Optimal Design for Count Data with Binary Predictors in Item Response Theory

U. Graßhoff, H. Holling and R. Schwabe

Abstract The Rasch Poisson counts model (RPCM) allows for the analysis of mental speed which represents a basic component of human intelligence. An extended version of the RPCM, which incorporates covariates in order to explain the difficulty, provides a means for modern rule-based item generation. After a short introduction into the extended RPCM we will develop locally D-optimal calibration designs for this model. Therefore, the RPCM is embedded in a particular generalized linear model. Finally, the robustness of the derived designs will be investigated.

1 Introduction

Reasoning, memory, creativity and mental speed belong to the most important factors of human intelligence (Jäger, 1984). Mental speed refers to the human ability to carry out mental processes, required for the solution of a cognitive task, at variable rates or increments of time. Usually, mental speed is measured by elementary tasks with low cognitive demands in which speed of response is primary. As Rasch

U. Graßhoff
Institut für Mathematische Stochastik, Otto-von-Guericke-Universität Magdeburg, PF 4120, 39016 Magdeburg, Germany, and Psychologisches Institut IV, Westfälische Wilhelms-Universität Münster, Fliednerstr. 21, 48149 Münster, Germany, e-mail: ulrike.grasshoff@ovgu.de

H. Holling
Psychologisches Institut IV, Westfälische Wilhelms-Universität Münster, Fliednerstr. 21, 48149 Münster, Germany, e-mail: holling@uni-muenster.de

R. Schwabe
Institut für Mathematische Stochastik, Otto-von-Guericke-Universität Magdeburg, PF 4120, 39016 Magdeburg, Germany, e-mail: rainer.schwabe@ovgu.de

(1960) already showed in his classical monograph, elementary cognitive tasks can be analyzed by the so-called Rasch Poisson counts model. Other successful applications of this model have been published by e. g. Jansen (1997) and Verhelst and Kamphuis (2009).

Typical items measuring mental speed can be differentiated by task characteristics or rules that correspond to cognitive operations to solve an item. The kind and amount of task characteristics influences the difficulty of the items. The task characteristics can be used to predict the task difficulty analogously to linear logistic models for reasoning items (Graßhoff et al., 2010).

2 Poisson model for count data

According to the Rasch Poisson count model the number of correct answers is assumed to follow a Poisson distribution with intensity $\lambda = \theta\sigma$, where θ is the ability of the test person and σ is the easiness of the test item. Obviously the (expected) number of correct answers will increase simultaneously with the ability of the person and the easiness of the task.

In the following we consider the calibration step for the test items, when the ability of the test person is assumed to be known. The dependence of the easiness of an item on the rules may then be specified by a Poisson regression (Poisson anova) model with exponential link.

More formally, the number of correct answers $Y(\mathbf{x})$ is Poisson distributed with intensity $\lambda(\mathbf{x};\beta) = \theta\exp(\mathbf{f}(\mathbf{x})^\top\beta)$, where $\mathbf{x}$ is the experimental setting ("rules"), which may be chosen from a specific experimental region $\mathscr{X}$, $\sigma = \exp(\mathbf{f}(\mathbf{x})^\top\beta)$ is the easiness of the item, $\mathbf{f} = (f_1,...,f_p)^\top$ is a vector of known regression functions, and $\beta \in \mathbb{R}^p$ the vector of unknown parameters to be estimated.

As rules may be applied or not, we will focus on the situation of a K-way layout with binary explanatory variables x_k, where $x_k = 1$, if the kth rule is applied, and $x_k = 0$ otherwise. In particular, if $x_k = 0$ for all rules k, a basic item is presented. The experimental setting is then $\mathbf{x} = (x_1,...,x_k) \in \{0,1\}^K$. As we assume no interactions, the vector of regression functions is $\mathbf{f}(\mathbf{x}) = (1,x_1,x_2,...,x_k)^\top$, and the parameter vector β consists of a constant term β_0 and the K main effects β_k. Thus $p = K+1$ and the expected response equals the intensity $\lambda(\mathbf{x};\beta) = \theta\exp(\beta_0 + \sum_{k=1}^K \beta_k x_k)$.

3 Information and design

For a single observation the Fisher information is $\mathbf{M}(\mathbf{x};\beta) = \lambda(\mathbf{x};\beta)\,\mathbf{f}(\mathbf{x})\,\mathbf{f}(\mathbf{x})^\top$, which depends on the particular setting $\mathbf{x}$ and additionally on β through the intensity. Consequently, the normalized information matrix equals $\mathbf{M}(\xi;\beta) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i;\beta)$ for an exact design ξ consisting of N design points $\mathbf{x}_1,...,\mathbf{x}_N$. For analytical ease we will make use 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). This approach seems appropriate, as typically the number N of items presented may be quite large. The information matrix is then more generally defined as $\mathbf{M}(\xi;\beta) = \sum_{i=1}^{n} w_i\,\lambda(\mathbf{x}_i;\beta)\,\mathbf{f}(\mathbf{x}_i)\,\mathbf{f}(\mathbf{x}_i)^\top$.

As common in generalized linear models the information matrix and, hence, optimal designs will depend on the parameter vector β. 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 βif it maximizes the determinant of the information matrix $\mathbf{M}(\xi;\beta)$.

In the present situation the intensity and, hence, the information is proportional to θ and $\exp(\beta_0)$ such that $\mathbf{M}(\xi;\beta) = \theta\exp(\beta_0)\mathbf{M}_0(\xi;\beta)$, where $\mathbf{M}_0(\xi;\beta)$ is the information matrix in the standardized situation $\theta = 1$ and $\beta_0 = 0$. Thus for a fixed person only $\det(\mathbf{M}_0(\xi;\beta))$ has to be optimized, and we will assume the standardized case ($\theta = 1$, $\beta_0 = 0$) without loss of generality throughout the remainder of the paper. If more than one test person is involved, then the same optimal design has to be applied to each of them. Note also that in the case that the choice of the test persons is at the disposition of the examiner the person with the highest ability provides the most information.

4 Two way-layout with binary predictors

Before starting the case of a two-way layout we notice that for the situation of only one rule ($K = 1$) the D-optimal design assigns equal weights $w_i^* = 1/2$ to the only two possible settings $x_1 = 1$ of application of the rule and $x_2 = 0$ of the basic item independently of β, as all (regular) designs are saturated.

Our main focus, however, is on $K = 2$ binary explanatory variables, where the number of parameters equals $p = 3$. Here the four possible settings are $\mathbf{x}_1 = (1,1)$, where both rules are applied, $\mathbf{x}_2 = (1,0)$ and $\mathbf{x}_3 = (0,1)$, where either only the first or the second rule is used, respectively, and $\mathbf{x}_4 = (0,0)$ for the basic item. Hence, any design ξ is completely determined by the corresponding weights $w_1,...,w_4$. Further we denote by $\lambda_i = \lambda(\mathbf{x}_i;\beta)$ the related intensities. Then the information matrix of a design ξ results in

$$\mathbf{M}(\xi;\beta) = \begin{pmatrix} \sum_{i=1}^4 w_i\lambda_i & w_1\lambda_1 + w_2\lambda_2 & w_1\lambda_1 + w_3\lambda_3 \\ w_1\lambda_1 + w_2\lambda_2 & w_1\lambda_1 + w_2\lambda_2 & w_1\lambda_1 \\ w_1\lambda_1 + w_3\lambda_3 & w_1\lambda_1 & w_1\lambda_1 + w_3\lambda_3 \end{pmatrix}$$

with a determinant equal to

$$\begin{aligned} \det(\mathbf{M}(\xi;\beta)) &= w_1w_2w_3\lambda_1\lambda_2\lambda_3 + w_1w_2w_4\lambda_1\lambda_2\lambda_4 \\ &\quad + w_1w_3w_4\lambda_1\lambda_3\lambda_4 + w_2w_3w_4\lambda_2\lambda_3\lambda_4 . \end{aligned}$$

Candidates for optimal designs will be either saturated designs on any three of these settings with corresponding weights $w_i = 1/3$ or "true" four-point designs with suitable positive weights for all four settings. As we will see later, all these cases may occur corresponding to the values for the effect sizes β_1 and β_2 of the two rules. For the saturated designs denote by ξ_{ij} the uniform three-point design on the setting (i, j) and its two adjacent settings $(i, 1-j)$ and $(1-i, j)$ for $i, j = 0, 1$. For example ξ_{00} is then the uniform design on $(0,0)$, $(0,1)$ and $(1,0)$.

For the present application it is reasonable to investigate the case $\beta_1 \leq 0$ and $\beta_2 \leq 0$, as it is to be expected that the application of a rule increases the difficulty and, hence, decreases the easiness of an item. Other parameter constellations can be treated by symmetry considerations.

Russell et al. (2009) treated the situation of continuous predictors. From their result we may conclude that in our setting the design ξ_{00}, which avoids the most difficult item, is locally *D*-optimal for $\beta_1 = \beta_2 = 2$.

For other non-positive values of β_1 and β_2 we can derive that the design ξ_{00} is locally *D*-optimal if and only if $\lambda_2\lambda_3\lambda_4 - \lambda_1\lambda_2\lambda_4 - \lambda_1\lambda_3\lambda_4 - \lambda_1\lambda_2\lambda_3 \geq 0$ by the celebrated equivalence theorem, see e. g. Silvey (1980). This condition is fulfilled if and only if $\beta_2 \leq \log((1-\exp(\beta_1))/(1+\exp(\beta_1)))$. Otherwise a "true" four-point design will be optimal.

By considerations of equivariance similar conditions can be derived for the other sign combinations in β, and we can state that some saturated design is locally *D*-optimal if and only if $|\beta_2| \geq \log((\exp(|\beta_1|)+1)/(\exp(|\beta_1|)-1))$.

In Figure 1 the parameter regions of β_1 and β_2 are depicted, where which saturated design is locally *D*-optimal. From this picture it can be seen that saturated designs are optimal, if the effect sizes are large, and then that level combination is avoided, which results in the lowest intensity. Conversely, for the interior diamond shaped region, where $|\beta_2| < \log((\exp(|\beta_1|)+1)/(\exp(|\beta_1|)-1))$, a "true" four-point design will be locally *D*-optimal. Similar results have been obtained by

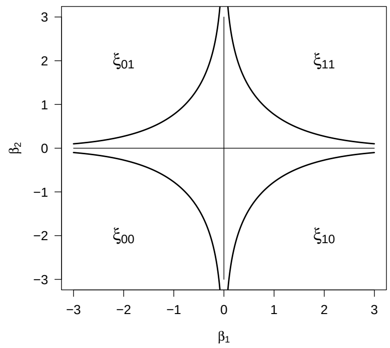

Fig. 1 Locally D-optimal designs in dependence on (β_1, β_2)

Yang et al. (2012) for binary response. In the case of vanishing effects, $\beta_1 = \beta_2 = 0$, the information matrix coincides with the corresponding linear model of a two-way layout, and the uniform design is optimal with weights $w_i = 1/4$ on all four level combinations $\mathbf{x}_1, ..., \mathbf{x}_4$.

Next we will consider two particular parameter constellations, where either one of the effect sizes vanishes or where both effect sizes are equal: For the first case we assume $\beta_1 = 0$. The case $\beta_2 = 0$ can be treated analogously. In this situation the intensity $\lambda(\mathbf{x};\beta)$ is constant in the first component, $\lambda_1 = \lambda_3 = \exp(\beta_2)$ and $\lambda_2 = \lambda_4 = 1$. Hence, according to Theorem 1 in Graßhoff et al. (2004) we obtain an optimal product-type design ξ^* defined by $\xi^*(\mathbf{x}) = \xi_2^*(x_2)/2$ and the marginal weight $v^* = \xi_2^*(1)$ maximizes $v(1-v)(1+(\lambda_1 - 1)v)$. If additionally also $\beta_2 = 0$, then $\lambda_1 = 1$ and the optimal marginal weight is $v^* = 1/2$, form which we recover the optimality of the uniform four-point design. If $\beta_2 \neq 0$, then $\lambda_1 \neq 1$ and the optimal weight can be calculated as

$$v^* \;=\; 1/2 \;+\; (\tau - 2\sqrt{\tau^2 - 3})/(6(\exp(\beta_2/2) - \exp(-\beta_2/2))\,,$$

where $\tau = \exp(\beta_2/2) + \exp(-\beta_2/2)$ Note that $1/3 < v^* < 2/3$. Consequently we get $1/6 < w_i^* < 1/3$ as $w_1^* = w_3^* = v^*/2$ and $w_2^* = w_4^* = (1 - v^*)/2$. The left panel of Figure 2 exhibits these weights as functions of β_2. The weights $w_1^* = w_3^*$ for $\mathbf{x}_1 = (1,1)$ and $\mathbf{x}_3 = (0,1)$ decrease, when β_2 tends to minus infinity, i. e. if these items become more difficult. Hence, more observations should be allocated to the other items $\mathbf{x}_2 = (1,0)$ and $\mathbf{x}_4 = (0,0)$ with lower difficulty.

An alternative parameter constellation, where we can explicitly determine the optimal weights, is the situation of equally sized effect sizes, $|\beta_2| = |\beta_1|$. In particular we consider the case $\beta_2 = \beta_1 = \beta$, which is relevant for our application. The case $\beta_2 = -\beta_1$ can again be treated by symmetry considerations. Here the intensities are $\lambda_1 = \exp(2\beta)$, $\lambda_2 = \lambda_3 = \exp(\beta)$ and $\lambda_4 = 1$. Due to symmetry considera-

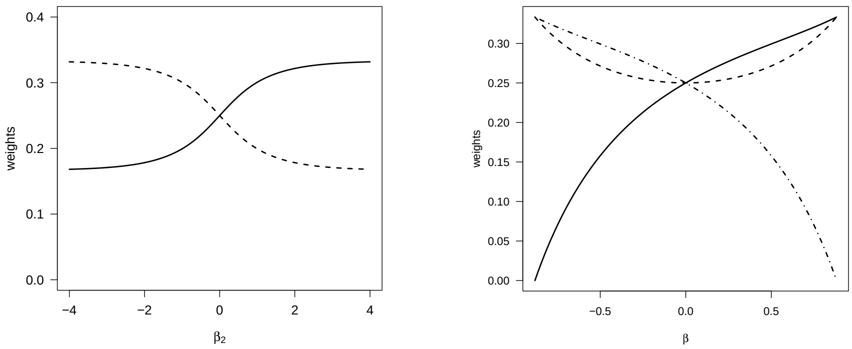

Fig. 2 Optimal weights $w_1^* = w_3^*$ (solid line) and $w_2^* = w_4^*$ (dashed line) for $\beta_1 = 0$ (left panel) and w_1^* (solid line), $w_2^* = w_3^*$ (dashed line) and w_4^* (dashed-dotted line) for $\beta_1 = \beta_2 = \beta$ (right panel)

tions with respect to swapping the factors we can conclude that the optimal weights satisfy $w_2^* = w_3^*$. The saturation condition above leads to $|\beta| \geq \log(\sqrt{2}+1) \approx 0.881$. Hence, for $\beta \leq -\log(1+\sqrt{2})$ the design ξ_{00} is locally D-optimal, while for $\beta \geq \log(1+\sqrt{2})$ this is true for the design ξ_{11}. For the intermediate case, $|\beta| < \log(1+\sqrt{2})$, the determinant is optimized by

$$w_2^* = w_3^* = \left(4\gamma + 2\sqrt{\gamma^2+12}\right) / \left(3(4-\gamma^2)\right),$$

where $\gamma = \exp(\beta) + \exp(-\beta) - 4$, and

$$w_{1,4}^* = 1/2 - w_2^* \pm (\exp(\beta) - \exp(-\beta))\, w_2^*/4.$$

The right panel of Figure 2 presents the weights of the locally D-optimal designs in dependence on β. The passage from an optimal design with four points to an optimal saturated design takes place continuously in the weights at the critical values $\beta = \pm\log(1+\sqrt{2})$, and the symmetry properties of the optimal weights become evident from the picture. Again the uniform four-point design can be recovered to be optimal for the case of vanishing effects ($\beta = 0$).

5 Robustness

Locally D-optimal designs may show a poor performance, if false initial values are specified for the parameters. Therefore a sensitivity analysis has to be performed, and we will compare the efficiency of a saturated design with the efficiency of the uniform four-point design, which is optimal for $\beta_1 = \beta_2 = 0$. As usual the D-efficiency of a design ξ is defined by $\mathrm{eff}(\xi;\beta) = (\det(\mathbf{M}(\xi;\beta)) / \det(\mathbf{M}(\xi_\beta^*;\beta)))^{1/p}$, where ξ_β^* denotes the locally D-optimal design at β and p is the dimension of the parameter vector (here $p = 3$).

In particular, we consider again the saturated design ξ_{00}. In the left panel of Figure 3 the efficiency is exhibited for the situation of one vanishing effect ($\beta_1 = 0$). The efficiency of the saturated design ξ_{00} (solid line) tends to 1 for β_2 to minus infinity and tends to 0, if β_2 goes to plus infinity. The efficiency of the uniform

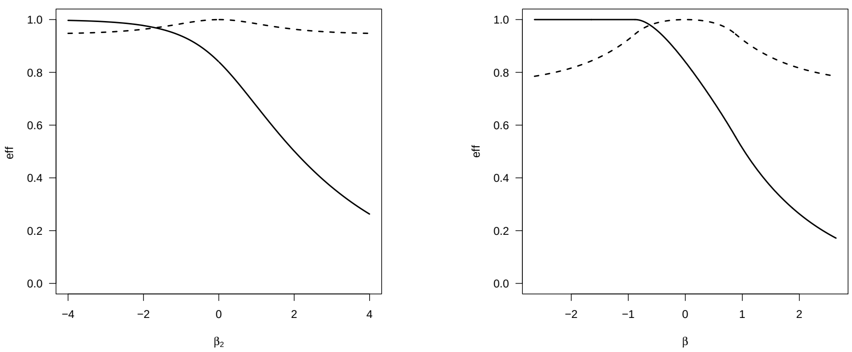

Fig. 3 Efficiencies of the saturated design ξ_{00} (solid line) and the uniform four-point design (dashed line) for $\beta_1 = 0$ (left panel) and for $\beta_1 = \beta_2 = \beta$ (right panel)

four-point design (dashed line) drops from 1 for $\beta_2 = 0$, where hi design is locally optimal, to $(27/32)^{1/3} \approx 0.945$, when $|\beta_2|$ tends to inifinity. In the right panel of Figure 3 the eficiency is plotted for equal effect sizes ($\beta_2 = \beta_1 = \beta$). The saturated design ξ_{00} is locally D-optimal and has, hence, efficiency 1 for $\beta \leq -\log(1+\sqrt{2})$. If β increases beyond this critical value, the efficiency of ξ_{00} decreases, and for $\beta \geq \log(1+\sqrt{2})$ the efficiency equals $\exp(-2\beta)^{1/3}$, which finally drops down to 0. For the efficiency of the uniform four-point design we observe again the value 1 at $\beta = 0$ and a lower bound of $3/4$, which is approached for $|\beta|$ to infinity. Thus the uniform four-point design seems to be essentially more robust to misspecifications of the parameter values than the saturated designs.

Finally, we note that the uniform four-point design is maximin efficient for symmetric parameter regions, which follows from a corresponding result in Graßhoff and Schwabe (2008), as this design is the only invariant design with respect to permutations of the levels. Similar arguments may also establish that the uniform four-point design is also optimal for weighted (“Bayesian”) criteria, when the weight function is symmetric in the parameters.

6 Conclusion

In this article we developed locally *D*-optimal designs for the Rasch Poisson counts model including two binary explanatory variables. If the effect sizes are large, saturated designs proved to be optimal. However, this condition implies, at least, a ratio of $(1+\sqrt{2})^2 \approx 5.83$ between the highest and the lowest intensity. Such a ratio is quite unrealistic in our applications of the RPCM for rule-based testing mental speed. Hence, four-point designs will be mostly required for corresponding calibration studies. For two particular parameter constellations optimal weights have been derived. For these cases it has been shown that uniform four-point designs are very robust. Since rule-based tests of mental speed often include more than two task characteristics, we will, as a next step, develop locally *D*-optimal designs for the RPCM with $K > 2$ binary explanatory variables.

Acknowledgements This research was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant HO 1286/6.

References

Graßhoff, U., H. Großmann, H. Holling, and R. Schwabe (2004). Design optimality for main effects in linear paired comparison models. *Journal of Statistical Planning and Inference 126*, 361–376.

Graßhoff, U., H. Holling, and R. Schwabe (2010). Optimal designs for linear logistic test models. In A. Givovagnoli, A. Atkinson, and B. Torsney (Eds.), *mODa 9 - Advances in Model-Oriented Design and Analysis*, pp. 97 – 104. Berlin: Physica.

Graßhoff, U. and R. Schwabe (2008). Optimal design for the Bradley-Terry paired comparison model. *Statistical Methods and Applications 17*, 275–289.

Jäger, A. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau 35*, 21–35.

Jansen, M. (1997). Rasch's model for reading speed with manifest explanatory variables. *Psychometrika 62*, 393–409.

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

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research.

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

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

Verhelst, N. D. and F. H. Kamphuis (2009). A Poisson-Gamma model for speed tests. Technical Report 2009-2, Cito, Arnhem.

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