Invariance and Equivariance in Experimental Design for Nonlinear Models

Martin Radloff and Rainer Schwabe

Abstract In this note we exhibit the usefulness of invariance considerations in experimental design in the context of nonlinear models.

1 Introduction

Invariance and equivariance are powerful tools in the construction of optimal designs of experiments. In this context equivariance means that simultaneous transformations on the design region and on the design considered leave the performance of the design unchanged. In particular, this induces that an optimal design on the original design region will be transformed to an optimal design on the transformed design region. If a design problem is additionally invariant with respect to a whole group of transformations (*symmetries*), then the search for an optimal design can be restricted to the invariant (*symmetrized*) ones. For standard linear models these properties are well-known and widely investigated (see e.g. Pukelsheim, 1993, chapter 13, Gaffke and Heiligers, 1996, or Schwabe, 1996, chapter 3). In nonlinear setups, however, additional features occur which need some clarification. In particular, because of the parameter dependence of the information matrix the introduction of an additional transformation on the parameter space is required. Starting with this parameter transformation Ford et al. (1992) introduced in their seminal paper equivariance under the name *canonical transformation* in the context of generalized linear models. Using this concept Ford et al. (1992) could characterize locally optimal designs for arbitrary pre-specified parameter values on the basis of standardized

Martin Radloff
Institute for Mathematical Stochastics, Otto-von-Guericke-University, PF 4120, 39016 Magdeburg, Germany, e-mail: martin.radloff@ovgu.de

Rainer Schwabe
Institute for Mathematical Stochastics, Otto-von-Guericke-University, PF 4120, 39016 Magdeburg, Germany, e-mail: rainer.schwabe@ovgu.de

models. This approach can be readily extended to other nonlinear situation as long as the intensity function of the information depends only on the value of the linear predictor. Here we put their approach bottom-up and start with transformations of the design region which makes argumentations more straightforward. First we introduce the model specifications and exhibit the local optimality of transformed designs in Section 2. In Section 4 we establish the optimality of invariant designs with respect to robust criteria like weighted or maximin optimality, which avoid parameter dependence, under the natural additional assumption that also the weight function or the parameter region of interest, respectively, is invariant.

2 Model, Transformation and Local Optimality

In the following sections it is required that the one-support-point information matrix $M(x,\beta)$ can be written in the form

$$M(x,\beta) = \lambda\left(f(x)^\top\beta\right) f(x)f(x)^\top$$

with an intensity function λ which only depends on the value of the linear predictor. The design point x is in the design region $\mathscr{X} \subseteq \mathbb{R}^k$ and the parameter β is in $\mathscr{B} \subseteq \mathbb{R}^p$. $f : \mathscr{X} \to \mathbb{R}^p$ is the regression function.

Hence the information matrix of the (generalized) design ξ with independent observations is

$$M(\xi,\beta) = \int_{\mathscr{X}} M(x,\beta)\xi(\mathrm{d}x) = \int_{\mathscr{X}} \lambda\left(f(x)^\top\beta\right) f(x)f(x)^\top\xi(\mathrm{d}x).$$

In generalized linear models (see McCulloch and Searle, 2004) or for example in censored data models (see Schmidt and Schwabe, 2015) this prerequisite is fulfilled.

Let g be a one-to-one-mapping of the design region $\mathscr{X}$. If there is a $p \times p$ matrix Q_g with $f(g(x)) = Q_g f(x)$ for all $x \in \mathscr{X}$, g induces a linear transformation of the regression function f and f is linearly equivariant.

These one-to-one-mappings of the design region $\mathscr{X}$ provide a group G, if every $g \in G$ induces a linear transformation of the regression function f. One example is the orthogonal group $O(k)$ or subgroups of it for a k-dimensional circle.

Here we only want to focus on finite groups of transformation, like the group generated by a 90° rotation in a 2-dimensional space.

The response function should be equivariant with respect to the transformation g of the design space. So it is necessary to transform the parameter space as well and to find a analogous parameter transformation $\tilde{g}$ with

$$\mu(g(x),\tilde{g}(\beta)) = \mu(x,\beta).$$

$\tilde{g}(\beta) = (Q_g^\top)^{-1}\beta$ is such a transformation for regular Q_g.

If not only the parameter but also the design points are transformed in this way, the information matrix has the form

$$\begin{aligned}
M(\xi^g, \tilde{g}(\beta)) &= \int_{\mathscr{X}} M(x, \tilde{g}(\beta)) \xi^g(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\Big(f(x)^\top \tilde{g}(\beta)\Big) f(x) f(x)^\top \xi^g(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\Big(f(g(x))^\top \tilde{g}(\beta)\Big) f(g(x)) f(g(x))^\top \xi(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\Big((Q_g f(x))^\top (Q_g^\top)^{-1}\beta\Big) Q_g f(x) (Q_g f(x))^\top \xi(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\Big(f(x)^\top Q_g^\top (Q_g^\top)^{-1}\beta\Big) Q_g f(x) f(x)^\top Q_g^\top \xi(\mathrm{d}x) \\
&= \int_{\mathscr{X}} \lambda\Big(f(x)^\top \beta\Big) Q_g f(x) f(x)^\top Q_g^\top \xi(\mathrm{d}x) \\
&= Q_g \int_{\mathscr{X}} \lambda\Big(f(x)^\top \beta\Big) f(x) f(x)^\top \xi(\mathrm{d}x)\, Q_g^\top \\
&= Q_g M(\xi, \beta) Q_g^\top \quad .
\end{aligned}$$

In linear models it is commonly known, that optimality criteria Φ are invariant,

$$\Phi(M(\xi^g, \tilde{g}(\beta))) = \Phi(Q_g M(\xi, \beta) Q_g^\top) = \Phi(M(\xi, \beta)),$$

if the information matrix is transformed by the matrices Q_g respectively the group G satisfying special properties. So, for example, Φ_q-optimality, $0 \le q \le \infty$, is invariant if the transformation matrix Q_g is orthonormal (see Schwabe, 1996, chapter 3). In the special case of D-optimality it is enough to have unimodal matrices Q_g.

Local optimality is not invariant, in general, because the information matrix depends on the parameter β.

There are two consequences. If ξ is a locally optimal design for β with respect to Φ then ξ^g is locally optimal for $\tilde{g}(\beta)$.

And, in situations, where $\tilde{g}(\beta) = \beta$ for all $\tilde{g}$ the local criterion Φ is invariant at β:

$$\Phi(M(\xi^g, \tilde{g}(\beta))) = \Phi(M(\xi^g, \beta)) = \Phi(M(\xi, \beta)) \ \text{ for all } g \in G.$$

By defining the symmetrized design of ξ with respect to the finite group G

$$\overline{\xi} := \frac{1}{\#G} \sum_{g \in G} \xi^g \ ,$$

which is an invariant design (with respect to G, $\overline{\xi}^g = \overline{\xi}\ \forall\ g \in G$), it can be shown like in the theory of linear models (see Schwabe, 1996, chapter 3) that for concave and (with respect to G) invariant criterion function Φ it is $\Phi(\overline{\xi}) \geq \Phi(\xi)$ for every possible design ξ. Hence locally optimal designs can be found in the class of invariant designs. This is in particular the case when $\beta = 0$, which is not surprising, because then the local criterion coincides with that in the corresponding linear model without intensity function.

The first consequence should be illustrated in a small example.

3 Example

In Russell et al. (2009) a theorem to find locally D-optimal designs for Poisson regression on a cuboid is given. We only want do consider the 2-dimensional case with $f(x_i)^\top \beta = (1, x_{1i}, x_{2i})(\beta_0, \beta_1, \beta_2)^\top = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ on the squared design region $[-1,1]^2$. With an initial guess $(\tilde{\beta}_0, 1, 2)$, $\tilde{\beta}_0$ arbitrary, for the parameter vector β a (locally) D-optimal design is given by the 3 equally weighted support points $(1,1)$, $(-1,1)$ and $(1,0)$.

A 90° rotation g of the design space induces a transformation of f with $f(g(x)) = Q_g f(x)$ and

$$Q_g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Because of $(Q_g^\top)^{-1} = Q_g$, it is $\tilde{g}(\beta) = Q_g \beta = (\beta_0, -\beta_2, \beta_1)^\top$ and the transformed initial guess $\tilde{g}((\tilde{\beta}_0, 1, 2)^\top) = (\beta_0, -2, 1)^\top$. The transformed design contains the 3 equally weighted support points $(-1,1)$, $(-1,-1)$ and $(0,1)$, which is identical to the (locally) D-optimal design given by Russell et al. (2009).

4 Weighted Optimality and Maximin Designs

Locally optimal designs depend on the initial values of the parameter β. A wrong choice, like an initial guess which is too far from the real parameter, may show a bad performance. So we regard two concepts to avoid this problem: the weighted optimality or (pseudo)-Bayesian designs and the maximin approach. (see Atkinson et al., 2007)

If a prior (probability) distribution π on the parameter space $\mathscr{B}$ with σ-algebra $\sigma_{\mathscr{B}}$ is available, a Bayesian optimal design can be found by maximizing

$$\Psi(\xi) = \int_{\mathscr{B}} \Phi(M(\xi, \beta))\pi(\mathrm{d}\beta)$$

with a criteria function Φ on the space of all information matrices.

The prior distribution π is invariant under g respectively $\tilde{g}$ if $\pi^{\tilde{g}} = \pi$ or to be precise $\pi(\tilde{g}^{-1}(B)) = \pi(B)$ for all $B \in \sigma_{\mathscr{B}}$. In further consequence it is necessary that the parameter space $\mathscr{B}$ is invariant under $\tilde{g}$, that is $\tilde{g}(\mathscr{B}) = \mathscr{B}$.

If moreover Φ is invariant for all $g \in G$ a transformation of the design ξ under g leads to the same value as without transformation. This invariance is used in $(*)$.

$$\begin{aligned}
\Psi(\xi^g) &= \int_{\mathscr{B}} \Phi(M(\xi^g, \beta))\pi(\mathrm{d}\beta) \\
&= \int_{\mathscr{B}} \Phi(M(\xi^g, \tilde{g}(\tilde{g}^{-1}(\beta))))\pi(\mathrm{d}\beta) \\
&\overset{(*)}{=} \int_{\mathscr{B}} \Phi(M(\xi, \tilde{g}^{-1}(\beta)))\pi(\mathrm{d}\beta) \\
&= \int_{\tilde{g}^{-1}(\mathscr{B})} \Phi(M(\xi, \beta))\pi^{\tilde{g}^{-1}}(\mathrm{d}\beta) \\
&\overset{(**)}{=} \int_{\mathscr{B}} \Phi(M(\xi, \beta))\pi(\mathrm{d}\beta) \\
&= \Psi(\xi)
\end{aligned}$$

One fact used in $(**)$ is that because of the finiteness of G there exists an $n \in \mathbb{N}$ with $\tilde{g}^{-1} = \tilde{g}^n$, hence $\pi^{\tilde{g}^{-1}} = \pi^{\tilde{g}^n} = (\pi^{\tilde{g}})^{\tilde{g}^{n-1}} = \pi^{\tilde{g}^{n-1}} = \ldots = \pi^{\tilde{g}} = \pi$.

In the case of the Bayesian D-criterion

$$\Psi_D(\xi) = \int_{\mathscr{B}} \log(\det(M(\xi, \beta)))\pi(\mathrm{d}\beta),$$

which is concave according to Firth and Hinde (1997), the prerequisites in section 2 for the symmetrization are fulfilled. Hence the optimal designs can be found in the class of invariant designs.

The second possibility to avoid the dependence of designs on the unknown value of the parameter β is the maximin design which corresponds to the invariant criterion Φ and maximizes

$$\Psi(\xi) := \inf_{\beta \in \mathscr{B}} \Phi(M(\xi, \beta)).$$

If $\mathscr{B}$ is also invariant under $\tilde{g}$ or equivalently under $\tilde{g}^{-1}$ then

$$\begin{aligned}
\Psi(\xi^g) &= \inf_{\beta \in \mathscr{B}} \Phi(M(\xi^g, \beta)) = \inf_{\beta \in \mathscr{B}} \Phi(M(\xi^g, \tilde{g}(\tilde{g}^{-1}(\beta)))) \\
&= \inf_{\beta \in \mathscr{B}} \Phi(M(\xi, \tilde{g}^{-1}(\beta))) = \inf_{\beta \in \tilde{g}^{-1}(\mathscr{B})} \Phi(M(\xi, \beta)) = \inf_{\beta \in \mathscr{B}} \Phi(M(\xi, \beta)) \\
&= \Psi(\xi)
\end{aligned}$$

Graßhoff and Schwabe (2008) used this fact not in the direct sense of optimality but of efficiency of maximin designs for the Bradley-Terry paired comparison model. The efficiency corresponding to the criterion Φ has the form

$$\mathrm{eff}_{\Phi}(\xi,\beta) = h(\Phi(M(\xi,\beta)),\Phi(M(\xi_{\beta}^{*},\beta)))$$

with a special function h and the locally Φ-optimal design ξ_{β}^{*} with respect to β. For example in the D-optimal case: $\mathrm{eff}_{D}(\xi,\beta) = (\frac{\det(M(\xi,\beta))}{\det(M(\xi_{\beta}^{*},\beta))})^{1/p}$.

If the criterion function Φ is homogeneous and concave, then the corresponding efficiency $\mathrm{eff}_{\Phi}(\xi,\beta)$ is obviously also concave in ξ for every β. As pointed out in Graßhoff and Schwabe (2008) then the criterion of maximin efficiency shares the property of concavity. This together with the invariance of the maximin efficience criterion

$$\Psi_{\mathrm{eff}_{\Phi}}(\xi) = \inf_{\beta\in\mathscr{B}} h(\Phi(M(\xi,\beta)),\Phi(M(\xi_{\beta}^{*},\beta)))$$

establishes that each design ξ is dominated by its symmetrization $\overline{\xi}$. Hence, globally maximin efficient designs can be found in the class of invariant designs if the parameter region $\mathscr{B}$ is invariant.

The invariance of the maximin efficience criterion can be easily seen by

$$\begin{aligned}
\Psi_{\mathrm{eff}_{\Phi}}(\xi^{g}) &= \inf_{\beta\in\mathscr{B}} h(\Phi(M(\xi^{g},\beta)),\Phi(M(\xi_{\beta}^{*},\beta)))\\
&= \inf_{\beta\in\mathscr{B}} h(\Phi(M(\xi^{g},\tilde{g}(\beta))),\Phi(M(\xi_{\tilde{g}(\beta)}^{*},\tilde{g}(\beta))))\\
&= \inf_{\beta\in\mathscr{B}} h(\Phi(M(\xi^{g},\tilde{g}(\beta))),\Phi(M(\xi_{\beta}^{*g},\tilde{g}(\beta))))\\
&= \inf_{\beta\in\mathscr{B}} h(\Phi(M(\xi,\beta)),\Phi(M(\xi_{\beta}^{*},\beta)))\\
&= \Psi_{\mathrm{eff}_{\Phi}}(\xi).
\end{aligned}$$

5 Example (continued)

We want to revisit the example in section 3 – the 2-dimensional Poisson regression without interactions. But the focus is to find an Bayesian D-optimal design containing exactly 3 equally weighted support points on the design region $[-1,1]^2$. The prior distributions for β_1 and β_2 should be a uniform distribution being symmetrical to 0, that is $\beta_1,\beta_2 \sim \mathrm{U}(-a,a)$, i.i.d., with $a \in (0,\infty)$. The prior distribution for β_0 can be any distribution. Then it can be shown that these 3 points have the form $\{(-1,1),(-1,-1),(1,d)\}$, $\{(-1,-1),(1,-1),(d,1)\}$, $\{(1,-1),(1,1),(-1,d)\}$ or $\{(1,1),(-1,1),(d,-1)\}$ with $d \in [-1,1]$.

Consider the group G generated by the 90° rotation $g_{90}:\mathbb{R}^2 \to \mathbb{R}^2$, $g_{90}(x_1,x_2) = (-x_2,x_1)$, and the reflection $g_{\mathrm{ref}}:\mathbb{R}^2 \to \mathbb{R}^2$, $g_{\mathrm{ref}}(x_1,x_2) = (-x_1,x_2)$. Then there are 8 transformations in G. The design space is invariant under g as well as the prior distribution of the parameter space is invariant under $\tilde{g}$ for all $g \in G$. So the designs $\{(-1,1),(-1,-1),(1,d)\}$, $\{(-1,1),(-1,-1),(1,-d)\}$, $\{(-1,-1),(1,-1),(d,1)\}$, $\{(-1,-1),(1,-1),(-d,1)\}$, $\{(1,-1),(1,1),(-1,d)\}$, $\{(1,-1),(1,1),(-1,-d)\}$, $\{(1,1),(-1,1),(d,-1)\}$ and $\{(1,1),(-1,1),(-d,-1)\}$ with fixed $d \in [-1,1]$ have the same value of the Bayesian criterion function. And if one of these designs is optimal (in the set of all 3-point designs), then all of these designs are optimal (in the set of all 3-point designs).

Changing only the design region to $\{-1,1\}^2$ the group G of transformation is only generated by g_{90} and the Bayesian D-optimal 3-point designs are $\{(-1,1),(-1,-1),(1,-1)\}$, $\{(-1,-1),(1,-1),(1,1)\}$, $\{(1,-1),(1,1),(-1,1)\}$ and $\{(1,1),(-1,1),(-1,-1)\}$. The symmetrized design consists of the 4 equally weighted support points $(1,1)$, $(-1,1)$,$(-1,-1)$ and $(1,-1)$. This is the only invariant design with respect to G, so it is optimal over all possible designs.

6 Discussion

Although locally optimal designs are not invariant, in general, symmetry considerations can be useful for weighted optimality or maximin criteria, when the weight function or the parameter region is invariant. This may substantially simplify the optimization problem since attention may be restricted to the class of invariant (*symmetric*) designs.

The above findings may be extended to infinite transformation groups like rotations on a circle or on a k-dimensional sphere. Then the averaging (*symmetrization*) has to be performed by the corresponding Haar (Lebesgue) measure which is uniform on the orbits. The resulting invariant designs are continuous and have to be discretized. For rotations this can typically be achieved without loss of information.

References

1. Atkinson, A.C., Donev, A.N., Tobias, R.D.: Optimum Experimental Designs, with SAS. Oxford University Press, Oxford (2007)
2. Firth, D., Hinde, J.: On Bayesian D-Optimum Design Criteria and the Equivalence Theorem in Non-Linear Models. J. R. Stat. Soc. Ser. B Methodol. **59**, 793–797 (1997)
3. Ford, I., Torsney, B., Wu, C.F.J.: The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems. J. R. Stat. Soc. Ser. B Methodol. **54**, 569–583 (1992)
4. Gaffke, N., Heiligers, B.: Approximate Designs for Polynomial Regression: Invariance, Admissibility, and Optimality. In: Ghosh, S., Rao, C.R. (eds.) Handbook of Statistics **13**, pp. 1149-1199. Elsevier, Amsterdam (1996)
5. Graßhoff, U., Schwabe, R.: Optimal designs for the Bradley-Terry paired comparison model. Stat. Methods Appt. **17**, 275–289 (2008)
6. McCulloch, C.E., Searle, S.R.: Generalized, Linear, and Mixed Models. Wiley (2004)
7. Pukelsheim, F.: Optimal design of experiments. Wiley-Interscience, New York (1993)

8. Russell, K.G., Woods, D.C., Lewis, S.M., Eccleston, J.A.: D-optimal designs for Poisson regression models. Stat. Sinica **19**, 721–730 (2009)
9. Schmidt, D., Schwabe, R.: On optimal designs for censored data. Metrika **78**, 237–257 (2015)
10. Schwabe, R.: Optimum Designs for Multi-Factor Models. Springer, New York (1996)