}^2 m_j\mu_j x_j^2 \end{pmatrix} \\ &= \frac{1}{1 + (e^{\sigma^2} - 1) \sum_{j=1}^2 m_j\mu_j} \begin{pmatrix} \sum_{j=1}^2 m_j\mu_j & \\ & \sum_{j=1}^2 m_j\mu_j x_j \end{pmatrix} \begin{pmatrix} \sum_{j=1}^2 m_j\mu_j & \sum_{j=1}^2 m_j\mu_j x_j \end{pmatrix} \\ &= \frac{1}{1 + (e^{\sigma^2} - 1)(m_1\mu_1 + m_2\mu_2)} \\ &\quad \times \begin{pmatrix} m_1\mu_1 + m_2\mu_2 & m_1\mu_1 x_1 + m_2\mu_2 x_2 \\ m_1\mu_1 x_1 + m_2\mu_2 x_2 & m_1\mu_1 x_1^2 + m_2\mu_2 x_2^2 + (e^{\sigma^2} - 1)m_1\mu_1 m_2\mu_2 (x_1 - x_2)^2 \end{pmatrix} \end{aligned}$$

A straightforward calculation leads to

$$\det(\mathbf{m}(\xi)) = \frac{m_1\mu_1 m_2\mu_2 (x_1 - x_2)^2}{1 + (e^{\sigma^2} - 1) \sum_{j=1}^2 m_j\mu_j} = \frac{m^2\mu_h^2 p_1 p_2 \tilde{\mu}_1 \tilde{\mu}_2 (x_1 - x_2)^2}{1 + (e^{\sigma^2} - 1) m\mu_h \sum_{j=1}^2 p_j \tilde{\mu}_j}$$

As, $(x_1 - x_2) = \frac{\ln \tilde{\mu}_1 - \ln \tilde{\mu}_2}{\beta_1}$, the representation follows

REFERENCES

- [1] Abdelbasit K.M and Plackett R.L., 1983. Experimental Designs for Binary Data. *J. Amer. Statist. Assoc.* 78, 90-98.
- [2] Biedermann S., Dette H. and Zhu W., 2006. Optimal Designs for Dose-Response Models With Restricted Design Spaces. *J. Amer. Statist. Assoc.* 101, 747-759.
- [3] Fedorov V.V. and Hackl P., 1997. Model-Oriented Design of Experiments. Lecture Note in Statistics. 125, Springer, Berlin.
- [4] McCullagh P. and Nelder J. A., 1998. Generalized Linear Models (Reprinted). Chapman and Hall, London.
- [5] McCulloch C.E. and Searle S.R., 2001. Generalized, Linear, Mixed Models. Wiley, New York.
- [6] Minkin, S., 1987. Optimal Designs for Binary Data. *J. Amer. Statist. Assoc.* 82, 1098-1103.
- [7] Minkin, S., 1993. Experimental Designs for Clonogenic Assay in Chemotherapy. *J. Amer. Statist. Assoc.* 88, 410-420.
- [8] Ouwens J.N.M. Mario, Tan E.S. Frans and Berger P.F. Martijn, 2006. A maximin criterion for the logistic random intercept model with covariates. *J. Statist. Plann. Inference* 136, 962-981.
- [9] Schmelter, T., 2007. The Optimality of Single-Group Designs for Certain Mixed Models. *Metrika.* 65, 183-193
- [10] Schott J. R., D., 1997. Matrix Analysis for Statistics. Wiley, New York.
- [11] Yanping W., Raymond H. M., Smith E. P. and Keying Ye, 2006. D-Optimal Designs for Poisson Regression Models. *J. Statist. Plann. Inference* 136, 2831-2845.