Efficient Paired Comparison Designs for Utility Elicitation

HEIKO GROSSMANN
University of Münster, Institute for Psychology IV, Fliednerstr. 21, D-48149 Münster, Germany

ULRIKE GRASSHOFF
Free University Berlin, Institute for Mathematics I, Arnimallee 2–6, D-14195 Berlin, Germany

HEINZ HOLLING
University of Münster, Institute for Psychology IV, Fliednerstr. 21, D-48149 Münster, Germany

RAINER SCHWABE
University of Tübingen, Department of Medical Biometry, Westbahnhofstr. 55, D-72070 Tübingen, Germany — &
Free University Berlin, Institute for Mathematics I

ABSTRACT: In applications often data are available from the comparison of two alternatives rather than the direct valuation of a single object on its own. Experiments have to be designed for these paired comparisons in a different way than for standard situations. In this note we deal with a problem arising from organizational psychology and present an application of design considerations to the estimation of utility functions.

KEYWORDS: Paired Comparisons; D–optimal design; utility function

1 Introduction

The Productivity Measurement and Enhancement System ProMES (Pritchard, 1990) is a well–known technique in industrial and organizational psychology for measuring and improving organizational productivity. In a recent analysis of a large number of utility functions used there the authors have found that a large proportion of these functions can be well approximated by the equation

$$\mu(t;\theta_0,\theta_1,\theta_2) = \theta_0 + \theta_1 t^{\theta_2},$$

where $\theta_2 > 0$ and the stimulus t may vary over the unit interval, $t \in [0,1]$. Note that this equation is non–linear in the parameter θ_2. By letting $z = -\ln t$ the model can be recognized as a rescaled version of the exponential growth curve associated with the Mitscherlisch law of diminishing returns $\theta_0 + \theta_1 e^{-\theta_2 z}$, $z \in [0,\infty)$. Optimal designs for this response have been investigated by Box and Lucas (1959).

As in various fields of statistical applications including psychological aspects of market research and psychophysical investigations in medicine it often happens in the present context that data are available from the comparison of two (or more) objects rather than the direct valuation of a single object on its own. Additionally, the technique of matching, i. e. of taking differences, is also commonly applied when individual effects occur in paired observations which cannot be modeled appropriately (individual intercepts, correlations, random effects). Evidently, for these *paired comparisons* the experiments should be designed in a different way than for standard experimental situations (see e. g. van Berkum, 1987, Großmann et al., 2000).

Therefore, we consider the problem of designing experiments for estimating the model parameters θ_1 and θ_2 from paired comparison data as follows: For the comparison of two alternative stimuli s and t within an ordered pair (s,t) the response is given by

$$Y(s,t) = \eta(s,t;\theta_1,\theta_2) + Z(s,t),$$

where

$$\eta(s,t;\theta_1,\theta_2) = \mu(s;\theta_0,\theta_1,\theta_2) - \mu(t;\theta_0,\theta_1,\theta_2) = \theta_1(s^{\theta_2} - t^{\theta_2})$$

is the mean response obtained in the comparison. The design region $\mathcal{X} = [0,1]^2$ is given by the Cartesian product as the stimuli presented may be chosen independently. Furthermore, the error distribution $Z(s,t)$ is assumed to belong to a zero mean exponential family and repeated observations are taken to be uncorrelated and homoscedastic.

As usually, in non–linear models (generalized) designs ξ are considered which are defined as finitely supported probability measures on the design region $\mathcal{X}$ (see Kiefer, 1974, for a motivation and discussion of such designs in linear settings). The quality of the design ξ will be assessed through the asymptotic information matrix

$$M(\xi;\theta) = \int_{\mathcal{X}} M(s,t;\theta)\,\xi(d(s,t)),$$

where

$$M(s,t;\theta) = \left(\frac{\partial}{\partial\theta_i}\eta(s,t;\theta)\frac{\partial}{\partial\theta_j}\eta(s,t;\theta)\right)_{i,j=1,2}$$

is the (asymptotic) information associated with the single comparison of the alternatives s and t. In general, no information is contained in a comparison of a stimuli with itself, $M(t,t;\theta) = 0$. The locally linearized regression functions $f_\theta(s,t) = \left(\frac{\partial}{\partial\theta_i}\eta(s,t;\theta)\right)_{i=1,2}$ are given by $s^{\theta_2} - t^{\theta_2}$ and $\theta_1(s^{\theta_2}\ln s - t^{\theta_2}\ln t)$ for $i = 1,2$, respectively. Note that under suitable regularity conditions the information matrix $M(\xi;\theta)$ is proportional to the inverse of the asymptotic covariance matrix for estimating θ. The most popular performance measure of a design ξ is the D–criterion which aims at the maximization of the determinant $|M(\xi;\theta)|$ of the information matrix and which is equivalent to the minimization of the determinant of the asymptotic covariance matrix. We will adopt this criterion in the sequel.
In contrast to the linear design theory the information matrix $M(\xi;\theta)$ depends on the unknown parameter vector θ. This dependence carries over to the problem of finding a D–optimal design. Hence, only locally D–optimal designs can be found for each θ, separately. As can be seen from the structure of the linearized regression functions this effect is crucial for the parameter θ_2 which acts non–linearly in the model. In contrast to that the linear parameter θ_1 comes in only as a constant factor which does not affect the optimization problem

2 Reduction to the Canonical Form

For every particular parameter value θ the transformation $g_\theta : (s,t) \to (s^{1/\theta_2}, t^{1/\theta_2})$ induces a correspondence $\xi \to \xi^{g_\theta}$ on the set of all designs on $\mathcal{X}$. Using this relation we have $|M(\xi;(1,1))| = \theta_2^2\theta_1^{-2}|M(\xi^{g_\theta};\theta)|$ and the left–hand side of this equation is maximized by a design ξ^* if and only if the right–hand side is maximized by the corresponding image under g_θ. Hence, it suffices to consider the design problem for $\theta_1 = \theta_2 = 1$ and the optimal design can be obtained by the inverse transformation g_θ^{-1} of the support. In this sense the transformation g_θ leads to a canonical form in the spirit of Ford, Torsney and Wu (1992) and Sitter and Torsney (1995). In the sequel we will write $M(\xi) = M(\xi;(1,1))$ for brevity.
Our approach to identify efficient designs is based on geometric ideas and rests on the well–known result that the support of a D–optimal design is included in the set of those points where a minimal circumscribing ellipsoid centered at the origin touches the induced design space (Sibson, 1972; Silvey and Titterington, 1973). This situation is depicted in Figure 1 for the induced design space $\{(s-t, s\ln s - t\ln t); (s,t) \in \mathcal{X}\}$.
If we ignore comparisons which are obtained by interchanging the alternatives of an ordered pair (s,t), in view of $M(t,s) = M(s,t)$, we can see that the support of a D–optimal design contains two or three points in accordance with Carathéodory's theorem (see Silvey, 1980). Due to the structure of the induced design region these points can be characterized by the three

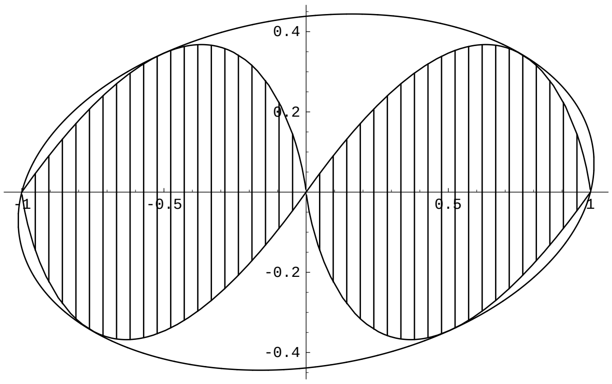

FIGURE 1. Induced design space with circumscribing ellipsoid.

settings $(1,0)$, $(\tau_1,0)$ and $(1,\tau_2)$ with $\tau_1,\tau_2 \in [0,1]$.

3 The Best Two–Point Design

In this section we consider two–point designs. As they have minimal support their optimal weights are distributed equally over the two supporting settings. Hence, by a majorization argument we can confine ourselves to designs of the form $\xi_2 = \frac{1}{2}\epsilon_{(\tau_1,0)} + \frac{1}{2}\epsilon_{(1,\tau_2)}$ where $\epsilon_{(s,t)}$ denotes the (one–point) Dirac measure on the setting (s,t). The corresponding value for the D–criterion is given by $|M(\xi_2)| = \frac{1}{4}((1-\tau_2)\tau_1 \ln\tau_1 + \tau_1\tau_2 \ln\tau_2)^2$. A closed form solution for the maximization problem does not seem to be feasible. Numerical optimization yields the optimal settings $\tau_1^* = 0.552$ and $\tau_2^* = 0.203$ with the determinant $|M(\xi_2^*)| = 0.0484$ for the corresponding design ξ_1^*.

Checking the condition of the celebrated equivalence theorem (Kiefer and Wolfowitz, 1960; White, 1973) reveals that ξ_2^* is not D–optimal within the set of all competing designs since the maximum of the variance function $d(s,t;\xi_2^*) = f_\theta^\top(s,t)M(\xi_2^*)f_\theta(s,t)$ exceeds the number $p=2$ of parameters. A closer inspection shows that the maximum is attained at the comparison $(1,0)$ of the extreme settings, $\sup_{s,t} d(s,t;\xi_2^*) = d(1,0;\xi_2^*) = 2.193$. According to Atwood (1969) a lower bound for the D–efficiency of ξ_2^* is given by $\mathrm{eff}_D(\xi_2^*) \geq \mathrm{eff}_G(\xi_2^*) = 2/\sup_{s,t} d(s,t;\xi_2^*) \geq 91.1\%$, where $\mathrm{eff}_G(\xi)$ is the G–efficiency. To our experience this lower bound is very conservative. Hence, we are prompted to consider three–point designs where the comparison $(1,0)$ of extreme levels is added to the support.

4 Efficient Three–Point Designs

In general, an optimal three–point design is of the form

$$\xi_3 = w_1 \epsilon_{(\tau_1,0)} + w_2 \epsilon_{(1,\tau_2)} + (1 - w_1 - w_2)\epsilon_{(1,0)}$$

due to geometric considerations. As a first approach it is tempting to consider the optimal settings obtained by Box and Lucas (1959) for the directly observable response, i. e. $\tau_1 = \tau_2 = e^{-1}$ besides the extreme settings 0 and 1. Note that these settings are also optimal in the presence of individual intercepts (Schwabe, 1995). The optimal weights are found to be the equi–distribution, $w_1 = w_2 = \frac{1}{3}$, and the corresponding value for the determinant of the information matrix, 0.0451, does not exceed the corresponding value for the best two–point design of the previous section. This observation is in accordance with results by Atkins and Cheng (1999) for correlated observations coming in pairs and by Großmann et al. (2000) in a polynomial setting. However, it is worth while mentioning that the setting $\tau_1 = \tau_2 = e^{-1}$ is optimal when equal weights $w_1 = w_2 = 1 - w_1 - w_2 = \frac{1}{3}$ are assumed.
To facilitate the further calculations we confine ourselves to designs with equal weights on the non–extreme comparisons, $w_1 = w_2 = w$, i. e.

$$\xi_3 = w\epsilon_{(\tau_1,0)} + w\epsilon_{(1,\tau_2)} + (1 - 2w)\epsilon_{(1,0)}.$$

As in the previous section the optimization problem for the determinant of the information matrix

$$M(\xi_3) = \begin{pmatrix} w(\tau_1^2 + (1-\tau_2)^2) + (1-2w) & w(\tau_1^2 \ln \tau_1 - (1-\tau_2)\tau_2 \ln \tau_2) \\ w(\tau_1^2 \ln \tau_1 - (1-\tau_2)\tau_2 \ln \tau_2) & w((\tau_1 \ln \tau_1)^2 + (\tau_2 \ln \tau_2)^2) \end{pmatrix}$$

does not seem to be amenable to a full analytical solution. However, a necessary condition for the optimal weight w in terms of the levels τ_1 and τ_2 is given by

$$w = -\frac{(\tau_1 \ln \tau_1)^2 + (\tau_2 \ln \tau_2)^2}{2[(\tau_1 \ln \tau_1)^2(\tau_2^2 - 2\tau_2 - 1) - 2\tau_1^2\tau_2(\tau_2 - 1)\ln \tau_1 \ln \tau_2 + (\tau_1^2 - 2)(\tau_2 \ln \tau_2)^2]}.$$

Using this equation in the calculation of $|M(\xi_3)|$ optimal levels $\tau_1^* = 0.506$ and $\tau_2^* = 0.242$ are found numerically with the optimal common weight $w^* = 0.417$. The determinant of the information matrix for the corresponding design ξ_3^* equals $|M(\xi_3^*)| = 0.0493$. The inspection of the variance function $d(s,t;\xi_3^*)$ for the design ξ_3^*, which is shown in Figure 2, yields $\sup_{s,t} d(s,t;\xi_3^*) = d(\tau_1^*, 0; \xi_3^*) = 2.0015$ and thus reveals the nearly optimality of ξ_3^*. In fact, as can be seen from $\mathrm{eff}_D \geq 2/\sup_{s,t} d(s,t;\xi_3^*) \geq 99.9\%$ the design is almost D–optimal.
Note that in the present setting of paired comparisons the variance function vanishes on the main diagonal, $d(t,t;\xi) = 0$. Hence, it is preferable to draw

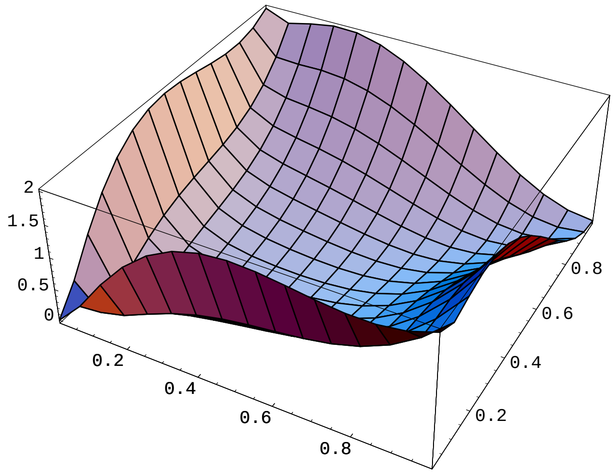

FIGURE 2. Variance function for the design ξ_3^*.

the surface of the variance function $d(s,t;\xi)$ rather than the sometimes more appealing information function $1/d(s,t;\xi)$.
In view of the efficiency of the design ξ_3^* the best two–point design ξ_2^* can be re–evaluated. By calculating the relative efficiency with respect to ξ_3^* a D–efficiency $\mathrm{eff}_D(\xi_2^*) \geq 99.0\%$ is obtained. Hence, the two–point design ξ_2^* is a good competitor when an even number of observations are to be taken. For practical applications the weights of a design have to be transformed to numbers of observations to be allocated to the corresponding settings. Apart from those fortunate situations when the weights are a multiple of $1/n$, where n is the sample size in the experiment, those weights have to be rounded. For example, in the present setting the optimal weight $w^* = 0.417$ may apparently be replaced by $w' = 2/5$. This results in a design ξ_3' with $|M(\xi_3')| = 0.0492$ and a D–efficiency which is still as high as $\mathrm{eff}_D(\xi_3') \geq 99.9\%$. Thus the design ξ_3' can be used as a highly efficient surrogate for the D–optimal design in the case that the number n of observations is a multiple of 5.

5 Discussion

We considered the model equation $Y(s,t) = \theta_1(s^{\theta_2} - t^{\theta_2}) + Z(s,t)$ of paired comparisons when the utility follows a logarithmically rescaled exponential growth curve. In this situation we have obtained a highly efficient ($\geq 99.9\%$) three–point design supported by the settings $(\tau_1^*, 0)$, $(1, \tau_2^*)$ and $(1, 0)$, where $\tau_1^* = 0.506$ and $\tau_2^* = 0.242$, with weights $2/5$, $2/5$ and $1/5$, respectively, when $\theta_2 = 1$. Similarly, a highly efficient ($\geq 99.0\%$) two–point design was found which is supported by the settings $(\tau_1^*, 0)$ and $(1, \tau_2^*)$, where, now, $\tau_1^* = 0.552$ and $\tau_2^* = 0.203$, with equal weights 1/2. For arbitrary θ_2 the efficient settings are determined by replacing τ_1^* and τ_2^* by their corresponding images $(\tau_1^*)^{\theta_2}$ and $(\tau_2^*)^{\theta_2}$ under the reversed canonical transformation $g_\theta^{-1}(t) = t^{\theta_2}$ while the weights remain unchanged.

Acknowledgments: The authors would like to thank Prof. R. D. Pritchard for providing a copy of the ProMES database. Part of this work was supported by the *Deutsche Forschungsgemeinschaft* under grant Ho 1286/2-1.

References

Atkins, J. E. and Cheng, C.-S. (1999). Optimal regression designs in the presence of random block effects. *J. Statist. Plann. Inference*, **77**, 321–335

Atwood, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist.*, **40**, 1570–1602.

Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in non–linear situations. *Biometrika*, **46**, 77–90

Ford, I., Torsney, B. and Wu, C. J. F. (1992). The use of a canonical form in the construction of locally optimal designs for non–linear problems. *J. R. Statist. Soc., Ser. B*, **54**, 569–583.

Großmann, H., Holling, H., Graßhoff, U. and Schwabe, R. (2000). Efficient designs for paired comparisons with a polynomial factor. In *Optimum Design 2000* (Atkinson, A. C., Bogacka, B. and Zhigljavsky, A. A., eds.) Kluwer (to appear).

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.*, **2**, 849–879.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. *Can. J. Math.*, **12**, 363–366.

Pritchard, R. D. (1990). *Measuring and improving organizational productivity: A practical guide.* Praeger, New York.

Schwabe, R. (1995). Designing experiments for additive nonlinear models. In *MODA 4 - Advances in Model–Oriented Data Analysis* (Kitsos, C. P. and Müller, W. G., eds.) Physica, Heidelberg, 77–85.

Sibson, R. (1972). Contribution to discussion of papers by H. P. Wynn and P. J. Laycock. *J. R. Statist. Soc., Ser. B*, **34**, 181–183.

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

Silvey, S. D. and Titterington, D. M. (1973). A geometric approach to optimal design theory. *Biometrika*, **60**, 21–32.

Sitter, R. R. and Torsney, B. (1995). D–optimal designs for generalized linear models. In *MODA 4 - Advances in Model–Oriented Data Analysis* (Kitsos, C. P. and Müller, W. G., eds.). Physica, Heidelberg, 87–102.

van Berkum, E. E. M. (1987). *Optimal Paired Comparison Designs for Factorial Experiments*, *CWI Tract*, **31**.

White, L. V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika*, **60**, 345–348.