



V-optimality of designs in random effects Poisson regression models

Mehrdad Niaparast¹ · Sahar MehrMansour¹ · Rainer Schwabe²

Received: 19 November 2021 / Accepted: 11 January 2023 / Published online: 30 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The knowledge of the Fisher information is a fundamental tool to judge the quality of an experiment. Unlike in linear and generalized linear models without random effects, there is no closed form for the Fisher information in the situation of generalized linear mixed models, in general. To circumvent this problem, we make use of the quasi-information in this paper as an approximation to the Fisher information. We derive optimal designs based on the V-criterion, which aims to minimize the average variance of prediction of the mean response. For this criterion, we obtain locally optimal designs in two specific cases of a Poisson straight line regression model with either random intercepts or random slopes.

Keywords Fisher information · Poisson regression · Quasi information · Random effects · V-optimality

1 Introduction

In a wide range of regression experiments, the response may be observed as count measurements which can be modeled by the Poisson distribution. An important application of the Poisson regression model is in cancer research. There, the clonogenic assay is a technique that is used to determine the effect of an anticancer drug on proliferating tumor cells. In these studies, a dose-response curve can be fitted to determine how

✉ Mehrdad Niaparast
niaparast@razi.ac.ir
Sahar MehrMansour
mehrmanous.sahar@razi.ac.ir
Rainer Schwabe
rainer.schwabe@ovgu.de

¹ Department of Statistics, Razi University, Kermanshah, Iran

² Institute for Mathematical Stochastics, Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany

much the dosage of a drug decreases the number of colony formations over different concentrations of the particular drug. Then, often the number of colonies is assumed to be Poisson distributed, and thus a Poisson regression model can be used to evaluate the results of the dose-response experiment (Minkin 1993).

The concept of generalized linear models (GLM; McCullagh and Nelder 1989) provides a fundamental idea to model such data. However, in statistical studies, the data set often reveals features that cause these models to be insufficient, in particular when observations are correlated that can be observed in clustered measurements (e.g., hierarchical data structures and longitudinal studies). To adjust for this clustering, generalized linear mixed models (GLMM; McCulloch and Searle 2001) can be used to generalize the GLMs by including random clusters and/or subject effects in the model equation to account for dependence in the data. In many applications, it is assumed that the random effects follow a normal distribution. In that case, it is well-known that, in general, the marginal likelihood of the GLMM cannot be expressed analytically in an explicit form. Hence, the maximum-likelihood (ML) estimation of the parameters has to be approximated numerically (Davidian and Giltinan 1995). Consequently, there is not sufficient knowledge about the statistical behavior of the ML estimators for GLMMs, which could serve as the basis for the design of experiments. Commonly the inverse of the Fisher information matrix is used to measure the performance of the ML estimator, at least in terms of its asymptotic properties. However, also the Fisher information cannot be expressed analytically in an explicit form. To resolve this, there are two different approaches to replacing the Fisher information matrix. One common approach is a direct approximation of the Fisher Information, for example, by linearization of the response function (see Mielke 2012). But Mielke and Schwabe (2010) showed that these Fisher information approximations might often be unreliable.

Another approach is to use alternative methods of estimation in GLMMs like Generalized Estimation Equations (GEE, see McCulloch and Searle 2001; Cameron and Trivedi 2013). In particular, the concept of quasi-likelihood can be identified as a GEE. It may serve as an attractive method to generate an efficient estimation of the parameters without using the full distributional assumptions. Niaparast (2009) used the quasi-likelihood method to obtain optimal designs in a Poisson regression model with random intercepts. There he employed results of Schmelter (2007) and proved for a range of criteria that an optimal design on an individual level is also optimal for the population parameters. Niaparast and Schwabe (2013) derived D-optimal designs for a Poisson regression model with random slope. In this article, we extend the results of Niaparast (2009) and Niaparast and Schwabe (2013) to obtain V-optimal designs in some Poisson regression models with random effects. The V-optimality criterion is a prediction-based optimality criterion that minimizes the average prediction variance of the estimated mean response. The V-optimality criterion is also often called I-optimality (Atkinson et al. 2007), Q-optimality (Bandemer 1977), or IMSE-optimality (Entholzner et al. 2005) in the literature.

As in all non-linear situations, there is a problem in finding optimal designs because of the dependence of the quasi-information matrix on unknown model parameters. Therefore, the local approach is adopted for which locally optimal designs are determined for prespecified nominal values of the parameters.

The structure of the paper is organized as follows: We first introduce the Poisson regression model with random effects in Sect. 2 and derive the quasi-information matrix for this model. The V-criterion is introduced in Sect. 3. In Sect. 4, we obtain V-optimal designs for two specific cases of straight line Poisson regression with either random intercepts or random slopes. Finally, the efficiency of a standard design is examined in a case study.

2 Model specification, information and design

In a Poisson regression model with random coefficients and log link, on the unit level it is assumed that the observation Y_{ij} of unit $i, i = 1, \dots, n$, at the j th replication, $j = 1, \dots, m_i$, at the experimental setting x_{ij} from the experimental region \mathcal{X} is distributed according to a Poisson distribution with intensity $\lambda_{ij} = \exp(\mathbf{f}^T(x_{ij})\mathbf{b}_i)$ where, in the linear component $\mathbf{f}^T(x_{ij})\mathbf{b}_i, \mathbf{f} = (f_0, f_1, \dots, f_{p-1})^T$ and $\mathbf{b}_i = (b_{0,i}, \dots, b_{p-1,i})^T$ are the p -dimensional vectors of known regression functions and individual coefficients, respectively. Conditional on the value of the individual coefficients, the observations Y_{ij} are assumed to be independent.

On the population level, the individual coefficients \mathbf{b}_i are assumed to be random iid multivariate normal with population mean parameter $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^T$ and variance covariance matrix $\boldsymbol{\Sigma}$ (see McCulloch and Searle 2001). For simplification, we further assume that $\boldsymbol{\Sigma}$ is known. Then, the marginal mean of the observation Y_{ij} is given by

$$E(Y_{ij}) = \mu(x_{ij})$$

where $\mu(x) = \exp(\mathbf{f}^T(x)\boldsymbol{\beta} + \sigma(x, x)/2)$ is the mean function and $\sigma(x, x') = \mathbf{f}^T(x)\boldsymbol{\Sigma}\mathbf{f}(x')$ is the dispersion function. Also, the marginal variance of the observation Y_{ij} can be calculated as

$$\text{Var}(Y_{ij}) = \mu(x_{ij}) + \mu(x_{ij})^2c(x_{ij}, x_{ij})$$

where $c(x, x') = \exp(\sigma(x, x')) - 1$ is the variance correction term.

Further the covariance of the observations within a unit is $\text{cov}(Y_{ij}, Y_{ik}) = \mu(x_{ij})\mu(x_{ik})c(x_{ij}, x_{ik})$. Hence for the vector $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})^T$, the mean and the variance covariance matrix of \mathbf{Y}_i are given by $E(\mathbf{Y}_i) = (\mu(x_{i1}), \dots, \mu(x_{im_i}))^T$ and

$$\text{Cov}(\mathbf{Y}_i) = \mathbf{A}_i + \mathbf{A}_i\mathbf{C}_i\mathbf{A}_i,$$

where $\mathbf{A}_i = \text{diag}\{\mu(x_{ij})\}_{j=1, \dots, m_i}$ is a $m_i \times m_i$ diagonal matrix with the observation means $\mu(x_{ij})$ as its non-zero, diagonal entries and $\mathbf{C}_i = (c(x_{ij}, x_{ik}))_{j,k=1, \dots, m_i}$ is the $m_i \times m_i$ matrix of variance correction terms. As $\mathbf{A}_i\mathbf{C}_i\mathbf{A}_i$ is the covariance matrix of the conditional expectation of \mathbf{Y}_i given \mathbf{b}_i and as \mathbf{A}_i is symmetric and non-singular, the matrix \mathbf{C}_i of variance correction terms is non-negative definite. The variance

covariance matrix $\text{Cov}(Y_i)$ depends on the population location parameters β only through the mean function.

Because of the integral in the marginal density with respect to the distribution of the random effects, it is not possible to derive a closed form for the likelihood function and the Fisher information. As an alternative approach, we will use the quasi-likelihood method.

By applying the quasi-likelihood approach, the amount of information contributed by the observations on the individual level can be described by the individual quasi-information matrix $\mathfrak{M}_{i,\beta}$,

$$\mathfrak{M}_{i,\beta} = \mathbf{F}_i^T \left(\mathbf{A}_i^{-1} + \mathbf{C}_i \right)^{-1} \mathbf{F}_i, \quad (1)$$

where $\mathbf{F}_i = (\mathbf{f}(x_{i1}), \dots, \mathbf{f}(x_{im_i}))^T$ and, on the right hand side, only \mathbf{A}_i depends on β . Then on the population level, the quasi-information matrix is given by $\mathfrak{M}_\beta = \sum_{i=1}^n \mathfrak{M}_{i,\beta}$. For more details, see McCullagh and Nelder (1989, Chap. 9).

On the individual level, we define individual approximate designs ξ_i as

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

where x_{i1}, \dots, x_{is_i} are mutually distinct experimental settings from the design region \mathcal{X} and $w_{ij} = m_{ij}/m_i$ are the proportions of observations taken at x_{ij} , $j = 1, \dots, s_i$, $\sum_{j=1}^{s_i} w_{ij} = 1$. Then, for the design ξ_i , the individual quasi-information matrix (1) can be represented as

$$\mathfrak{M}_\beta(\xi_i) = \mathbf{F}_{\xi_i}^T \left(\mathbf{A}_{\xi_i}^{-1} + \mathbf{C}_{\xi_i} \right)^{-1} \mathbf{F}_{\xi_i}, \quad (2)$$

where $\mathbf{F}_{\xi_i} = (\mathbf{f}(x_{i1}), \dots, \mathbf{f}(x_{is_i}))^T$, $\mathbf{A}_{\xi_i} = \text{diag}(m_i w_{ij} \mu(x_{ij}))_{j=1, \dots, s_i}$ and $\mathbf{C}_{\xi_i} = (\mathbf{c}(x_{ij}, x_{ik}))_{j,k=1, \dots, s_i}$ (see Niaparast and Schwabe 2013).

Similarly, we define population approximate designs ζ by

$$\zeta = \left\{ \begin{array}{l} \xi_1, \dots, \xi_q \\ \nu_1, \dots, \nu_q \end{array} \right\},$$

where ξ_l , $l = 1, \dots, q$, are mutually distinct individual designs and $\nu_l > 0$, $\sum_{l=1}^q \nu_l = 1$, are the proportions of individuals that are observed under the corresponding individual designs. Based on the fact that \mathbf{C}_{ξ_i} is non-negative definite, Niaparast and Schwabe (2013) proved the convexity of the individual quasi-information. Following the argumentation in Schmelter (2007), it is sufficient to consider the class of single group designs, $\zeta = \left\{ \begin{array}{l} \xi \\ 1 \end{array} \right\}$, where all individuals are observed under the same individual design, to obtain an optimal design in the class of all population designs, when for all units the same number $m_i = m$ of observations are taken. Hence, we have to optimize designs only on the individual level. To simplify notation, we may,

thus, omit the index i in the individual design, $\xi = \left\{ \begin{matrix} x_1, \dots, x_s \\ w_1, \dots, w_s \end{matrix} \right\}$ throughout the remainder of the text.

Our main interest in this paper is to find optimal designs for estimating the predicted response averaged over the design region with minimal variance. For the present model, the predictor $\hat{\mu}_x$ of the mean response at a given point x is given by

$$\hat{\mu}_x = \exp \left(\mathbf{f}^T(x) \hat{\boldsymbol{\beta}} + \sigma(x, x)/2 \right),$$

where $\hat{\boldsymbol{\beta}}$ is the quasi-likelihood estimator of $\boldsymbol{\beta}$.

The mean response $\mu(x)$ and, hence, its predictor $\hat{\mu}_x$ does not have a linear structure. Therefore, the variance of the predictor has to reflect the non-linearity, and the asymptotic variance of the predictor has to be obtained by the delta method.

Lemma 1 *The asymptotic variance $\text{Avar}(\hat{\mu}_x)$ of the predictor $\hat{\mu}_x$ of the mean response $\mu(x)$ can be represented as*

$$\text{Avar}(\hat{\mu}_x) = \mu^2(x) \mathbf{f}^T(x) \mathfrak{M}_{\boldsymbol{\beta}}^{-1}(\xi) \mathbf{f}(x).$$

Proof According to McCullagh (1983), $\hat{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$ and asymptotic covariance matrix $\text{Avar}(\hat{\boldsymbol{\beta}}) = \mathfrak{M}_{\boldsymbol{\beta}}^{-1}(\xi)$ equal to the inverse of the quasi-information matrix. Let $g(\boldsymbol{\beta}) = \mu(x) = \exp(\mathbf{f}^T(x)\boldsymbol{\beta} + \sigma(x, x)/2)$. By the delta method, $g(\hat{\boldsymbol{\beta}})$ is asymptotically normal with asymptotic variance $\nabla g(\boldsymbol{\beta})^T \text{Avar}(\hat{\boldsymbol{\beta}}) \nabla g(\boldsymbol{\beta})$, where $\nabla g(\boldsymbol{\beta}) = \exp(\mathbf{f}^T(x)\boldsymbol{\beta} + \sigma(x, x)/2) \mathbf{f}(x) = \mu(x) \mathbf{f}(x)$ is the the gradient vector of continuous partial derivatives of g . Hence, the statement of the lemma follows.

3 Optimality criterion and equivalence theorem

We consider the V-criterion

$$\phi(\xi) = \int \text{Avar}(\hat{\mu}_x) \nu(dx),$$

of the averaged asymptotic variance for the mean response with respect to a measure ν on the design region \mathcal{X} . Commonly this measure ν is chosen to be uniform on \mathcal{X} (continuous or discrete depending on the structure of \mathcal{X}). For example, if x is a continuous variable and ν is uniform on the interval $\mathcal{X} = [a, b]$, then the V-criterion can be written as

$$\phi(\xi) = \int_a^b \text{Avar}(\hat{\mu}_x) dx.$$

More generally, the measure ν may be any (finite) measure on \mathcal{X} or, in the case of extrapolation, it might be supported by an even larger set \mathcal{X}' of settings x for which

the model equations are assumed to be valid. A design ξ^* will be called V-optimal if it minimizes the V-criterion $\phi(\xi)$. Note that here, as in all nonlinear models, the V-criterion depends on the nominal value of the parameter vector β and, hence, the V-optimal design will be a local solution (at β). In the appendix, it is shown that the V-criterion is a convex function in ξ .

For the given model, the V-criterion can be represented as

$$\begin{aligned} \phi(\xi) &= \int \text{Avar}(\hat{\mu}_x) \nu(dx) = \int \mu^2(x) \mathbf{f}^T(x) \mathfrak{M}_\beta^{-1}(\xi) \mathbf{f}(x) \nu(dx) \\ &= \int \text{tr} \left(\mathfrak{M}_\beta^{-1}(\xi) \mu^2(x) \mathbf{f}(x) \mathbf{f}^T(x) \right) \nu(dx) = \text{tr} \left(\mathfrak{M}_\beta^{-1}(\xi) \mathbf{B} \right), \end{aligned} \tag{3}$$

where the matrix $\mathbf{B} = \int \mu^2(x) \mathbf{f}(x) \mathbf{f}^T(x) \nu(dx)$ does not depend on ξ .

The V-criterion and, hence, the V-optimal design depend on β through the mean function $\mu(x)$. These designs will be called locally optimal (at β). Finding a V-optimal design requires knowledge of the values of the parameters. A common approach is to use an initial guess of the parameter values as nominal values which will be adopted here.

An important tool for checking the optimality of a design ξ^* for any convex and differentiable design criterion, is the equivalence theorem. This equivalence theorem has been originally proven by Kiefer and Wolfowitz (1960) for linear models in the case of D-optimality. The equivalence theorem has been extended to nonlinear models by White (1973) and to general convex criteria by Whittle (1973). A fundamental ingredient in the equivalence theorem is the sensitivity function, which constitutes the non-constant part of the directional derivative and plays a major role in optimal design theory (see Atkinson et al. 2007). The equivalence theorem for V-optimality in the given model is considered in the following theorem. For this we introduce the vector $\mathbf{c}_{\xi,x} = (c(x_j, x))_{j=1,\dots,s}$ of joint (correlation) correction terms for the settings x_1, \dots, x_s of a design ξ with a further setting x . This is part of the sensitivity function

$$\begin{aligned} d(x, \xi) &= m\mu(x) \left(\mathbf{f}(x) - \mathbf{F}_\xi^T \left(\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi \right)^{-1} \mathbf{c}_{\xi,x} \right)^T \mathfrak{M}_\beta^{-1}(\xi) \mathbf{B} \mathfrak{M}_\beta^{-1}(\xi) \\ &\quad \left(\mathbf{f}(x) - \mathbf{F}_\xi^T \left(\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi \right)^{-1} \mathbf{c}_{\xi,x} \right). \end{aligned}$$

Theorem 1 ξ^* is V-optimal for the Poisson regression model with random effects if (and only if)

$$\begin{aligned} d(x, \xi^*) &\leq \text{tr} \left(\mathfrak{M}_\beta^{-1}(\xi^*) \mathbf{B} \right) \\ &\quad - \text{tr} \left(\mathfrak{M}_\beta^{-1}(\xi^*) \mathbf{F}_{\xi^*}^T \left(\mathbf{A}_{\xi^*}^{-1} + \mathbf{C}_{\xi^*} \right)^{-1} \mathbf{C}_{\xi^*} \left(\mathbf{A}_{\xi^*}^{-1} + \mathbf{C}_{\xi^*} \right)^{-1} \mathbf{F}_{\xi^*} \mathfrak{M}_\beta^{-1}(\xi^*) \mathbf{B} \right), \end{aligned} \tag{4}$$

for all $x \in \mathcal{X}$.

Moreover, if ξ^* is V-optimal, then equality holds in (4) for all support points x of ξ^* .

The proof is given in the appendix.

4 Poisson regression with a single linear component

Often in dose-response experiments the mean response curve can be represented by a straight line relationship. In this case, a simple Poisson regression model with random effects may be fitted.

In this section, we are going to investigate two types of a straight line Poisson regression model with $p = 2$ population parameters: The simple Poisson regression model with random intercepts and the simple Poisson regression model with random slopes. These models can be characterized by their intensities (mean response)

$$\lambda_{ij} = \exp(b_{0,i} + \beta_1 x_{ij}) \tag{5}$$

and

$$\lambda_{ij} = \exp(\beta_0 + b_{1,i} x_{ij}), \tag{6}$$

respectively. In the former, the individual coefficients have only influence on the overall level of the response, and hence $b_{0,i}$ expresses as the random intercept with mean β_0 and the variance σ^2 . In the context of dose-response curves in cancer research, it is reasonable to assume that the slope β_1 is negative, because the purpose of using the drug is to eradicate the tumor cells. The dosage x will commonly measured on a standardized scale, $x \in [0, 1]$ or $x \in \{0, 1\}$, where $x = 0$ indicates placebo or a control level.

It can be shown that in models with random intercepts the linear optimality criteria do not depend on the random variation of the intercept whereas this may be in challenge when the effect of the explanatory variable varies across individuals. The latter situation is described in (6) in which $b_{1,i}$ is the random slope effect with mean β_1 and variance σ^2 .

For these two models, the vector of regression functions is given by $f(x) = (1, x)^T$ and $\beta = (\beta_0, \beta_1)^T$. In this case the weighting matrix B in the V-criterion can be written as $B = \begin{pmatrix} B_0 & B_1 \\ B_1 & B_2 \end{pmatrix}$, where $B_k = \int_{\mathcal{X}} x^k \mu^2(x) \nu(dx)$ denotes the k th moment with respect to $\mu^2(\cdot) \nu$, $k = 0, 1, 2$.

4.1 Random intercepts

We first consider the Poisson model with random intercepts, where $b_{0,i} \sim N(\beta_0, \sigma^2)$ and the slope $b_{1,i} = \beta_1$ for all units i is a fixed effect. This condition can be described formally by the dispersion matrix $\Sigma = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}$. The dispersion function and the

variance correction term are $\sigma(x, x') = \sigma^2$ and $c(x, x') = \exp(\sigma^2) - 1$, respectively. In this case the functions $\sigma(x, x')$ and $c(x, x')$ are constant on x .

Theorem 2 *For the Poisson regression with random intercept, the V-optimal design only depends on the slope β_1 and does not depend on the mean intercept β_0 , the variance σ^2 of the random intercept, and the number m of observations per unit. Moreover, the V-optimal design for the random intercepts model coincides with that for the corresponding fixed effects model without random effects.*

Proof According to Niaparast (2010, Lemma 4.1.3), the information matrix can be represented as $\mathfrak{M}_\beta(\xi) = \left((F_\xi^T A_\xi F_\xi)^{-1} + U \right)^{-1}$, where $U = \begin{pmatrix} \exp(\sigma^2) - 1 & 0 \\ 0 & 0 \end{pmatrix}$ does not depend on ξ . Then

$$\phi(\xi) = \text{tr} \left(\mathfrak{M}_\beta^{-1}(\xi) \mathbf{B} \right) = \text{tr} \left((F_\xi^T A_\xi F_\xi)^{-1} \mathbf{B} \right) + \text{tr}(U \mathbf{B}).$$

The last expression does not depend on the design and can be ignored. Now let $\check{\mu}(x) = \exp(\beta_1 x)$, $\check{A}_\xi = \text{diag}\{w_j \check{\mu}(x_j)\}_{j=1, \dots, s}$, and $\check{\mathbf{B}} = \begin{pmatrix} \check{B}_0 & \check{B}_1 \\ \check{B}_1 & \check{B}_2 \end{pmatrix}$ with $\check{B}_k = \int x^k \check{\mu}^2(x) \nu(dx)$, $k = 0, 1, 2$. Then,

$$\text{tr} \left((F_\xi^T A_\xi F_\xi)^{-1} \mathbf{B} \right) = \frac{\exp(\beta_0 + \sigma^2/2)}{m} \text{tr} \left((F_\xi^T \check{A}_\xi F_\xi)^{-1} \check{\mathbf{B}} \right),$$

and minimization of $\phi(\xi)$ is equivalent to minimization of $\check{\phi}(\xi) = \text{tr} \left((F_\xi^T \check{A}_\xi F_\xi)^{-1} \check{\mathbf{B}} \right)$ which does not depend on β_0 , σ^2 and m . Furthermore,

$$\phi(\xi) = a\phi_0(\xi) + c, \tag{7}$$

where $a = \exp(\sigma^2/2)$ and $c = \text{tr}(U \mathbf{B}) = \exp(2\beta_0 + \sigma^2)(\exp(\sigma^2) - 1)\check{B}_0$. Obviously a and c do not depend on ξ and $\phi_0(\xi) = \frac{\exp(\beta_0)}{m} \check{\phi}(\xi)$ can be identified as the V-criterion in the corresponding fixed effects model without random intercepts by formally letting $\sigma^2 = 0$. This proves the theorem.

Example 1 (Binary regressor). We consider the simple Poisson regression model with random intercepts and binary predictor $x \in \mathcal{X} = \{0, 1\}$, where $x = 1$ denotes treatment and $x = 0$ control. Here, the designs $\xi = \xi_w = \begin{Bmatrix} 0 & 1 \\ 1 - w & w \end{Bmatrix}$ are uniquely characterized by the weight w allocated to $x = 1$, $F_\xi = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is the same for all designs, and $\check{A}_\xi = \begin{pmatrix} 1 - w & 0 \\ 0 & w \exp(\beta_1) \end{pmatrix}$ with the notation in the proof of Theorem 2. Note that, here, the uniform measure ν can be chosen as the counting measure on $\mathcal{X} = \{0, 1\}$ which assigns weight 1 to each of the settings 0 and 1. Then

$\check{B} = \begin{pmatrix} 1 + \exp(2\beta_1) & \exp(2\beta_1) \\ \exp(2\beta_1) & \exp(2\beta_1) \end{pmatrix}$, and we have to minimize

$$\check{\phi}(\xi) = \frac{\exp(\beta_1)}{w} + \frac{1}{1-w} \tag{8}$$

with respect to w . The derivative is $d\check{\phi}(\xi_w)/dw = -(\exp(\beta_1)/w^2) + (1/(1-w)^2)$. Solving for a root satisfying $0 \leq w \leq 1$ leads to the optimal weight $w^* = 1/(1 + \exp(-\beta_1/2))$ at $x = 1$.

The optimal weight w^* is increasing in β_1 , equal to $1/2$ for $\beta_1 = 0$, and counter-symmetric with respect to sign change of β_1 , i.e. if w^* is optimal for β_1 , then $1 - w^*$ is optimal for $-\beta_1$. In accordance with the findings in Niaparast (2009), the V-optimal design does not depend on the variance σ^2 of the random intercept and, hence, coincides with the V-optimal design in the corresponding fixed effects model without random effects. For example, if we assume a nominal value of $\beta_1 = -2$ for the slope, the optimal weight at $x = 1$ is $w^* = 0.269$.

Example 2 (Continuous regressor). Let $\mathcal{X} = [0, 1]$ and ν uniform on \mathcal{X} . It can be shown that for the present model the V-optimal design is supported on two points x_1^* and x_2^* (see Schmidt 2019). Therefore, we search for the optimal design ξ^* among

all two point designs for which ξ^* is of the form $\xi = \begin{Bmatrix} x_1 & x_2 \\ 1-w & w \end{Bmatrix}$ and $x_1 < x_2$.

According to the proof of Theorem 2, $\check{A}_\xi = \begin{pmatrix} (1-w)\exp(\beta_1 x_1) & 0 \\ 0 & w\exp(\beta_1 x_2) \end{pmatrix}$ and

$\check{B} = \begin{pmatrix} \check{B}_0 & \check{B}_1 \\ \check{B}_1 & \check{B}_2 \end{pmatrix}$ with $\check{B}_k = \int_0^1 x^k \check{\mu}^2(x) dx$, $k = 0, 1, 2$, such that

$$\begin{aligned} \check{B}_0 &= (\exp(2\beta_1) - 1) / (2\beta_1), \\ \check{B}_1 &= ((2\beta_1 - 1)\exp(2\beta_1) + 1) / (2\beta_1)^2, \\ \check{B}_2 &= ((4\beta_1^2 - 4\beta_1 + 2)\exp(2\beta_1) - 2) / (2\beta_1)^3. \end{aligned}$$

Solutions for the V-optimal designs are listed in Table 1 for different values of β_1 , where w^* denotes the optimal weight in x_2^* .

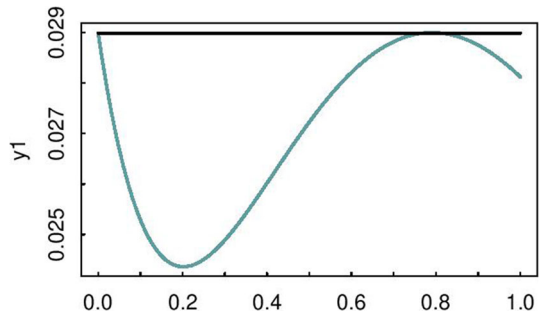
When $|\beta_1|$ is small, the support points will be the endpoints $x_1^* = 0$ and $x_2^* = 1$. For larger values $|\beta_1|$ of the slope, an internal support point x^* will come in together with $x_2^* = 1$ when β_1 is positive and $x_1^* = 0$ when β_1 is negative. Hence, always the setting $x_1^* = 0$ or $x_2^* = 1$ is included in the optimal design which has highest intensity, respectively.

The conditions of Theorem 2 can be checked to confirm that the designs in Table 1 are V-optimal. For example, in the case of a nominal value $\beta_1 = -2$ for the slope, the optimal design is given by $\xi^* = \begin{Bmatrix} 0.000 & 0.790 \\ 0.465 & 0.535 \end{Bmatrix}$. Figure 1 shows that the sensitivity function of ξ^* attains its maximum at the support points $x_1^* = 0.000$ and $x_2^* = 0.790$ of ξ^* .

Table 1 Locally V-optimal settings x_1^* , x_2^* , and weights w^* at x_2^* for Example 2

slope β_1	design x_1^*	x_2^*	w^*
-5	0.000	0.335	0.561
-4	0.000	0.418	0.560
-3	0.000	0.550	0.556
-2	0.000	0.790	0.535
-1	0.000	1.000	0.502
0	0.000	1.000	0.500
1	0.000	1.000	0.498
2	0.210	1.000	0.465
3	0.450	1.000	0.444
4	0.582	1.000	0.440
5	0.665	1.000	0.439

Fig. 1 Sensitivity function of ξ^* for $\beta_1 = -2$ (Example 2)



Note that also here the optimal designs are counter-symmetric with respect to sign change of β_1 , i.e. if the design $\begin{Bmatrix} x_1^* & x_2^* \\ 1 - w^* & w^* \end{Bmatrix}$ is optimal for β_1 , then the design $\begin{Bmatrix} 1 - x_2^* & 1 - x_1^* \\ w^* & 1 - w^* \end{Bmatrix}$ is optimal for $-\beta_1$. Further, the optimal designs can be transferred to arbitrary intervals $\mathcal{X} = [a, b]$ by means of the concept of equivariance (see Idaïis et al. 2021).

4.2 Random slopes

In the simple Poisson regression with random slopes, (6), the individual slope $b_{1,i}$ is assumed to be normally distributed with mean β_1 and variance σ^2 , and the intercept $b_{0,i} = \beta_0$ is constant across units. The covariance structure of the random effect $b_i = (\beta_0, b_{1,i})^T$ is $\Sigma = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix}$ which results in $\sigma(x, x') = \sigma^2 x x'$ for the dispersion function and $c(x, x') = \exp(\sigma^2 x x') - 1$ for the variance correction. For simplicity, as

in Example 1 for the model with random intercepts, we consider also here the situation of a design region which is restricted to only two settings.

Example 3 (Binary regressor). As in Example 1 we treat the case of a design region $\mathcal{X} = \{0, 1\}$ which only consists of two settings $x_1 = 0$ and $x_2 = 1$. Hence, again only designs ξ of the form $\xi_w = \begin{Bmatrix} 0 & 1 \\ 1-w & w \end{Bmatrix}$ have to be considered, where w is the proportion of observations which are taken at $x_2 = 1$, $0 < w < 1$. To find the optimal design ξ^* , it is sufficient to obtain the optimal weight w^* . In this situation $F_\xi = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ is as defined in Example 1, $A_\xi = m \exp(\beta_0) \begin{pmatrix} 1-w & 0 \\ 0 & w \exp(\beta_1 + \sigma^2/2) \end{pmatrix}$, and $C_\xi = \begin{pmatrix} 0 & 0 \\ 0 & \exp(\sigma^2) - 1 \end{pmatrix}$. Here $F = F_\xi$ and $C = C_\xi$ do not depend on ξ , and the quasi-information matrix can be simplified as follows.

Lemma 2 *Let $\mathcal{X} = \{0, 1\}$. For the Poisson regression model with random slopes, the quasi-information matrix can be represented as*

$$\mathfrak{M}_\beta(\xi) = \left((F^T A_\xi F)^{-1} + C \right)^{-1}.$$

Proof Since $C = F^{-1} C F^{-T}$, where F^{-T} denotes the inverse of F^T , we have

$$F^T (A_\xi^{-1} + C)^{-1} F = \left(F^{-1} A_\xi^{-1} F^{-T} + F^{-1} C F^{-T} \right)^{-1} = \left((F^T A_\xi F)^{-1} + C \right)^{-1}.$$

Then the V-criterion can be simplified.

Theorem 3 *Let $\mathcal{X} = \{0, 1\}$. For the simple Poisson regression with random slopes, the V-optimal design does not depend on the mean intercept β_0 and the number m of observations per unit.*

Proof Using Lemma 2, we obtain

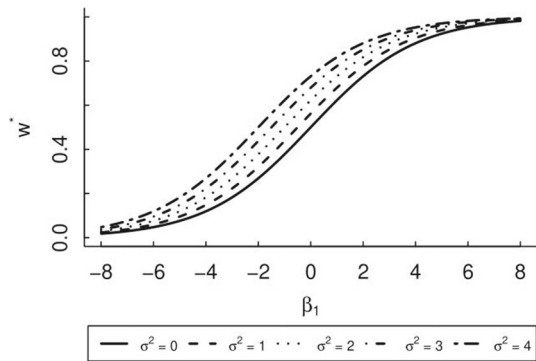
$$\phi(\xi) = \text{tr} \left(\mathfrak{M}_\beta^{-1}(\xi) \mathbf{B} \right) = \text{tr} \left((F^T A_\xi F)^{-1} \mathbf{B} \right) + \text{tr}(C \mathbf{B}).$$

The last term of the above statement is independent of ξ and can be disregarded for optimization. Define here $\check{\mu}(x) = \exp(\beta_1 x + \sigma^2 x^2/2)$ and accordingly $\check{A}_\xi = \text{diag}\{w_j \check{\mu}(x_j)\}_{j=1, \dots, s}$, $\check{\mathbf{B}} = \left(\int x^{i+j-2} \check{\mu}^2(x) \nu(dx) \right)_{i=1,2}^{j=1,2}$, and $\check{\phi}(\xi) = \text{tr} \left((F^T \check{A}_\xi F)^{-1} \check{\mathbf{B}} \right)$. Then $A = m \exp(\beta_0) \check{A}$, $B = \exp(2\beta_0) \check{B}$, and $\text{tr} \left((F^T A_\xi F)^{-1} B \right) = (\exp(\beta_0)/m) \check{\phi}(\xi)$. Hence, the minimization of $\phi(\xi)$ is the same as the minimization of $\check{\phi}(\xi)$. Obviously, $\check{\phi}(\xi)$ does neither depend on β_0 nor on m .

Inserting F , \check{A}_ξ and \check{B} in $\check{\phi}(\xi)$, we get

$$\check{\phi}(\xi) = \frac{1}{1-w} + \frac{\exp(\beta_1 + \sigma^2/2)}{w}, \tag{9}$$

Fig. 2 Optimal weight w^* at $x = 1$ for the model with random slopes (Example 3)



which is minimized at $w^* = 1 / (1 + \exp(-\beta_1/2 - \sigma^2/4))$. For example, in the case of nominal values $\beta_1 = -2$ and $\sigma^2 = 1$ for the mean slope, the optimal weight at $x = 1$ is given by $w^* = 0.321$.

Note that when $\sigma^2 = 0$ (no random effect), the optimal weight coincides with the fixed effects solution in Example 1. Figure 2 exhibits the optimal weights for different values of σ^2 and β_1 .

This figure shows that an increase of the mean slope β_1 or the variance σ^2 of the random slopes leads to an increase in the optimal weight w^* at $x = 1$.

5 Efficiency considerations

In many cases, it is prevalent that a standard design will be applied in the experiments. Therefore, we consider the uniform design $\xi_0 = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}$ on the endpoints of a standardized design region as a candidate.

To judge the gain of the design optimization, it is of interest to evaluate the efficiency of the design ξ_0 with respect to the V-optimal designs. In general, the V-efficiency of a design ξ is defined by

$$\text{eff}(\xi) = \frac{\phi(\xi^*)}{\phi(\xi)} \tag{10}$$

relative to the V-optimal design ξ^* . The efficiency of the design ξ can be interpreted as the proportion of patients needed, when the V-optimal design ξ^* is used, to obtain the same value of the V-criterion as for the design ξ under comparison. Note that the efficiency $\text{eff}(\xi)$ may depend on the parameters through both the non-linearity of the models as well as the presence of random effects.

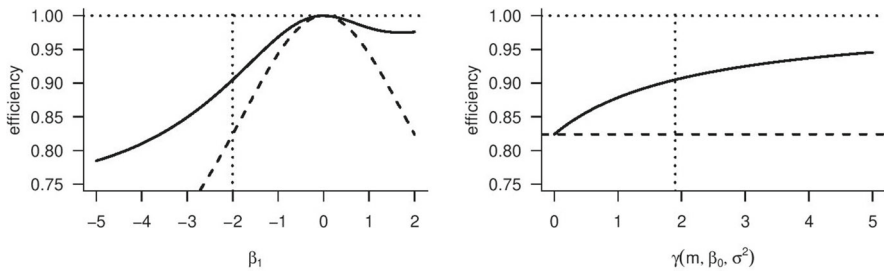


Fig. 3 Efficiency of ξ_0 in dependence on β_1 (left panel) and γ (right panel) in the random intercept (solid line) and the fixed effects (dashed line) model for a binary regressor (Example 4)

5.1 Random intercepts

First, we consider the efficiency for models with random intercepts. The following theorem describes the dependence upon the parameters.

Theorem 4 For model (5), the efficiency depends on β_1 and on the remaining parameters only through $\gamma = m \exp(\beta_0 + \sigma^2/2) (\exp(\sigma^2) - 1) \geq 0$.

Proof By (7), $\text{eff}(\xi)$ can be written as

$$\text{eff}(\xi) = \text{eff}_0(\xi) + (1 - \text{eff}_0(\xi)) \frac{\gamma \check{B}_0}{\check{\phi}(\xi) + \gamma \check{B}_0}, \tag{11}$$

where $\text{eff}_0(\xi) = \phi_0(\xi^*)/\phi_0(\xi) = \check{\phi}(\xi^*)/\check{\phi}(\xi)$ is the efficiency in the corresponding fixed effects model. Since $\text{eff}_0(\xi)$, $\check{\phi}(\xi)$ and \check{B}_0 only vary in terms of β_1 , the claim follows.

Equation (11) provides a lower bound $\text{eff}(\xi) \geq \text{eff}_0(\xi)$ for the efficiency $\text{eff}(\xi)$ in terms of its counterpart in the fixed effects model. This lower bound is attained only if either ξ is optimal in the fixed effects model ($\text{eff}_0(\xi) = 1$) or if there are no random effects ($\sigma^2 = 0$).

Example 4 (Binary regressor). In the situation of Example 1 only two experimental settings $x = 0, 1$ are available. We assume $m = 100$, $\sigma^2 = 1$ and nominal values $\beta_0 = -5$ and $\beta_1 = -2$ for the parameters resulting in $\gamma = 1.909$. For these values, the optimal weight at $x = 1$ is $w^* = 0.269$ (see Example 1), and the efficiency of ξ_0 amounts to $\text{eff}(\xi_0) = 0.905$ compared to its efficiency $\text{eff}_0(\xi_0) = 0.824$ in the fixed effects model. Although the optimal weight depends only on β_1 , the efficiency is also affected by the other parameters through γ .

The efficiency $\text{eff}(\xi_0)$ of the standard design ξ_0 is exhibited in the two panels of Fig. 3 in dependence on β_1 and γ while the other parameter is held fixed to its nominal value, respectively, together with $\text{eff}_0(\xi_0)$. For orientation, the nominal values are indicated there by vertical dotted lines in the corresponding panels.

As can be seen from Fig. 3, the efficiency of ξ_0 increases in β_1 for negative values of β_1 , attains its maximum 1 at $\beta_1 = 0$ where the design ξ_0 is (locally) V-optimal (see

Example 1), and initially decreases for positive values of β_1 . With respect to the other parameters β_0, σ^2 , and m , the efficiency is increasing through γ .

In the situation of Example 2 with a continuous regressor on $\mathcal{X} = [0, 1]$, the behavior of the efficiency is qualitatively the same as in Example 4 and will, hence, not be reported in detail here.

5.2 Random slopes

For models with random slopes, we consider again a binary regressor as in Subject. 4.2. Then the parameter dependence of the efficiency can be characterized as follows.

Theorem 5 *Let $\mathcal{X} = \{0, 1\}$. For model (6), the efficiency depends on β_1 and σ^2 and on the remaining parameters only through $\gamma = m \exp(\beta_0) \geq 0$.*

Proof According to the proof of Theorem 3 and with the notation used there, the efficiency equals

$$\begin{aligned} \text{eff}(\xi) &= \frac{\check{\phi}(\xi^*) + \gamma(\exp(\sigma^2) - 1)\check{B}_2}{\check{\phi}(\xi) + \gamma(\exp(\sigma^2) - 1)\check{B}_2} \\ &= \check{\text{eff}}(\xi) + \left(1 - \check{\text{eff}}(\xi)\right) \frac{\gamma(\exp(\sigma^2) - 1)\check{B}_2}{\check{\phi}(\xi) + \gamma(\exp(\sigma^2) - 1)\check{B}_2} \end{aligned} \tag{12}$$

where $\check{\text{eff}}(\xi) = \check{\phi}(\xi^*)/\check{\phi}(\xi)$ is a reduced version of the efficiency ignoring the additive term $\text{tr}(\mathbf{C}\mathbf{B})$ in the criterion ϕ . Since $\check{\text{eff}}(\xi), \check{\phi}(\xi)$ and \check{B}_2 only vary in terms of β_1 and σ^2 , the claim follows.

Example 5 (Binary regressor). For the situation of the random slopes model in Example 3, we assume $m = 100, \sigma^2 = 1$ and nominal values $\beta_0 = -5$ and $\beta_1 = -2$ the parameters resulting in $\gamma = 0.674$ similar to Example 4. For these values, the optimal weight at $x = 1$ is $w^* = 0.321$ (see Example 3), and the efficiency of ξ_0 amounts to $\text{eff}(\xi_0) = 0.916$. Although the optimal weight depends only on β_1 and σ^2 , the efficiency is also affected by the other parameters through γ .

In Fig. 4 we present a graphical view of the efficiency $\text{eff}(\xi_0)$ for the standard design ξ_0 in dependence on β_1, σ^2 and γ while the other parameters are held fixed to their nominal values, respectively. As before, the corresponding nominal values are indicated in the panels by vertical dotted lines.

In this figure, it can be seen that the efficiency of ξ_0 increases in β_1 up to $\beta_1 = -\sigma^2/2$, where the efficiency attains its maximum value 1 because ξ_0 is (locally) V-optimal for this parameter combination (see Example 3). For larger values of β_1 the efficiency is initially decreasing. With respect to σ^2 , the efficiency is increasing up to $\sigma^2 = -2\beta_1$, where again ξ_0 is (locally) V-optimal and the efficiency becomes 1. For larger values of σ^2 the efficiency stays close to 1. Lastly, the efficiency is slightly increasing in m and β_0 through γ .

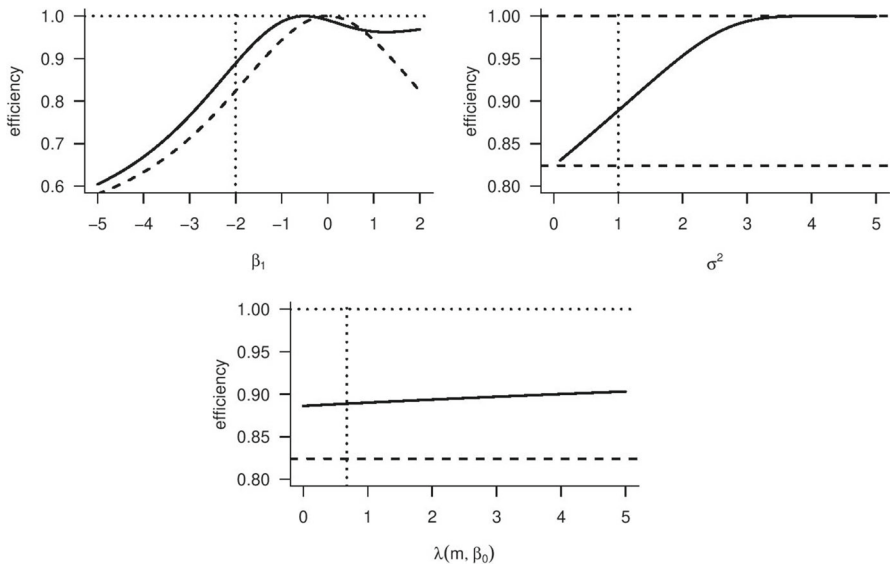


Fig. 4 Efficiency of ξ_0 in dependence on β_1 (top left panel), σ^2 (top right panel) and γ (bottom panel) in the random slope model for a binary regressor (Example 5)

6 Discussion

The main purpose of this paper is to illustrate the problem of optimal design for prediction in the Poisson regression model with random effects using V-criterion. We apply the quasi-likelihood technique to derive the information matrix, which is applied to characterize V-criterion for this model. This criterion has been discussed in detail for the Poisson model with a single linear component on specific experimental regions, including binary and continuous. We also investigated the equivalence theorem to confirm the optimality of obtained designs. Besides, we discussed efficiency for the saturated standard design in terms of the different values of parameters.

Appendix A Proof of the convexity of the V-criterion

Let ξ_1 and ξ_2 be two designs. For any $\alpha \in [0, 1]$, Niaparast and Schwabe (2013) showed

$$\mathfrak{M}_\beta((1 - \alpha)\xi_1 + \alpha\xi_2) \geq (1 - \alpha)\mathfrak{M}_\beta(\xi_1) + \alpha\mathfrak{M}_\beta(\xi_2) \tag{B1}$$

in the sense of Loewner ordering of non-negative definiteness. This requires that the matrices C_ξ of variance correction terms are non-negative definite which follows from the corresponding property of C_i mentioned in Sect. 2.

By standard inversion formulas the convexity of the inverse follows,

$$\mathfrak{M}_\beta^{-1}((1 - \alpha)\xi_1 + \alpha\xi_2) \leq (1 - \alpha)\mathfrak{M}_\beta^{-1}(\xi_1) + \alpha\mathfrak{M}_\beta^{-1}(\xi_2).$$

Hence, because $\text{tr}(\mathbf{A}\mathbf{B})$ is linear in \mathbf{A} , the V-criterion is a convex functional in ξ .

Appendix B Proof of Theorem 1

In order to obtain an equivalence theorem we consider the directional derivative of the trace $\phi(\xi) = \text{tr}(\mathfrak{M}_\beta^{-1}(\xi)\mathbf{B})$. The directional derivative of $\phi(\xi)$ in the direction of ξ' , $F_\phi(\xi, \xi')$ is

$$F_\phi(\xi, \xi') = \frac{d}{d\alpha} \text{tr}((\mathfrak{M}_\beta^{-1}((1 - \alpha)\xi + \alpha\xi'))\mathbf{B})|_{\alpha=0^+} \tag{B1}$$

Since $(\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} = \mathbf{A}_\xi - \mathbf{A}_\xi(\mathbf{I} + \mathbf{C}_\xi\mathbf{A}_\xi)^{-1}\mathbf{C}_\xi\mathbf{A}_\xi$ where \mathbf{I} is the identity matrix, the quasi-information matrix in Eq. (2) can be represented as

$$\mathfrak{M}_\beta(\xi) = \mathbf{F}_\xi^T \left(\mathbf{A}_\xi - \mathbf{A}_\xi(\mathbf{I} + \mathbf{C}_\xi\mathbf{A}_\xi)^{-1}\mathbf{C}_\xi\mathbf{A}_\xi \right) \mathbf{F}_\xi. \tag{B2}$$

As in Niaparast and Schwabe (2013) we will derive a representation of the quasi-information matrix of the convex combination of ξ and ξ' . For this we define the weighted joint intensity matrix $\mathbf{A}_{\xi, \xi'}(\alpha) = \begin{pmatrix} (1 - \alpha)\mathbf{A}_\xi & 0 \\ 0 & \alpha\mathbf{A}_{\xi'} \end{pmatrix}$ for two designs ξ and ξ' , $0 \leq \alpha \leq 1$, where \mathbf{A}_ξ and $\mathbf{A}_{\xi'}$ are diagonal matrices with the weighted response means $(m_{j\mu_j})$ as diagonal elements corresponding to ξ and ξ' respectively, $\mathbf{F}_{\xi, \xi'} = \begin{pmatrix} \mathbf{F}_\xi^T & \mathbf{F}_{\xi'}^T \end{pmatrix}^T$ the joint reduced design matrix for the designs ξ and ξ' and by $\mathbf{C}_{\xi, \xi'} = \begin{pmatrix} \mathbf{C}_\xi & \mathbf{\Gamma}_{\xi, \xi'} \\ \mathbf{\Gamma}_{\xi, \xi'}^T & \mathbf{C}_{\xi'} \end{pmatrix}$ the combined correction matrix, which contains the mixed correction terms for ξ and ξ' in $\mathbf{\Gamma}_{\xi, \xi'} = (c(x, x'))$, where x and x' are the support points of ξ and ξ' , respectively.

Then regarding (B2), the quasi-information matrix of the convex combination of ξ and ξ' is as follow

$$\begin{aligned} \mathfrak{M}_\beta((1 - \alpha)\xi + \alpha\xi') &= \mathbf{F}_{\xi, \xi'}^T [\mathbf{A}_{\xi, \xi'}(\alpha) \\ &\quad - \mathbf{A}_{\xi, \xi'}(\alpha)(\mathbf{I} + \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha))^{-1}\mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha)] \mathbf{F}_{\xi, \xi'}. \end{aligned}$$

Then,

$$\begin{aligned} F_\phi(\xi, \xi') &= \text{tr} \left(\frac{d}{d\alpha} (\mathbf{F}_{\xi, \xi'}^T [\mathbf{A}_{\xi, \xi'}(\alpha) - \mathbf{A}_{\xi, \xi'}(\alpha)(\mathbf{I} + \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha))^{-1} \right. \\ &\quad \left. \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha)] \mathbf{F}_{\xi, \xi'}) \mathbf{B} \right) |_{\alpha=0^+} \\ &= \text{tr} [-\mathfrak{M}_\beta^{-1}((1 - \alpha)\xi + \alpha\xi') (\mathbf{F}_{\xi, \xi'}^T ([\mathbf{A}'_{\xi, \xi'}(\alpha) - \mathbf{A}'_{\xi, \xi'}(0) \\ &\quad (\mathbf{I} + \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha))^{-1}\mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha) + \mathbf{A}_{\xi, \xi'}(\alpha) \\ &\quad (\mathbf{I} + \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha))^{-1}\mathbf{C}_{\xi, \xi'}\mathbf{A}'_{\xi, \xi'}(\alpha)(\mathbf{I} + \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha))^{-1} \\ &\quad \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha) - \mathbf{A}_{\xi, \xi'}(\alpha)(\mathbf{I} + \mathbf{C}_{\xi, \xi'}\mathbf{A}_{\xi, \xi'}(\alpha))^{-1}\mathbf{C}_{\xi, \xi'}\mathbf{A}'_{\xi, \xi'}(\alpha)] \end{aligned}$$

$$F_{\xi, \xi'} \mathfrak{M}_\beta^{-1}((1 - \alpha)\xi + \alpha\xi')B|_{\alpha=0^+}, \tag{B3}$$

where $A'_{\xi, \xi'}(\alpha)$ is the derivative of $A_{\xi, \xi'}(\alpha)$ w.r.t. α . We obtain for $\alpha = 0$ that

$$(I + C_{\xi, \xi'} A_{\xi, \xi'}(0)) = \begin{pmatrix} I + C_\xi A_\xi & 0 \\ \Gamma_{\xi, \xi'}^T A_\xi & I \end{pmatrix}$$

is lower block triangular as well its inverse

$$(I + C_{\xi, \xi'} A_{\xi, \xi'}(0))^{-1} = \begin{pmatrix} (I + C_\xi A_\xi)^{-1} & 0 \\ -\Gamma_{\xi, \xi'}^T A_\xi (I + C_\xi A_\xi)^{-1} & I \end{pmatrix}.$$

We obtain by multiplication of the block matrix

$$F_{\xi, \xi'}^T A'_{\xi, \xi'}(0) F_{\xi, \xi'} = -F_\xi^T A_\xi F_\xi + F_{\xi'}^T A_{\xi'} F_{\xi'}$$

and also we obtain

$$F_{\xi, \xi'}^T A'_{\xi, \xi'}(0) (I + C_{\xi, \xi'} A_{\xi, \xi'}(0))^{-1} C_{\xi, \xi'} A_{\xi, \xi'}(0) F_{\xi, \xi'} \\ = -F_\xi^T A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi F_\xi - F_{\xi'}^T A_{\xi'} \Gamma_{\xi, \xi'}^T A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi F_\xi \\ + F_{\xi'}^T A_{\xi'} \Gamma_{\xi, \xi'}^T A_{\xi'} F_{\xi'}.$$

By the same way we have

$$F_{\xi, \xi'}^T A_{\xi, \xi'}(0) (I + C_{\xi, \xi'} A_{\xi, \xi'}(0))^{-1} C_{\xi, \xi'} A'_{\xi, \xi'}(0) (I + C_{\xi, \xi'} A_{\xi, \xi'}(0))^{-1} C_{\xi, \xi'} A_{\xi, \xi'}(0) F_{\xi, \xi'} \\ = -F_\xi^T A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi F_\xi \\ - F_\xi^T A_\xi (C_\xi A_\xi + I)^{-1} \Gamma_{\xi, \xi'}^T A_{\xi'} \Gamma_{\xi, \xi'}^T A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi F_\xi \\ + F_{\xi'}^T A_{\xi'} (C_\xi A_\xi + I)^{-1} \Gamma_{\xi, \xi'}^T A_{\xi'} \Gamma_{\xi, \xi'}^T A_\xi F_\xi,$$

and also

$$F_{\xi, \xi'}^T A_{\xi, \xi'}(0) (I + C_{\xi, \xi'} A_{\xi, \xi'}(0))^{-1} C_{\xi, \xi'} A'_{\xi, \xi'}(0) F_{\xi, \xi'} \\ = -F_\xi^T A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi F_\xi + F_{\xi'}^T A_{\xi'} (C_\xi A_\xi + I)^{-1} \Gamma_{\xi, \xi'}^T A_{\xi'} F_{\xi'}.$$

Inserting these results into (B3) we obtain

$$F_\phi(\xi, \xi') = \text{tr}([-\mathfrak{M}_\beta^{-1}(\xi)((F_{\xi'} - \Gamma_{\xi, \xi'}^T [A_\xi (C_\xi A_\xi + I)^{-1}] F_\xi)^T A_{\xi'}(F_{\xi'} \\ - \Gamma_{\xi, \xi'}^T [A_\xi - A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi] F_\xi) - F_\xi^T [A_\xi \\ - A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi] F_\xi + F_\xi^T A_\xi (C_\xi A_\xi + I)^{-1} C_\xi \\ [A_\xi - A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi] F_\xi) \mathfrak{M}_\beta^{-1}(\xi) B].$$

By $(A_\xi^{-1} + C_\xi)^{-1} = A_\xi - A_\xi (C_\xi A_\xi + I)^{-1} C_\xi A_\xi$, it follows that

$$F_\phi(\xi, \xi') = \text{tr}([-\mathfrak{M}_\beta^{-1}(\xi)((F_{\xi'} - \Gamma_{\xi, \xi'}^T (A_\xi^{-1} + C_\xi)^{-1} F_\xi)^T \\ A_{\xi'}(F_{\xi'} - \Gamma_{\xi, \xi'}^T (A_\xi^{-1} + C_\xi)^{-1} F_\xi) - F_\xi^T (A_\xi^{-1} + C_\xi)^{-1} F_\xi$$

$$+ \mathbf{F}_\xi^T (\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} \mathbf{C}_\xi (\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} \mathbf{F}_\xi) \mathfrak{M}_\beta^{-1}(\xi) \mathbf{B}).$$

The directional derivative $F_\phi(\xi, \xi')$ is linear in ξ' . Therefore, it suffices to consider one-point designs ξ_x which assign all observations to a single setting x . For such one-point designs the directional derivative reduces to

$$\begin{aligned} F_\phi(\xi, \xi_x) = & \text{tr}([-\mathfrak{M}_\beta^{-1}(\xi)((\mathbf{f}(x) - \mathbf{F}_\xi^T (\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} \mathbf{c}_{\xi,x}) m \mu(x) \\ & (\mathbf{f}(x) - \mathbf{F}_\xi^T (\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} \mathbf{c}_{\xi,x})^T - \mathbf{F}_\xi^T (\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} \mathbf{F}_\xi \\ & + \mathbf{F}_\xi^T (\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} \mathbf{C}_\xi (\mathbf{A}_\xi^{-1} + \mathbf{C}_\xi)^{-1} \mathbf{F}_\xi) \mathfrak{M}_\beta^{-1}(\xi) \mathbf{B}). \end{aligned}$$

According to the general equivalence theorem a design ξ^* is optimal, if and only if

$$F_\phi(\xi^*, \xi_x) \geq 0; \quad \forall x \in \mathcal{X}$$

Hence, the result follows.

References

- Atkinson AC, Donev AN, Tobias RD (2007) Optimum experimental designs. Oxford University Press, Oxford
- Bandemer H (1977) Theorie und Anwendung der optimalen Versuchsplanung I. Akademie-Verlag, Berlin
- Cameron AC, Trivedi PK (2013) Regression analysis of count data. Cambridge University Press, Cambridge
- Davidian M, Giltinan D (1995) Nonlinear models for repeated measurement data. Chapman and Hall, London
- Entholzner M, Benda N, Schmelter T, Schwabe R (2005) A note on designs for estimating population parameters. *Biometr Lett-Listy Biometryczne* 42:25–41
- Groß J (2003) Linear regression. Springer, Berlin
- Idais O and Schwabe R (2021) Equivariance and invariance for optimal designs in generalized linear models exemplified by a class of gamma models. *J Stat Theory Pract*
- Kiefer J, Wolfowitz J (1960) The equivalence of two extremum problem. *Can J Math* 12:363–366
- McCullagh P (1983) Quasi-likelihood functions. *Ann Stat* 11:59–67
- McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London
- McCulloch CE, Searle SR (2001) Generalized linear and mixed models. Wiley, New York
- Mielke T (2012) Approximations of the Fisher information for the construction of efficient experimental designs in nonlinear mixed effects models. Ph.D. Thesis. Otto-von-Guericke University Magdeburg
- Mielke T, Schwabe R (2010) Some considerations on the Fisher information in nonlinear mixed effects models. In: Giovagnoli A, Atkinson AC, Torsney B, May C (eds) *mODa9-advances in model-oriented design and analysis*. Physica, Heidelberg, pp 129–136
- Minkin S (1993) Experimental designs for clonogenic assay in chemotherapy. *J Am Stat Assoc* 88:410–420
- Niaparast M (2009) On optimal design for a Poisson regression model with random intercept. *Statist Probab Lett* 79:741–747
- Niaparast M (2010) Optimal designs for mixed effects Poisson regression models. *J Stat Plann Inference*. Ph.D. Thesis. Otto-von-Guericke-University of Magdeburg
- Niaparast M, Schwabe R (2013) Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients. *J Stat Plann Inference* 143:296–306
- Schmelter T (2007) The optimality of single-group designs for certain mixed models. *Metrika* 65:183–193
- Schmidt D (2019) Characterization of c-, L- and ϕ_k -optimal designs for a class of non-linear multiple-regression models. *J Roy Stat Soc B* 81:101–120
- White LV (1973) An extension of the general equivalence theorem to nonlinear models. *Biometrika* 60:345–348

Whittle P (1973) Some general points in the theory of optimal experimental design. *J Roy Stat Soc B* 35:123–130

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.