Some Curiosities in Optimal Designs for Random Slopes

Thomas Schmelter[1], Norbert Benda[2], and Rainer Schwabe[3]

[1] Otto-von-Guericke-University, Institute for Mathematical Stochastics, PF 4120, D–39 016 Magdeburg, Germany / Schering AG, Clinical Statistics Europe, D–13 342 Berlin, Germany Thomas.Schmelter@schering.de
[2] Schering AG, Clinical Statistics Europe, D–13 342 Berlin, Germany Norbert.Benda@schering.de
[3] Otto-von-Guericke-University, Institute for Mathematical Stochastics, PF 4120, D–39 016 Magdeburg, Germany rainer.schwabe@mathematik.uni-magdeburg.de

Summary. The purpose of this note is to show by a simple example that some of the favourite results in optimal design theory do not necessary carry over if random effects are involved. In particular, the usage of the popular D-criterion appears to be doubtful.

Key Words: Optimal design, mixed linear model, random coefficient regression

MSC classification: 62K05, 62J05

1 Introduction

Mixed models have attracted growing interest in the biosciences, when replicated measurements are available from different individuals. While the corresponding statistical analysis is well-developed, only few results are available on optimal designs for such experiments. For a recent survey on the particular setting of random coefficient regression see Entholzner et al. (2005).

The most popular criterion in applications is the D-criterion in analogy to fixed effects only models, where it has some nice properties but which do not necessarily carry over to mixed models. Fedorov and Hackel (1997, p. 75) provide an equivalence theorem for this situation which has been extended by Schmelter (2006). While results are quite obvious for random intercept models (Schwabe and Schmelter, 2006), the optimization may lead to apparently counter-intuitive solutions, if there is randomness in the treatment effects (see e.g. Fedorov and Leonov, 2004).

In the present note we will indicate how various standard criteria are influenced by the presence of random individual effects.

2 The Model

To keep notations as simple as possible we discuss a straight line regression model on the unit interval in which only the slopes are affected by random effects. More specifically, we consider n individuals with m observations each, and the jth observation Y_{ij} of individual i is described by

$$Y_{ij} = \mu + b_i\, x_{ij} + \varepsilon_{ij} \ ,$$

where x_{ij} is the corresponding experimental setting, $0 \leq x_{ij} \leq 1$, $i = 1, \ldots, n, j = 1, \ldots, m$. The individual random slopes b_i are assumed to be *iid* with unknown population mean β and known variance σ_β^2. A typical example for a bunch of the conditional individual mean response lines $\mu + b_i x$ is given in the spaghetti plot of Figure 1. Our interest will be only in the population parameters μ and β or, equivalently, the mean response $\mu + \beta x$ across the individuals, rather than in prediction.

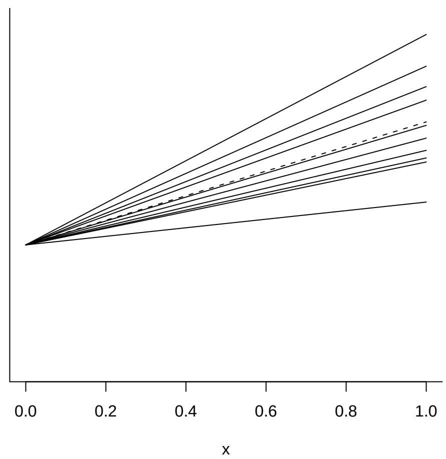

Figure 1: Population (dashed line) and individual mean response curves (solid lines)

Furthermore the observational errors ε_{ij} are assumed to be homoscedastic (*iid*) with zero mean and known variance σ^2 and to be independent of the random slope parameters b_i. We define the dispersion factor $d = \sigma_\beta^2/\sigma^2$ as the variance ratio of the slope compared to the observational error.

Denote by $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})^\top$ and $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})$ the vector of observations and corresponding experimental settings, respectively, and by $\mathbf{F}_i = (\mathbf{1}_m \, \mathbf{x}_i)$ the associated design matrix for individual i. Here $\mathbf{1}_m$ is a vector of length m with all entries equal to one. The covariance matrix $\mathbf{V}_i$ of the observation $\mathbf{Y}_i$ is given by $\mathbf{V}_i = \sigma^2(\mathbf{I}_m + \mathbf{F}_i \left(\begin{smallmatrix} 0 & 0 \\ 0 & d \end{smallmatrix}\right) \mathbf{F}_i^\top) = \sigma^2(\mathbf{I}_m + d\, \mathbf{x}_i \, \mathbf{x}_i^\top)$, where $\mathbf{I}_m$ is the m-dimensional identity matrix. For simplicity we will assume furtheron without loss of generality that σ^2 is equal to 1.

The weighted least squares estimator $(\sum_{i=1}^n \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i)^{-1} \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{Y}_i$ is the best linear unbiased estimator for $(\mu, \beta)^\top$ and its covariance matrix is equal to the inverse of the information matrix $\mathbf{M}_d = \sum_{i=1}^n \mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i$ which depends on the variance ratio d through $\mathbf{V}_i^{-1} = \mathbf{I}_m - \frac{d}{1+d\mathbf{x}_i^\top \mathbf{x}_i} \mathbf{x}_i \, \mathbf{x}_i^\top$ as well as on the experimental settings $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Note that the individual information $\mathbf{F}_i^\top \mathbf{V}_i^{-1} \mathbf{F}_i$ is equal to $\frac{m}{1+md\,\nu_{i2}} \begin{pmatrix} 1 + m\, d(\nu_{i2} - \nu_{i1}^2) & \nu_{i1} \\ \nu_{i1} & \nu_{i2} \end{pmatrix}$, where $\nu_{ik} = \frac{1}{m} \sum_{j=1}^m x_{ij}^k$ denotes the kth moment of the experimental setting $\mathbf{x}_i$.

For estimating the mean response $\mu + \beta x$ over the design region $(0 \le x \le 1)$ the variance function is given by $v_d(x) = \operatorname{var}(\hat{\mu} + \hat{\beta} x) = (1, x)\, \mathbf{M}_d^{-1} \binom{1}{x}$.

3 Optimal Design

Design optimality aims at finding the best experimental settings $\mathbf{x}_i$ to maximise the information $\mathbf{M}_d = \mathbf{M}_d(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ or, equivalently, to minimisze the covariance matrix $\mathbf{M}_d^{-1}$. As uniform matrix optimization is not possible, there are various competing optimality criteria which are real-valued functionals of the information $\mathbf{M}_d$. One of the most popular is the D-criterion which aims at maximizing the determinant of $\mathbf{M}_d$. For fixed effects only models (i.e. $d = 0$) the D-optimality is equivalent to optimization with respect to the G-criterion, which aims at minimizing the maximum $\max_{0 \le x \le 1} v_d(x)$ of the variance function over the design region according to the Kiefer-Wolfowitz (1960) equivalence theorem within the setup of approximate designs. To avoid discretizations we will deal with such a generalised setup throughout this section. According to Schmelter (2006) optimal designs can be found among those which are uniform across the individuals, i.e. $\mathbf{x}_i = \mathbf{x}$ and, hence, $\mathbf{F}_i = \mathbf{F}$ for all i. Then the covariance simplifies to $\mathbf{M}_d^{-1} = \frac{1}{n} \left((\mathbf{F}^\top \mathbf{F})^{-1} + \left(\begin{smallmatrix} 0 & 0 \\ 0 & d \end{smallmatrix}\right) \right)$ (see Entholzner et al., 2005). Due to majorization (see e.g. Pukelsheim, 1993, p. 101) we can confine the search for optimal designs to those with observations at the extreme settings $x = 0$ and $x = 1$. Candidates for an optimal design are, thus, characterised by the number m_1 or, equivalently, by the proportion $w = m_1/m$ of observations at the experimental setting $x = 1$, while $m_0 = (1 - w)m$ observations are made at the baseline, $x = 0$, for each individual. The corresponding covariance matrices can be calculated as

$$\mathbf{M}_d^{-1} = \frac{1}{nm} \frac{1}{w(1-w)} \begin{pmatrix} w & -w \\ -w & 1 + mdw(1-w) \end{pmatrix} .$$

For the optimization we also allow generalised proportions w which are not necessarily multiples of $1/m$.

Theorem 1. *The D-optimal proportion w_D^* at $x = 1$ equals* $(1+\sqrt{md+1})^{-1}$.

Proof. The determinant of $\mathbf{M}_d^{-1}$ is proportional to $(1 + mdw)/(w(1-w))$ which is minimised by w_D^*.

□

The optimal proportion w_D^* varies continuously with the variance ratio d, and w_D^* tends to zero as d tends to infinity (see Figure 2).

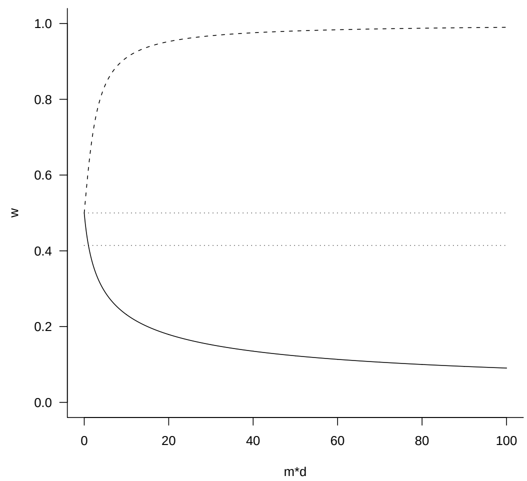

Figure 2: Optimal proportions of observations in $x = 1$: D-optimal (solid line), G-optimal (dashed line), β- and $IMSE$-optimal (upper horizontal dotted line), and A-optimal (lower horizontal dotted line)

It has to be noted that all optimal designs depend on the number m of replications and the variance ratio d only through their product $m \cdot d$ as the standardised information matrix $nm\mathbf{M}_d$ does.

Theorem 2. *The G-optimal proportion w_G^* at $x = 1$ equals* $\frac{1}{2}(1 - 2(md)^{-1} + \sqrt{1 + 4(md)^{-2}})$ *if $d > 0$, and $w_G^* = 1/2$ for $d = 0$.*

Proof. As the variance function $v_d(x) = \frac{1}{nm} \frac{1}{w(1-w)} (w - 2wx + (1 + m\,dw(1 - w))x^2)$ is a polynomial of degree 2 with positive leading term ($0 < w < 1$), its maximum is attained either at $x = 0$ or $x = 1$, or both, i.e. $\max_{0 \le x \le 1} v_d(x) =$

$\max(v_d(0), v_d(1))$. Now, the standardised variance $nmv_d(0) = (1-w)^{-1}$ is strictly increasing in w while $nmv_d(1) = w^{-1} + md$ is strictly decreasing in w. Thus, $\min_{0<w<1} \max(v_d(0),\, v_d(1))$ is attained when $v_d(0) = v_d(1)$, i.e. $(1-w)^{-1} = w^{-1} + md$, which is solved by w_G^*.

□

The optimal proportion w_G^* varies continuously in d, but, in contrast to the D-optimal proportion, it tends to one as d tends to infinity (see Figure 2). Thus D- and G-optimal proportions are very sensitive to the variance ratio d and differ essentially if d is large.

Linear criteria, however, which are of the form $\mathrm{tr}(\mathbf{A}\,\mathbf{M}_d^{-1})$ for some fixed positive semidefinite matrix $\mathbf{A}$ are not affected by the variance ratio d, because $n^{-1}\mathrm{tr}(\mathbf{A}\,\mathbf{M}_d^{-1}) = \mathrm{tr}(\mathbf{A}(\mathbf{F}^\top\,\mathbf{F})^{-1}) + da_{22}$ decomposes into the corresponding criterion $\mathrm{tr}(\mathbf{A}(\mathbf{F}^\top\mathbf{F})^{-1})$ of the fixed effects only model and a design independent constant da_{22}, where a_{22} is the lower right entry in $\mathbf{A}$. Hence, for such criteria the optimal design is independent of d. Typical examples are the c-criterion for the slope β, $c^\top\,\mathbf{M}_d^{-1}\,c$, where $c = (0,1)^\top$, the integrated mean squared error (IMSE) criterion, $\int_0^1 v_d(x)\,dx$ or the A-criterion, $\mathrm{tr}(\mathbf{M}_d^{-1})$.

Theorem 3. *The β- and IMSE-optimal proportions $w_\beta^* = w_{IMSE}^*$ are equal to $1/2$. The A-optimal proportion w_A^* is equal to $\sqrt{2}-1$.*

To judge the impact of design optimization one is tempted to calculate the efficiency of the proportion $w_0 = 1/2$ which is simultaneously D- and G-optimal for the fixed effects only model, i.e. $d = 0$, when the variance ratio increases. The D-efficiency, which is $(\det \mathbf{M}_d(w_0)/\det \mathbf{M}_d\,(w_D^*))^{1/2} = (1 + \sqrt{md+1})/\sqrt{4+2md}$, decreases slowly to $1/\sqrt{2} > 0.70$ if the variance ratio d becomes large. But the G-efficiency which is equal to $(w_G^{*-1} + md)/(2+md)$ shows a strange behaviour. If d increases the G-efficiency drops very quickly to about 0.86 and, then, increases again and tends ultimately to one for d going to infinity. This strange limiting behaviour may be explained by the fact, that the G-efficiency for w_0 is bounded by $(1+md)/(2+md)$ from below. The efficiencies are plotted in Figure 3.

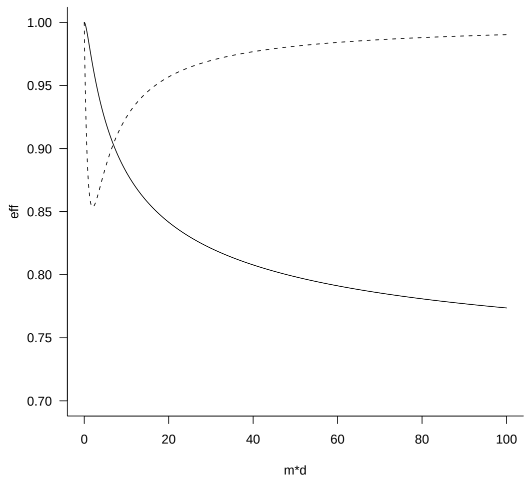

Figure 3: Efficiency of the equireplicated design ($w = 0.5$) : D-criterion (solid line) and G-criterion (dashed line)

4 Discretization

For applications it must be required that the proportion w is a multiple of $1/m$, i.e. that the number $m_1 = w \cdot m$ of observations at $x = 1$ is an integer. In general, optimal numbers m_1^* can be found by rounding w^*m to the next smaller or larger integer $[w^*m]$ or $[w^*m] + 1$, respectively, where w^* is the optimal generalised proportion obtained in the previous section and $[\cdot]$ denotes the integer part.

For example, in the case $m = 2$ the optimal value m_1^* for the number of observations at $x = 1$ is equal to 1 independently of d for every reasonable criterion considered in section 3 due to estimability requirements. In the case $m = 4$ the D-optimal number m_1^* is equal to 2 for small variance ratios $d \leq 1/2$ and has to be chosen as 1 for larger values, $d \geq 1/2$. Note that in this case the generalised solution $w_D^*\, m$ yields the integer value 1 for $d = 2$.

In the situation of non-integer w^*m it may additionally turn out that it is more favourable to apply non-uniform designs in which experimental settings may differ from individual to individual. It seems reasonable that for a certain proportion α of individuals m_1 is chosen to be equal to $[w^*m] + 1$ while for the remaining $(1 - \alpha) \cdot n$ individuals m_1 equals $[w^*m]$ in order to improve the performance of the design. For example, in the case $m = 4$ it is D-optimal to have a proportion of $1 - d$ individuals with $m_1 = 2$ observations at 1 and a proportion of d individuals with $m_1 = 1$ as long as the variance ratio d is smaller than 1. For larger d, $d \geq 1$, the uniform design becomes D-optimal with $m_1 = 1$ for all individuals.

For $m = 2$ it can be shown by the multivariate version of the equivalence theorem (see Fedorov, 1972, p. 212) that the uniform design with $m_1 = 1$ is simultaneously D-optimal for all values of the variance ratio d.

Analogous findings can be obtained for the G-criterion.

5 Discussion

In the present simple model of straight line regression with random slopes neither the commonly used D-criterion nor its pretended counterpart the G-criterion seem to show a reasonable behaviour, in particular, if the variablility is large. While the D-criterion gives solutions which are lightweight in the sense that most observations are made where it is easy, i.e. where the variance is small, the G-criterion overemphasises the difficult observations, where the variation is large and which cannot be substantially reduced by increasing the number of intraindividual replications. Moreover, the G-criterion exhibits a strange non-monotonic efficiency behaviour. In fact, it can be shown that for every regular design its G-efficiency tends to one if the variance ratio tends to infinity. This indicates that with respect to the G-criterion all designs are equally good - or equally bad - if d is large.

Although this last statement also applies to linear criteria like the IMSE-criterion they seem to be a reasonable compromise and, in particular, have the advantage to result in optimal designs which are independent of the magnitude of the variance ratio.

References

1. Entholzner, M. — Benda, N. — Schmelter, T. — Schwabe, R. (2005): A note on designs for estimating population parameters. *Listy Biometryczne - Biometrical Letters* **42**, 25–41.
2. Fedorov, V.V. (1972): *Theory of Optimal Experiments.* Academic Press, New York.
3. Fedorov, V. V. — Hackl, P. (1997): *Model-Oriented Design of Experiments. Lecture Notes in Statistics* **125**. Springer, New York.
4. Fedorov, V. V. — Leonov, S. (2004): Optimal designs for regression models with forced measurements at baseline. In *mODa 7 - Advances in Model-Oriented Design and Analysis* (A. Di Bucchianico, H. Läuter and H. P. Wynn, eds.). Physica, Heidelberg, 61–69.
5. Kiefer, J. — Wolfowitz, J. (1960): The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12**, 363–366.
6. Pukelsheim, F. (1993): *Optimal Design of Experiments.* Wiley, New York.

7. Schmelter, T. (2006): The optimality of single-group designs for certain mixed models. *Metrika (to appear)*.
8. Schwabe, R.— Schmelter, T. (2006): On optimal designs in random intercept models. (Unpublished manuscript).