Optimal Design for the Rasch Poisson-Gamma Model

Ulrike Graßhoff, Heinz Holling and Rainer Schwabe

Abstract The Rasch Poisson counts model is an important model for analyzing mental speed, an fundamental component of human intelligence. Here, the parameter representing the ability of a person is specified as random with an underlying Gamma distribution. For an extended version of this model which incorporates two binary covariates in order to explain the difficulty of an item we will develop locally *D*-optimal calibration designs.

1 Introduction

Mental speed refers to the human ability to carry out mental processes, required for the solution of a cognitive task and belongs to the most important factors of human intelligence. Usually, mental speed is measured by elementary tasks with low cognitive demands in which speed of response is primary. As Rasch (1960) already showed in his classical monograph, elementary cognitive tasks can be analyzed by the Rasch Poisson counts model. For calibration of the tasks a generalization of the model, the so-called Rasch Poisson-Gamma model, is adequate which incorporates random effects related to the test persons (see e. g. 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

Ulrike Graßhoff
Humboldt University, School of Business and Economics, Unter den Linden 6, D–10 099 Berlin, Germany, e-mail: grasshou@hu-berlin.de

Heinz Holling
Psychologisches Institut IV, Westfälische Wilhelms-Universität, Fliednerstr. 21, D–48 149 Münster, Germany, e-mail: holling@uni-muenster.de

Rainer Schwabe
Otto-von-Guericke University, Institute for Mathematical Stochastics, PF 4120, D–39 016 Magdeburg, Germany, e-mail: rainer.schwabe@ovgu.de

amount of task characteristics influences the difficulty of the items. To determine this influence an optimal choice of the items may help to improve the statistical anlysis.

2 The Poisson-Gamma model for count data

In the Rasch (1960) Poisson count model the number of solved tasks in a test is assumed to follow a Poisson distribution with mean (*intensity*) $\lambda = \theta\sigma$, where the person parameter θ represents the ability of the respondent and the item parameter σ specifies the easiness of the test item. The ability and the easiness can be estimated in a two step procedure: First, in a calibration step, the item parameter is estimated conditional on known person parameters; and in a second step, the person parameter is estimated based on the item parameters previously obtained. In these two steps the items and the persons may be investigated separately. As in Graßhoff et al. (2013) we will consider here the calibration step, but we will now allow the ability θ of a respondent to be random with known prior distribution. Usually in this context a Gamma distribution is assumed for the ability, which leads to the Poisson-Gamma model (see e. g. Verhelst and Kamphuis, 2009).

To be more specific the conditional distribution of the number Y of correct answers given the ability $\Theta = \theta$ is Poisson with mean $\theta\sigma$, and the ability Θ is Gamma distributed with shape parameter $A > 0$ and inverse scale parameter $B > 0$.

Then the unconditional probabilities of Y can be obtained by integration of the joint density $\frac{\theta^y}{y!} e^{-\theta\sigma} \cdot \frac{B^A}{\Gamma(A)} \theta^{A-1} \exp(-B\theta)$ with respect to θ. It is well-known that the resulting distribution is (generalized) negative binomial with success probability $B/(\sigma + B)$ and (generalized) number of successes A which for integer A models the number of failures before the Ath success occurs The corresponding probability function is $P(Y = y) = \frac{\Gamma(y+A)}{y!\Gamma(A)} (\frac{B}{(\sigma+B)})^A (\frac{\sigma}{(\sigma+B)})^y$ with expectation $\mathrm{E}(Y) = \frac{A}{B}\sigma$ and variance $\mathrm{Var}(Y) = \frac{\sigma+B}{B}\mathrm{E}(Y)$. Thus the negative binomial is a common distribution to model overdispersion in count data. For fixed expectation, $A/B = \lambda$ say, the limiting distribution is again Poisson when B tends to infinity.

As in the Poisson count model the easiness σ will be connected with the linear predictor based on the rules applied by the log link, $\sigma = \exp(\mathbf{f}(\mathbf{x})^\top \beta)$, where $\mathbf{x}$ is the experimental setting which may be chosen from a specific experimental region $\mathscr{X}$, $\mathbf{f} = (f_1, ..., f_p)^\top$ is a vector of known regression functions, and β the p-dimensional vector of unknown parameters to be estimated. Hence, the number of correct answers $Y(\mathbf{x})$ is Poisson-Gamma distributed with $E(Y(\mathbf{x})) = \frac{A}{B} \exp(\mathbf{f}(\mathbf{x})^\top \beta)$

The items are generated in such a way that K different rules may be applied or not. Thus the explanatory variables x_k are binary, where $x_k = 1$, if the kth rule is applied, and $x_k = 0$ otherwise, $k = 1, ..., K$. The experimental settings $\mathbf{x} = (x_1, ..., x_K) \in \{0, 1\}^K$ constitute a binary K-way layout. In the particular case $\mathbf{x} = \mathbf{0}$ a basic item is presented. We assume that there are no interactions between the rules. Then the vector of regression functions can be specified by $\mathbf{f}(\mathbf{x}) = (1, x_1, x_2, ..., x_K)^\top$, and the parameter vector $\beta) = (\beta_0, \beta_1, ..., \beta_K)^\top$ of dimension $p = K+1$ consists of a constant term β_0 and the K main effects β_k related to the application of the k rules. In the present application it can be assumed that $\beta_k \leq 0$ for the main effects because the application of a rule will typically increase the difficulty and, hence, decrease the easiness of an item.

Additionally we only consider the situation that each test person receives exactly one item in order to ensure independence of the observations.

3 Information and design

The impact of an experimental setting on the quality of the maximum likelihood estimator of the parameter vector β is measured by the Fisher information matrix $\mathbf{M}(\mathbf{x};\beta)$. Similar to generalized linear models the calculation of the Fisher information results in $\mathbf{M}(\mathbf{x};\beta) = \lambda(\mathbf{x};\beta)\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top$, which depends on the particular setting $\mathbf{x}$ and additionally on β through the intensity

$$\lambda(\mathbf{x};\beta) = A\frac{\exp(\mathbf{f}(\mathbf{x})^\top\beta)}{\exp(\mathbf{f}(\mathbf{x})^\top\beta)+B}. \tag{1}$$

Consequently, for an exact design ξ consisting of N design points $\mathbf{x}_1, ..., \mathbf{x}_N$, the normalized information matrix equals $\mathbf{M}(\xi;\beta) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{M}(\mathbf{x}_i;\beta)$. For analytical ease we will use 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 $\sum_{i=1}^{n} w_i = 1$. This approach is apparently appropriate, as typically the number N of items presented may be quite large. The information matrix is then more generally defined as

$$\mathbf{M}(\xi;\beta) = \textstyle\sum_{i=1}^{n} w_i\lambda(\mathbf{x}_i;\beta)\mathbf{f}(\mathbf{x}_i)\mathbf{f}(\mathbf{x}_i)^\top. \tag{2}$$

The information matrix and, hence, optimal designs will depend on the parameter vector β through the intensity. For measuring the quality of a design we will use the popular D-criterion: A design ξ will be called locally D-optimal at β if it maximizes the determinant of the information matrix $\mathbf{M}(\xi;\beta)$.

We can factorize the information matrix $\mathbf{M}(\xi;\beta) = A\mathbf{M}_0(\xi;\beta)$, where $\mathbf{M}_0(\xi;\beta)$ is the information in the standardized situation $A = 1$, $\beta_0 = 0$ with B replaced by $B\exp(-\beta_0)$. Hence, only $\det(\mathbf{M}_0(\xi;\beta))$ has to be optimized, and we assume the standardized case ($A = 1$ and $\beta_0 = 0$) without loss of generality throughout the remainder of the paper.

4 Two binary predictors

First we note that for the situation of only one rule ($K = 1$) it is obvious that the D-optimal design assigns equal weights $w_i^* = 1/2$ to the two possible settings $x = 1$ of application of the rule and $x = 0$ of the basic item independently of β.

For $K = 2$ rules the four possible settings are $(1,1)$, where both rules are applied, $(1,0)$ and $(0,1)$, where either only the first or the second rule is used, respectively, and $(0,0)$ for the basic item. Hence, any design ξ is completely determined by the corresponding weights w_{11}, w_{10}, w_{01} and w_{00}, respectively. Further we denote by $\lambda_{x_1x_2} = \lambda((x_1,x_2);\beta)$ the related intensities for given parameter $B > 0$. It is worthwhile noting that the information matrix of any design depends on the parameters only through the intensities. Hence all optimality results will be obtained in terms of the intensities and can be transferred to any such model with binary predictors.

Denote by ξ_0 the saturated design with equal weights $1/3$ on the three settings $(0,0)$, $(1,0)$ and $(0,1)$, in which at most one rule is applied. For non-positive values $\beta_k \leq 0$ the optimality condition of Graßhoff et al. (2013) can be reformulated in terms of the inverse intensities.

Theorem 1. *The design ξ_0 is locally D-optimal if and only if*

$$\lambda_{00}^{-1} + \lambda_{10}^{-1} + \lambda_{01}^{-1} \leq \lambda_{11}^{-1}\,. \tag{3}$$

For the present Poisson-Gamma model condition (3) can be rephrased as

$$\beta_2 \leq \log\left(\frac{B(1-\exp(\beta_1))}{2\exp(\beta_1)+B(1+\exp(\beta_1))}\right) \tag{4}$$

in terms of the parameters. If the condition is not satisfied then also items have to be used in which both rules are applied. The parameter regions of β_1 and β_2, where the saturated design ξ_0 is locally D-optimal, are displayed in Figure 1a for particular values of B. If the point (β_1,β_2) lies below the corresponding line, then ξ_0 is optimal. It is easy to see that the boundary condition (4) tends to the Poisson case, $\beta_2 \leq \log((1-\exp(\beta_1))/(1+\exp(\beta_1)))$, when B tends to infinity.

For illustrative purposes we will consider two particular parameter constellations, where optimal weights can be determined explicitly even when condition (3) is not satisfied. First we consider the situation of equal effect sizes, $\beta_1 = \beta_2 = \beta \leq 0$ indicated by the dotted line in Figure 1a. Due to symmetry considerations with respect to interchanging the predictors we can conclude that the optimal weights satisfy $w_{10}^* = w_{01}^*$. For given parameter $B > 0$ the condition (3) yields that ξ_0 is optimal if $\beta \leq \beta_c = \log((\sqrt{2B(B+1)} - B)/(B+2))$. For $0 \geq \beta > \beta_c$ all four settings have to be presented in an optimal design. In that case the optimal weights can be calculated to

$$w_{10}^* = w_{01}^* = \frac{4\gamma + 2\sqrt{\gamma^2 + 12\rho_0\rho_1}}{3(4\rho_0\rho_1 - \gamma^2)}, \tag{5}$$

where $\rho_0 = \lambda_{10}/\lambda_{00}$ and $\rho_1 = \lambda_{10}/\lambda_{11}$ are intensity ratios, $\gamma = \rho_0 + \rho_1 - 4$,

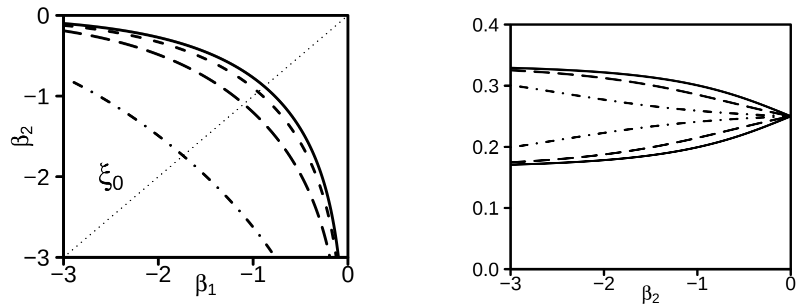

Fig. 1 a (left panel): Regions for locally D-optimal designs in dependence on (β_1, β_2) for $B = 0.1$ (dashes and dots), $B = 1$ (long dashes), $B = 4$ (short dashes), and $B = \infty$ (Poisson; solid line); b (right panel): Optimal weights $w^*_{11} = w^*_{01} \leq 0.25$ and $w^*_{10} = w^*_{00} \geq 0.25$ for $B = 0.1$ (dashes and dots), $B = 1$ (dashed line) and $B = \infty$ (Poisson; solid line) in the case $\beta_1 = 0$

$$w^*_{00} = \frac{1}{2} - w^*_{10} + \frac{1}{4}(\rho_1 - \rho_0)w^*_{10}$$
$$\text{and} \quad w^*_{11} = \frac{1}{2} - w^*_{10} - \frac{1}{4}(\rho_1 - \rho_0)w^*_{10}. \tag{6}$$

For the Poisson-Gamma model the optimal weights are displayed in Figure 2 in dependence on β for various values of B. It can be seen that the equally weighted four point design $\bar{\xi}$ (with $\bar{w}_{x_1,x_2} = 1/4$) can be recovered to be optimal for the case of zero effects ($\beta = 0$).

Alternatively, when one of the effects vanishes, say $\beta_1 = 0$, which corresponds to the vertical axis in Figure 1a, the intensities are constant in the first component. Then because of symmetry also the optimal weights are constant in x_1 and can be calculated to

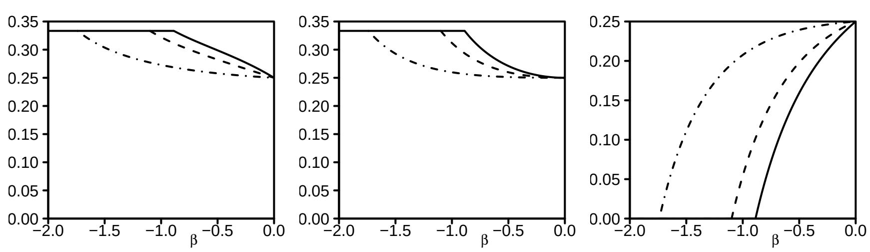

Fig. 2 Optimal weights w^*_{00} (left panel), $w^*_{10} = w^*_{01}$ (middle panel) and w^*_{11} (right panel) for $B = 0.1$ (dashes and dots), $B = 1$ (dashed line) and $B = \infty$ (Poisson; solid line) in the case $\beta_1 = \beta_2 = \beta$

$$w_{11}^* = w_{01}^* = \frac{2\lambda_{00} - \lambda_{11} - \sqrt{\lambda_{00}^2 - \lambda_{00}\lambda_{11} + \lambda_{11}^2}}{6(\lambda_{00} - \lambda_{11})} \tag{7}$$

and $w_{10}^* = w_{00}^* = 1/2 - w_{11}^*$. For the Poisson-Gamma model Figure 1b exhibits these weights as functions of β_2 for various values of B. The weights $w_{11}^* = w_{01}^*$ decrease, when these items become more difficult, and approach $1/6$ for $\beta_2 \to -\infty$. Hence, more observations will be allocated to the items with lower difficulty.

5 Efficiency

Locally D-optimal designs may show poor performance, if erroneous initial values are specified for the parameters. Similar to Graßhoff et al. (2013) we will perform a sensitivity analysis for the Poisson-Gamma model which is additionally influenced by the value of B.

We consider the parameter constellations of the preceding section for which we can determine the optimal weights explicitly. For the case of equal effect sizes ($\beta_1 = \beta_2 = \beta$) we display the D-efficiency of the saturated design ξ_0 in the left panel and the efficiency of the equally weighted four point design $\bar{\xi}$ that is optimal for $\beta_1 = \beta_2 = 0$ in the right panel of Figure 3 for various values of B. If $\beta \leq \log((\sqrt{2B(B+1)} - B)/(B+2)$, the design ξ_0 is locally D-optimal and has, hence, efficiency 100%. When β increases beyond this critical value, the efficiency of ξ_0 decreases to 83% for $\beta = 0$ for all B. Moreover, for ξ_0 the efficiency decreases when B gets smaller. For $\bar{\xi}$ the efficiency obviously equals 100% for $\beta = 0$ and decreases to 75% when β tends to $-\infty$. Here the efficiency increases when B gets smaller which seems to be reasonable because for small values of B the intensities are nearly constant as in the corresponding linear model, where $\bar{\xi}$ is D-optimal.

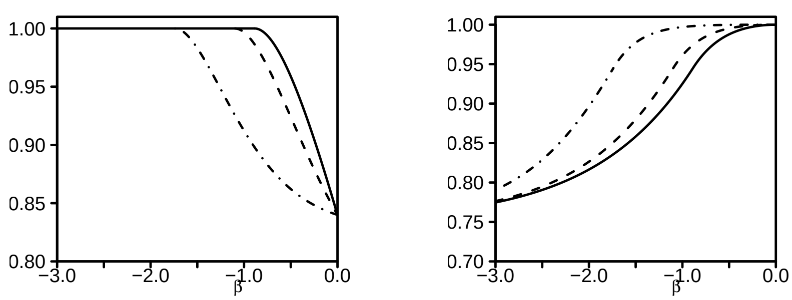

Fig. 3 Efficiency of ξ_0 (left panel) and $\bar{\xi}$ (right panel) for $B = 0.1$ (dashes and dots), $B = 1$ (dashed line) and $B = \infty$ (Poisson; solid line) in the case $\beta_1 = \beta_2 = \beta$

Similarly, for the case $\beta_1 = 0$ we display the D-efficiency of the saturated design ξ_0 in the left panel and the efficiency of the equally weighted four point design $\bar{\xi}$ in the right panel of Figure 4 for various values of B. For ξ_0 the efficiency attains again the minimal value of 83% for the situation of no effects ($\beta_2 = 0$) and approaches 100% when $\beta_2 \to -\infty$ for all B. For $\bar{\xi}$ the efficiency goes from 100% for $\beta_2 = 0$ to 94% when $\beta \to -\infty$ for all B. It can again be noticed that for fixed β_2 the efficiency of ξ_0 decreases and of $\bar{\xi}$ increases when B gets smaller.

From this figures we may conclude that in the case of equal effects $\beta_1 = \beta_2 = \beta$ the minimal efficiency is 83% for ξ_0 and 75% for $\bar{\xi}$, where the minimum is attained at $\beta = 0$ for ξ_0 and at $\beta \to -\infty$ for $\bar{\xi}$, respectively.

The maximin efficient design ξ_m^* can be obtained by easy but tedious computations. It turns out that ξ_m^* is the same for all B and allocates equal weights $w_{00}^* = w_{10}^* = w_{01}^* = 0.31$ to the settings of the saturated design ξ_0 and the remaining weight $w_{11}^* = 0.07$ to the setting in which both rules are applied. More precisely, the maximin efficient design $\xi_m^* = (111\,\xi_0 + 44\,\bar{\xi})/155$ is a mixture of the two limiting locally optimal designs ξ_0 and $\bar{\xi}$, where the coefficients do not depend on B. Moreover, ξ_m^* is also maximin efficient in the Poisson model. As can be seen from Figure 5 (left panel) the minimal efficiency of ξ_m^* equals 93% for all B and is attained both for $\beta = 0$ and $\beta \to -\infty$. Additionally it appears that ξ_m^* is locally optimal for some β which varies with B. Moreover, as shown in the right panel of Figure 5 the efficiency of ξ_m^* is also at least 93% when one of the effects vanishes. Hence, we conjecture that ξ_m^* is maximin efficient for all $\beta_1, \beta_2 \leq 0$.

6 Discussion

As pointed out the performance of a design is completely determined in a model with binary predictors by the intensities for the single settings. Thus optimal de-

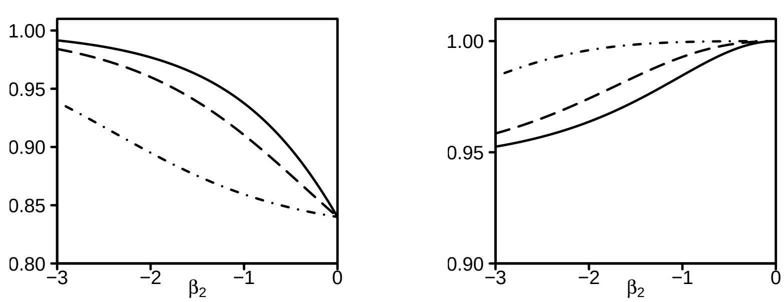

Fig. 4 Efficiencies of ξ_0 (left panel) and $\bar{\xi}$ (right panel) for $B = 0.1$ (dashes and dots), $B = 1$ (dashed line) and $B = \infty$ (Poisson; solid line) in the case $\beta_1 = 0$

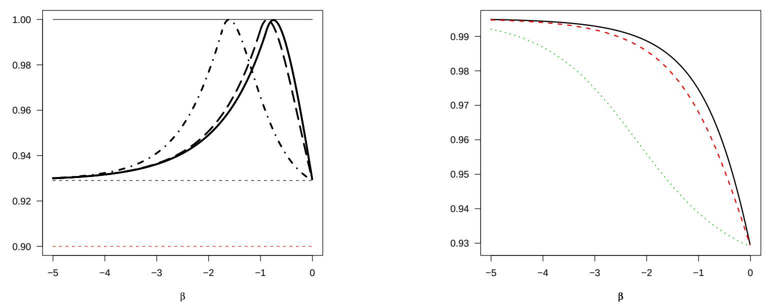

Fig. 5 Efficiency of ξ_m^* for $B = 0.1$ (dashes and dots), $B = 2$ (dashed line) and $B = \infty$ (Poisson; solid line) in the case $\beta_1 = \beta_2 = \beta$ (left panel) and $\beta_1 = 0$ (right panel)

signs can be obtained, in general, in terms of these intensities computed from the parameters using the model equations or, vice versa, conditions on the intensities can be transferred back to these parameters. Extensions to higher dimensions are under consideration and require, in general, higher order conditions.

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

References

1. Graßhoff, U., Holling, H., Schwabe, R.: Optimal design for count data with binary predictors in item response theory. In: Uciński, D., Atkinson, A.C., Patan, M. (eds) mODa10 -Advances in Model-Oriented Design and Analysis, pp. 117-124. Springer, Cham (2013)
2. Rasch, G.: Probabilistic Models for Some Intelligence and Attainment Tests. Danish Institute for Educational Research, Copenhagen (1960)
3. Verhelst, N.D., Kamphuis, F.H.: A Poisson-Gamma model for speed tests. Measurement and Research Department Reports **2009-2**. Cito, Arnhem (2009)