The Poisson Model with Three Binary Predictors: When are Saturated Designs Optimal?

Ulrike Grasshoff [1], Heinz Holling [2,*], Rainer Schwabe [3]

Abstract. In this paper, Poisson regression models with three binary predictors are considered. These models are applied to rule-based tasks in educational and psychological testing. To efficiently estimate the parameters of these models locally D-optimal designs will be derived. Eight out of all 70 possible saturated designs are proved to be locally D-optimal in the case of active effects. Two further saturated designs which are the classical fractional factorial designs turn out to be locally D-optimal for vanishing effects.

1 Introduction

Many educational and psychological tests, e. g. measuring human abilites, yield count data. Usually, such tests contain items with different difficulties. The difficulties are often determined by certain binary characteristics or rules of the items. In many cases the data of such tests are distributed according to a Poisson distribution. Thus, a Poisson regression model with binary explanatory variables is the adequate statistical model to describe the data of such tests.

Such a Poisson regression model can be considered as a particular case of a generalized linear model with canonical exponential link. This facilitates the calculation of the likelihood and of the information matrix. The latter may serve as a characteristic of the quality of a design, i. e. for the choice of the explanatory variables in experimental settings, because the asymptotic covariance matrix of the maximum likelihood estimator is proportional to the inverse of the information matrix.

For continuous predictors optimal designs have been derived by Rodríguez-Torreblanca and Rodríguez-Díaz (2007) for one explanatory variable and by Russell et al. (2009) in the case of additive linear effects of the explanatory variables in the linear predictor. Wang et al. (2006) derived numerical results, when there is an additional interaction term which describes a synergetic or antagonistic ef-

[1]Humboldt University, School of Business and Economics, Unter den Linden 6, D–10 099 Berlin, Germany
[2]Institute of Psychology, University of Münster, Fliednerstr. 21, D–48 149 Münster, Germany
[3]Otto-von-Guericke University, Institute for Mathematical Stochastics, PF 4120, D–39 016 Magdeburg, Germany
*corresponding author, e-mail: holling@uni-muenster.de

fect. As in all underlying models which are non-linear in their parameters the optimal designs may depend on the true parameter values, which results in the determination of locally optimal designs.

In the case of two binary explanatory variables Grasshoff et al. (2013) characterize optimal designs, when there are additive effects. A similar result has been obtained by Yang et al. (2012) for binary response. For count data this result is extended in Grasshoff et al. (2014) to an arbitrary number K of explanatory variables. In particular they established that certain saturated designs are optimal, when the effect sizes are sufficiently large.

Since saturated designs result in experiments which can be performed quite easily, the natural question arises, which saturated designs may be optimal and under which parameter constellations. In the present note we will show that in the case of $K = 3$ binary predictors only that particular class of saturated designs described in Grasshoff et al. (2014) may be optimal - with the only exceptional situation of vanishing effect sizes, where a 2^{3-1} fractional factorial design is optimal. All other saturated designs cannot be locally optimal under any parameter settings.

2 Model description, information and design

For each observation of counts the response variable Y is assumed to follow a Poisson distribution $\mathrm{Po}(\lambda)$ with intensity $\lambda = \lambda(\mathbf{x};\boldsymbol{\beta})$, which depends on the experimental settings $\mathbf{x}$ of the explanatory variables and a parameter vector $\boldsymbol{\beta}$ of interest describing the effects of these explanatory variables. Under the exponential link this dependence is specified by $\lambda(\mathbf{x};\boldsymbol{\beta}) = \exp(\mathbf{f}(\mathbf{x})^\top\boldsymbol{\beta})$, where $\mathbf{f}$ is a vector of known regression functions.

In the present setting of binary explanatory variables we code these by a base level $x_k = 0$ and an active level $x_k = 1$ for each variable x_k. Here we focus on $K = 3$ predictors with additive effects, hence $\mathbf{x} = (x_1, x_2, x_3) \in \{0, 1\}^3$ and the regression function is $\mathbf{f}(\mathbf{x}) = (1, x_1, x_2, x_3)^\top$. Then the intensity decomposes according to $\lambda(\mathbf{x};\boldsymbol{\beta}) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$. The parameter vector $\boldsymbol{\beta} = (\beta_0, ..., \beta_3)^\top$ has $p = 4$ components, where β_0 is a baseline parameter and β_k is the effect of the kth explanatory variable which results in a relative change of intensity by the factor $\exp(\beta_k)$, when the kth variable is active.

In the present situation of a generalized linear model with canonical link the Fisher information of a single observation equals $\mathbf{M}(\mathbf{x};\boldsymbol{\beta}) = \lambda(\mathbf{x};\boldsymbol{\beta})\,\mathbf{f}(\mathbf{x})\,\mathbf{f}(\mathbf{x})^\top$, which depends on the setting $\mathbf{x}$ and additionally on $\boldsymbol{\beta}$ through the intensity. Under the assumption of independent observations the normalized information matrix is defined by $\mathbf{M}(\xi;\boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i;\boldsymbol{\beta})$ for an exact design ξ consisting of N design points $\mathbf{x}_1, ..., \mathbf{x}_N$.

For analytical purposes we will make use of the concept of approximate designs ξ with mutually different design points $\mathbf{x}_1, ..., \mathbf{x}_n$, say, and corresponding (real valued) weights $w_i = \xi(\mathbf{x}_i) \geq 0$ with $\sum_{i=1}^{n} w_i = 1$ in the spirit of Kiefer (1974). For such an approximate design the information matrix is more generally defined as $\mathbf{M}(\xi;\boldsymbol{\beta}) = \sum_{i=1}^{n} w_i\, \lambda(\mathbf{x}_i;\boldsymbol{\beta})\, \mathbf{f}(\mathbf{x}_i)\, \mathbf{f}(\mathbf{x}_i)^\top$.

As common in generalized linear models the information matrix and, hence, optimal designs will depend on the parameter vector $\boldsymbol{\beta}$. For measuring the quality of a design we will use the popular D-criterion. More precisely, a design ξ will be called locally D-optimal at $\boldsymbol{\beta}$ if it maximizes the determinant of the information matrix $\mathbf{M}(\xi;\boldsymbol{\beta})$.

For the present Poisson model the intensity and, hence, the information is proportional to the baseline intensity $\exp(\beta_0)$, i. e. $\mathbf{M}(\xi;\boldsymbol{\beta}) = \exp(\beta_0)\mathbf{M}_0(\xi;\boldsymbol{\beta})$, where $\mathbf{M}_0(\xi;\boldsymbol{\beta})$ is the information matrix in the standardized situation when $\beta_0 = 0$. Thus design optimization does not depend on β_0, and $\det(\mathbf{M}_0(\xi;\boldsymbol{\beta}))$ has to be maximized only in dependence on β_1,β_2 and β_3, which means that we can assume the standardized case ($\beta_0 = 0$) in the following.

3 Saturated designs

A design is called saturated, if the number n of distinct design points is equal to the number of parameters ($n = p$). For saturated designs it is well-known that the D-optimal weights are uniform ($w_i = 1/p$). Hence, optimization in the class of saturated designs has only to be done with respect to the choice of the settings $\mathbf{x}_1, ..., \mathbf{x}_p$.

In the present situation the dimension is $p = 4$, and we want to characterize which of the saturated designs can be optimal, i. e. which choice $\mathbf{x}_{i_1}, ..., \mathbf{x}_{i_4}$ of four out of the eight possible settings $\mathbf{x}_0, ..., \mathbf{x}_7 \in \{0,1\}^3$ results in a locally D-optimal design for any parameter constellation $\boldsymbol{\beta}$. For notational reasons we enumerate the possible settings $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$ according to the reversed binary number representation $i = \sum_{k=1}^{3} x_{ik} 2^{k-1}$ for the index i by their components x_{ik}. This means for example $\mathbf{x}_0 = (0,0,0)$, $\mathbf{x}_4 = (0,0,1)$ and $\mathbf{x}_7 = (1,1,1)$. We can visualize these eight design points as the vertices of a three dimensional cube with edge length 1 placed in the first octant (see Figure 1). Further we denote by $\lambda_i = \lambda(\mathbf{x}_i;\boldsymbol{\beta})$ the intensities of the eight possible settings.

In total, there are $\binom{8}{4} = 70$ different saturated designs with uniform weights. Denote by $\mathcal{I} = \{i_1, ..., i_4\}$ the index set of settings in such a saturated design and by $\xi_{\mathcal{I}}$ the design itself.

First of all we can exclude 12 of the 70 saturated designs, for which the design points are located on a plane, i. e. a two-dimensional affine subspace, and which, hence, result in a singular information matrix with determinant equal to zero. These excluded designs consist of the six faces of the cube, for example $\mathcal{I} = \{0,1,4,5\}$, and the six diagonal planes connecting two opposite edges, for example $\mathcal{I} = \{0,1,6,7\}$.

The remaining 58 designs can be assigned to four different equivalence classes

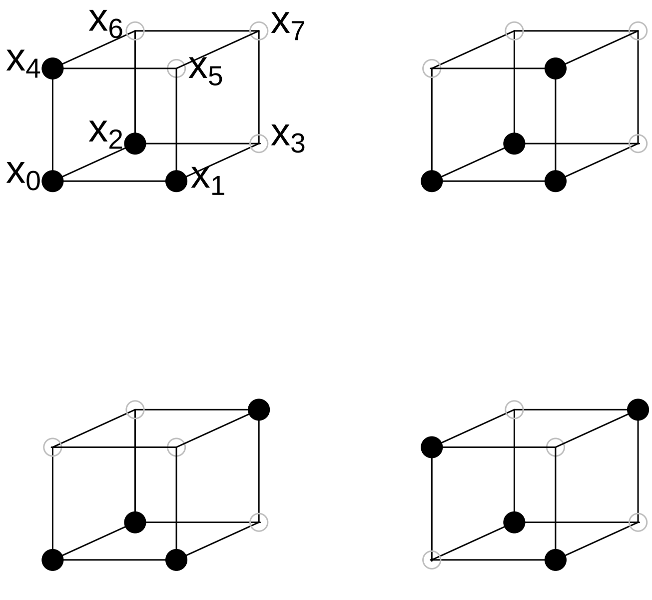

Figure 1: The four relevant types of saturated designs on $\mathcal{X} = \{0,1\}^3$

with respect to permutations of the levels $\{0,1\}$ for each explanatory variable x_k and among the explanatory variables $k = 1,2,3$ themselves. For each of these equivalence classes a representative is exhibited in Figure 1.

The eight saturated designs in the first class (represented in the upper left panel of Figure 1) can be characterized by a vertex $\mathbf{x} = (x_1, x_2, x_3)$ as follows: These "tripod-type" designs contain $\mathbf{x}$ as a central vertex ("head") and additionally all three adjacent vertices ("legs") $(1 - x_1, x_2, x_3)$, $(x_1, 1 - x_2, x_3)$ and $(x_1, x_2, 1 - x_3)$, respectively, as settings and will be denoted by $\xi_{\mathbf{x}}$. For example, the design $\xi_{\mathbf{0}}$ is exhibited in the upper left panel of Figure 1 and is described by the index set $\mathcal{I} = \{0,1,2,4\}$. In Grasshoff et al. (2014) it is shown that the saturated design $\xi_{\mathbf{0}}$ is optimal when all effect sizes are negative and their modulus is sufficiently large. Using symmetry considerations the latter condition can be extended to all "tripod-type" designs $\xi_{\mathbf{x}}$, when the appropriate "head" $\mathbf{x} = (x_1, x_2, x_3)$ is chosen for which $x_k = 0$, when β_k is negative, and $x_k = 1$, when β_k is positive. In particular, this means that $\mathbf{x}$ is the setting with the highest intensity ($\lambda(\mathbf{x}; \boldsymbol{\beta}) = \max_{i=0,\ldots 7} \lambda_i$).

The second class are "snake-type" designs represented in the upper right panel of Figure 1, where the four settings are located on a linear graph not contained in a plane. By symmetry there are 24 saturated designs of that type, and a representative is shown in Figure 1 for $\mathcal{I} = \{0,1,2,5\}$. In the third class the saturated designs consist of three settings on one of the faces of the cube and the isolated opposite vertex as the fourth setting. Also in this class there are 24 different designs, which are equivalent with respect to symmetries. One representative is exhibited in the lower left panel of Figure 1 with $\mathcal{I} = \{0, 1, 2, 7\}$. Saturated designs of these two types can never be locally D-optimal, as will be proved in the subsequent section.

The last two saturated designs are the classical fractional factorial designs. One of these is reported in the lower right panel of Figure 1 , where $\mathcal{I} = \{1, 2, 4, 7\}$. The other fractional factorial design is supported by the complementary four design points ($\mathcal{I} = \{0, 3, 5, 6\}$). These fractional factorial designs are well-known to be D-optimal in the related linear model ($\mathrm{E}(Y(\mathbf{x})) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$) with three binary predictors and no interaction. In the case of no active effects, $\beta_1 = \beta_2 = \beta_3 = 0$, the information matrix of the present Poisson count model turns out to be proportional to that in the linear model for any design ξ. Hence, the fractional factorial designs can be seen to be locally D-optimal in the Poisson model for vanishing effects ($\beta_k = 0$) of the predictors. For all other parameter constellations the fractional factorial cannot be locally D-optimal, as will be indicated in the next section.

4 Proofs of non-optimality

Due to symmetry considerations it suffices to show that the representatives given in the previous section cannot be locally D-optimal for any parameter value $\boldsymbol{\beta}$, To do so we will make use of the celebrated Kiefer-Wolfowitz equivalence theorem (see Silvey, 1980) in the version of Fedorov (1972) which incorporates explicitly an intensity function λ.

Denote by $\psi_{\boldsymbol{\beta}}(\mathbf{x}; \xi) = \lambda(\mathbf{x}; \boldsymbol{\beta})\, \mathbf{f}(\mathbf{x})^\top \mathbf{M}(\xi; \boldsymbol{\beta})^{-1} \mathbf{f}(\mathbf{x})$ the sensitivity function for a design ξ given $\boldsymbol{\beta}$. Then the equivalence theorem states that the design ξ is locally D-optimal at $\boldsymbol{\beta}$ if and only if $\psi_{\boldsymbol{\beta}}(\mathbf{x}; \xi) \leq p$ for all possible settings $\mathbf{x}$. Note that, in general, for a saturated design equality is attained,$\psi_{\boldsymbol{\beta}}(\mathbf{x}; \xi) = p$, on its support points $\mathbf{x}_{i_1}, ..., \mathbf{x}_{i_p}$. Hence, the inequality has only to be checked for the remaining settings $\mathbf{x} \neq \mathbf{x}_{i_1}, ..., \mathbf{x}_{i_p}$. In the following we will suppress the dependence on the parameter vector $\boldsymbol{\beta}$ to facilitate the notation.

For a saturated design $\xi_\mathcal{I}$ the information matrix $\mathbf{M}(\xi_\mathcal{I}) = \frac{1}{p} \mathbf{F}_\mathcal{I}^\top \boldsymbol{\Lambda}_\mathcal{I} \mathbf{F}_\mathcal{I}$ can be decomposed into a product of the essential design matrix $\mathbf{F}_\mathcal{I} = (\mathbf{f}(\mathbf{x}_{i_1}), ..., \mathbf{f}(\mathbf{x}_{i_p}))^\top$ and the diagonal matrix $\boldsymbol{\Lambda}_\mathcal{I} = \mathrm{diag}(\lambda(\mathbf{x}_{i_1}), ..., \lambda(\mathbf{x}_{i_p}))$ of corresponding intensities. Note that these matrices are $p \times p$ square matrices and can be individually inverted. Because of $\mathbf{M}(\xi_\mathcal{I})^{-1} = p\, \mathbf{F}_\mathcal{I}^{-1} \boldsymbol{\Lambda}_\mathcal{I}^{-1} \mathbf{F}_\mathcal{I}^{-\top}$ the sensitivity function simplifies, and the condition of the equivalence theorem reduces to

$$\tfrac{1}{p}\, \psi(\mathbf{x}_i; \xi_\mathcal{I}) = \lambda(\mathbf{x}_i)\, (\mathbf{F}_\mathcal{I}^{-\top} \mathbf{f}(\mathbf{x}_i))^\top \boldsymbol{\Lambda}_\mathcal{I}^{-1} (\mathbf{F}_\mathcal{I}^{-\top} \mathbf{f}(\mathbf{x}_i)) \leq 1$$

for all $i \notin \mathcal{I}$, where $\mathbf{F}^{-\top}$ denotes the inverse of the transpose of $\mathbf{F}$.

For the present situation we recall that $p = 4$ and $\lambda(\mathbf{x}_i) = \lambda_i$, where $\lambda_0 = 1$, $\lambda_3 = \lambda_1\lambda_2$, $\lambda_5 = \lambda_1\lambda_4$, $\lambda_6 = \lambda_2\lambda_4$ and $\lambda_7 = \lambda_1\lambda_2\lambda_4$.

Now for the "snake-type" design we consider the representative $\xi_{\mathcal{I}}$ of Figure 1 specified by $\mathcal{I} = \{0, 1, 2, 5\}$. For this design the essential design matrix and its inverse are given by

$$\mathbf{F}_{\mathcal{I}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{F}_{\mathcal{I}}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

Then $\mathbf{F}_{\mathcal{I}}^{-\top}\mathbf{f}(\mathbf{x}) = (1 - x_1 - x_2, x_1 - x_3, x_2, x_3)^\top$, and the conditions of the equivalence theorem become

$$\lambda(\mathbf{x})((1 - x_1 - x_2)^2 + (x_1 - x_3)^2/\lambda_1 + x_2^2/\lambda_2 + x_3^2/\lambda_5) \leq 1.$$

In the case $\mathbf{x}_3 = (1, 1, 0)$ this condition reduces to $\lambda_1\lambda_2 + \lambda_1 + \lambda_2 \leq 1$, and in the case $\mathbf{x}_4 = (0, 0, 1)$ the condition is equivalent to $\lambda_1\lambda_4 + \lambda_4 + 1 \leq \lambda_1$. This leads to a contradiction, because these conditions require $\lambda_1 < 1$ for $\mathbf{x}_3$ and $\lambda_1 > 1$ for $\mathbf{x}_4$, respectively. Consequently there exists no $\boldsymbol{\beta}$ for which the "snake-type" saturated design $\xi_{\mathcal{I}}$ can be locally D-optimal.

Similarly, for the designs with an isolated setting we consider the representative $\xi_{\mathcal{I}}$ of Figure 1 specified by $\mathcal{I} = \{0, 1, 2, 7\}$. The essential design matrix and its inverse are

$$\mathbf{F}_{\mathcal{I}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{F}_{\mathcal{I}}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Then $\mathbf{F}_{\mathcal{I}}^{-\top}\mathbf{f}(\mathbf{x}) = (1 - x_1 - x_2 + x_3, x_1 - x_3, x_2 - x_3, x_3)^\top$, and the conditions of the equivalence theorem become

$$\lambda(\mathbf{x})((1 - x_1 - x_2 + x_3)^2 + (x_1 - x_3)^2/\lambda_1 + (x_2 - x_3)^2/\lambda_2 + x_3^2/\lambda_7) \leq 1.$$

In the case $\mathbf{x}_3 = (1, 1, 0)$ this condition reduces again to $\lambda_1\lambda_2 + \lambda_1 + \lambda_2 \leq 1$, and in the case $\mathbf{x}_5 = (1, 0, 1)$ the condition is equivalent to $\lambda_1\lambda_2\lambda_4 + \lambda_1\lambda_4 + 1 \leq \lambda_2$. This leads to a contradiction, because these conditions require $\lambda_2 < 1$ for $\mathbf{x}_3$ and $\lambda_2 > 1$ for $\mathbf{x}_5$, respectively. Consequently there exists no $\boldsymbol{\beta}$ for which the saturated design $\xi_{\mathcal{I}}$ with an isolated setting can be locally D-optimal.

Finally, for the fractional factorial design $\xi_{\mathcal{I}}$ of Figure 1, which is specified by $\mathcal{I} = \{1, 2, 4, 7\}$, the essential design matrix and its inverse are

$$\mathbf{F}_{\mathcal{I}} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{F}_{\mathcal{I}}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix}.$$

For $i = 0, 3, 5$, and 6 the conditions of the equivalence theorem are equivalent to

$$\lambda_i\,(\lambda_2\lambda_4 + \lambda_1\lambda_4 + \lambda_1\lambda_2 + 1) \leq 4\lambda_1\lambda_2\lambda_4\,.$$

For $i = 3$ this condition reduces to $\lambda_2\lambda_4 + \lambda_1\lambda_4 + \lambda_1\lambda_2 + 1 \leq 4\lambda_4$, which requires $\lambda_4 > 1/4$. If we rearrange this inequality to $\lambda_2(\lambda_4 + \lambda_1) \leq 4\lambda_4 - \lambda_1\lambda_4 - 1$. the right hand side provides an upper bound $\lambda_1 < (4\lambda_4 - 1)/\lambda_4$ for λ_1. Similarly, if we rearrange the condition for $i = 0$, then $\lambda_1\lambda_4 + 1 \leq \lambda_2(4\lambda_1\lambda_4 - \lambda_1 - \lambda_4)$ provides a lower bound $\lambda_1 > \lambda_4/(4\lambda_4 - 1)$. These two bounds can only be satisfied simultaneously if $\lambda_4 > 1/3$.

Moreover, the above conditions yield the following inequalities

$$\frac{\lambda_1\lambda_4 + 1}{4\lambda_1\lambda_4 - \lambda_1 - \lambda_4} \leq \lambda_2 \leq \frac{4\lambda_4 - \lambda_1\lambda_4 - 1}{\lambda_1 + \lambda_4}\,,$$

which cannot be satisfied simultaneously if $1/3 < \lambda_4 < 1$.

Analogously, from the conditions for $i = 5$ and 6, respectively, we obtain the inequality $1/(4 - \lambda_4) < 4 - \lambda_4$, which requires $\lambda_4 < 3$, and

$$\frac{\lambda_1\lambda_4 + 1}{4 - \lambda_1 - \lambda_4} \leq \frac{4\lambda_1 - \lambda_1\lambda_4 - 1}{\lambda_1 + \lambda_4}\,.$$

This inequality cannot be satisfied for $1 < \lambda_4 < 3$. Hence, local D-optimality can only be achieved, when $\lambda_4 = 1$, i.e. $\beta_3 = 0$.

Symmetry considerations lead to the same result for λ_1 and λ_2. As a consequence fractional factorial designs are locally D-optimal only if $\beta_1 = \beta_2 = \beta_3 = 0$.

Acknowledgement. This work was partly supported by grant Ho1286-6 of the Deutsche Forschungsgemeinschaft.

References

Fedorov, V.V. (1972): *Theory of Optimal Experiments.* Academic Press, New York.

Graßhoff, U., Holling, H. and Schwabe, R. (2013): Optimal design for count data with binary predictors in item response theory. In: D. Uciński, A.C. Atkinson and M. Patan (Eds): *mODa10 -Advances in Model-Oriented Design and Analysis*, Springer, Cham, 117–124.

Graßhoff, U., Holling, H. and Schwabe, R. (2014): Optimal Design for the Rasch Poisson Counts Model with Multiple Binary Predictors. *Technical Report.*

Kiefer, J. (1974): General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, **2**, 849–879.

Rodríguez-Torreblanca, C. and Rodríguez-Díaz, J.M. (2007): Locally D- and c-optimal designs for Poisson and negative binomial regression models. *Metrika*, **66**, 161-172.

Russell, K., Woods, D., Lewis, S. and Ecclestone, J. (2009): D-optimal designs for Poisson regression models. *Statistica Sinica*, **19**, 721–730.

Schwabe, R. (1996): *Optimum Designs for Multi-Factor Models.* Springer, New York.

Silvey, D. (1980): *Optimal Design.* Chapman & Hall, London.

Wang Y., Myers R.H., Smith E.P. and Ye K. (2006): D-optimal Designs for Poisson Regression Models. it Journal of Statistical Planning and Inference, **136**, 2831–2845.

Yang, J., Mandal, A. and Majumdar, D. (2012): Optimal designs for two-level factorial experiments with binary response. *Statistica Sinica*, **22**, 885–907.