Journal of Statistical Planning and Inference 142 (2012) 1108–1113

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Optimal design for linear regression models in the presence of heteroscedasticity caused by random coefficients ☆

Ulrike Graßhoff [a], Anna Doebler [b], Heinz Holling [b], Rainer Schwabe [a,*]

[a] Otto-von-Guericke University, Institute for Mathematical Stochastics, PF 4120, D-39 016 Magdeburg, Germany
[b] Westfälische Wilhelms-Universität, Psychologisches Institut IV, Fliednerstr. 21, D-48 149 Münster, Germany

ARTICLE INFO

Article history:
Received 26 July 2011
Received in revised form
22 November 2011
Accepted 25 November 2011
Available online 8 December 2011

Keywords:
Optimal design
Linear mixed model
Random coefficient regression

ABSTRACT

Random coefficients may result in heteroscedasticity of observations. For particular situations, where only one observation is available per individual, we derive optimal designs based on the geometry of the design locus.

1. Introduction

In the social sciences and biosciences random effects play a growing role, whenever different individuals are involved in a study. While in fixed effects models typically only additive errors are taken into account, the situation has to change, when the coefficients of the regression function may vary randomly across those individuals. Our approach here is motivated by a validation problem in intelligence testing, when only one observation is made available per individual. Freund and Holling (2008) analyzed the impact of reasoning and creativity on the grade point average (GPA) for students' scholastic achievements based on data from the standardization sample of the Berlin Structure of Intelligence Test for Youth: Assessment of Talent and Giftedness (BIS-HB). Since the effect of both variables on school performance may vary between different classrooms, a random effects model incorporating the explanatory variables on two levels (level 1: students and level 2: classrooms) was specified. Thus, the results allow for a more detailed interpretation of the role of different variables in the context of predicting scholastic achievement.

Such "sparse" observations have also been considered by Patan and Bogacka (2007) in a population pharmacokinetics setup. In those applications the response is usually non-linear. However, to solve the corresponding design problem it is advisable to have some knowledge of the influence of random coefficients already in the linear model setup. This is the motivation of the present investigation.

In the linear setup the random coefficient model with a single observation can be reformulated as a heteroscedastic fixed effects model with a specific structure of the variance function. If the covariance matrix of the random coefficients is regular it can be easily verified that the corresponding standardized design locus is included on the surface of an ellipsoid generated by that covariance matrix. In the case of high variability this ellipsoid may coincide with the smallest

☆ Work supported by Grant HO 1286/6-1 of the Deutsche Forschungsgemeinschaft.
* Corresponding author.
E-mail address: rainer.schwabe@ovgu.de (R. Schwabe).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2011.11.025

circumscribing one. Then multiple solutions to the design problem are possible as indicated in the discussion by Silvey (1972) and Sibson (1972). This phenomenon will be demonstrated by some simple but illustrative examples.

2. Model description

We consider a random coefficient regression model $Y_i(\mathbf{x}_i)=\mathbf{f}(\mathbf{x}_i)^\top\mathbf{b}_i$. The dependence of the observations Y_i on the experimental settings $\mathbf{x}_i$ is given by the p-dimensional vector of known regression functions $\mathbf{f}$ and the statistically independent vectors $\mathbf{b}_i$ of random coefficients, which come from a normal distribution, $\mathbf{b}_i \sim \mathrm{N}_p(\boldsymbol{\beta},\mathbf{D})$, with mean vector $\boldsymbol{\beta}$ and dispersion matrix $\mathbf{D}$. The design problem is to choose the experimental settings $\mathbf{x}_i$ from the design region $\mathcal{X}$ for estimating the population location parameters $\boldsymbol{\beta}$, while the dispersion matrix $\mathbf{D}$ is assumed to be known.

In this note we suppose that all observations Y_i are independent, i.e. only one observation is made for each realization $\mathbf{b}_i$ of the random coefficients. Moreover, we assume here that an intercept is included in the model $(f_1(\mathbf{x})\equiv 1)$ such that additive observational errors can be subsumed into the random intercept, which will be done in what follows.

This model can be rewritten as a heteroscedastic linear fixed effects model

$$Y_i(\mathbf{x}_i)=\mathbf{f}(\mathbf{x}_i)^\top\boldsymbol{\beta}+\varepsilon_i, \tag{1}$$

where the observational errors $\varepsilon_i=\mathbf{f}(\mathbf{x}_i)^\top(\mathbf{b}_i-\boldsymbol{\beta})\sim \mathrm{N}(0,\sigma^2(\mathbf{x}_i))$ are independent and their variance function is defined by $\sigma^2(\mathbf{x})=\mathbf{f}(\mathbf{x})^\top\mathbf{D}\mathbf{f}(\mathbf{x})$. Within this heteroscedastic linear model the information equals $\mathbf{M}(\mathbf{x})=\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top/\sigma^2(\mathbf{x})$ for each single setting $\mathbf{x}\in\mathcal{X}$. Then for a design ξ the standardized information matrix is defined by $\mathbf{M}(\xi)=\sum_{j=1}^m \xi(\mathbf{x}_j)\mathbf{M}(\mathbf{x}_j)$, where $\xi(\mathbf{x}_j)$ is the proportion of observations at setting $\mathbf{x}_j$, $\sum_{j=1}^m \xi(\mathbf{x}_j)=1$.

Note that the covariance matrix for the weighted least squares estimator $\hat{\boldsymbol{\beta}}$, which is the best unbiased estimator for $\boldsymbol{\beta}$ and coincides with the maximum likelihood estimator in the present setting, equals the inverse of the information matrix.

To compare different designs we consider the most popular criterion, the D-criterion, with respect to which a design ξ^* is D-optimal, if it maximizes the determinant of the information matrix. This is equivalent to the minimization of the volume of a confidence ellipsoid for $\boldsymbol{\beta}$. In the setting of approximate designs, for which the proportions $\xi(\mathbf{x})$ are not necessarily multiples of $1/n$, where n denotes the sample size, the D-optimality of a design ξ^* can be established by the well-known Kiefer–Wolfowitz equivalence theorem (see Fedorov, 1972, for a suitable version): A design ξ^* is D-optimal, if $\mathbf{f}(\mathbf{x})^\top\mathbf{M}(\xi^*)^{-1}\mathbf{f}(\mathbf{x})/\sigma^2(\mathbf{x})\le p$, uniformly in $\mathbf{x}\in\mathcal{X}$. When we substitute $\sigma^2(\mathbf{x})=\mathbf{f}(\mathbf{x})^\top\mathbf{D}\mathbf{f}(\mathbf{x})$ into this relation and rearrange terms, D-optimality is achieved, if

$$\delta(\mathbf{x};\xi^*)\ge 0 \tag{2}$$

for all $\mathbf{x}\in\mathcal{X}$, where $\delta(\mathbf{x};\xi)=\mathbf{f}(\mathbf{x})^\top(p\mathbf{D}-\mathbf{M}(\xi)^{-1})\mathbf{f}(\mathbf{x})$ is the suitably transformed sensitivity function. Moreover, equality is attained in (2) for design points, where $\xi^*(\mathbf{x})>0$.

3. Linear regression on the standard interval

In the situation of linear regression we have observations $Y_i=b_{i0}+b_{i1}x$. The vector of regression functions $\mathbf{f}$ is given by $\mathbf{f}(x)=(1,x)^\top$, and the random coefficient vector is $\mathbf{b}_i=(b_{i0},b_{i1})^\top$. We assume that the setting x may be chosen from the symmetric standard interval $\mathcal{X}=[-1,1]$, and we consider the design problem with an underlying general covariance matrix

$$\mathbf{D}=\begin{pmatrix} d_0 & d_{01}\\ d_{01} & d_1\end{pmatrix}.$$

Here the variances of the random coefficients are denoted by $d_0=\mathrm{Var}(b_{i0})$ and $d_1=\mathrm{Var}(b_{i1})$, respectively, while $d_{01}=\mathrm{Cov}(b_{i0},b_{i1})$ is the covariance of the two random coefficients. The variance function equals $\sigma^2(x)=d_0+2d_{01}x+d_1x^2$ for each setting x.

For any $x_1,x_2\in[-1,1]$ we introduce the two-point designs ξ_{x_1,x_2} on x_1 and x_2 with equal weights $\xi_{x_1,x_2}(x_1)=\xi_{x_1,x_2}(x_2)=1/2$ as candidates for the optimal designs. Their corresponding information matrix is given by

$$\mathbf{M}(\xi_{x_1,x_2})=\frac{1}{2\sigma^2(x_1)\sigma^2(x_2)}\begin{pmatrix}\sigma^2(x_1)+\sigma^2(x_2) & x_1\sigma^2(x_2)+x_2\sigma^2(x_1)\\ x_1\sigma^2(x_2)+x_2\sigma^2(x_1) & x_1^2\sigma^2(x_2)+x_2^2\sigma^2(x_1)\end{pmatrix} \tag{3}$$

with determinant $\det(\mathbf{M}(\xi_{x_1,x_2}))=(x_1-x_2)^2/(4\sigma^2(x_1)\sigma^2(x_2))$. Maximizing the determinant of $\mathbf{M}(\xi_{x_1,x_2})$ with respect to x_1 and $x_2\in[-1,1]$ leads to the following solutions.

If $d_0\ge d_1$, the endpoints $x_1^*=1$ and $x_2^*=-1$ are optimal. In this case the information matrix results in

$$\mathbf{M}(\xi_{1,-1})=\frac{1}{(d_0+d_1)^2-4d_{01}^2}\begin{pmatrix} d_0+d_1 & -2d_{01}\\ -2d_{01} & d_0+d_1\end{pmatrix}. \tag{4}$$

Since $2\mathbf{D}-\mathbf{M}(\xi_{1,-1})^{-1}=\mathrm{diag}(d_0-d_1,d_1-d_0)$ we obtain for the sensitivity $\delta(x;\xi_{1,-1})=(d_0-d_1)(1-x^2)$, and the inequality (2) is satisfied, which proves the D-optimality of the design $\xi_{1,-1}$.