Listy Biometryczne - Biometrical Letters
Vol. 34 (1997), No. 1, 27-34

Relationships between some classes of optimality criteria

Barbara Bogacka and Stanisław Mejza

Department of Mathematical and Statistical Methods,
Agricultural University of Poznań,
Wojska Polskiego 28, 60-637 Poznań, Poland

SUMMARY

The paper deals with three wide classes of optimality criteria of experimental designs: Universal Optimality, involving the classical definition of convexity of a function, General Optimality, involving a combination of differentiable functions and B-Universal Optimality, where the Schur convexity is used. Two theorems on relationships between pairs of these classes are proved and some corollaries drawn. The applicability of the considered criteria to designs under fixed and also mixed linear models makes the results general.

KEY WORDS: optimality criteria, experimental designs, information matrix

1. Introduction and preliminaries

Let D be a class of experimental designs under a linear model

$$\{\mathbf{y}, \mathbf{X}^T\boldsymbol{\tau}, \mathbf{V}(\boldsymbol{\sigma})\}, \tag{1}$$

where $\mathbf{y}$ is an n-dimensional observable random vector with an expected value $E(\mathbf{y}) = \mathbf{X}^T\boldsymbol{\tau}$ and a dispersion matrix $\mathrm{Cov}(\mathbf{y}) = \mathbf{V}(\boldsymbol{\sigma}) = \sum_{i=1}^{m}\mathbf{V}_i\sigma_i^2$, which depends on an unknown vector $\boldsymbol{\sigma} = (\sigma_1^2, ..., \sigma_m^2)^T$ of variance components; $\mathbf{V}_i$ are nonnegative definite known matrices, $\boldsymbol{\tau}$ denotes a v-dimensional vector of fixed parameters, and $\mathbf{X}$ is a known design matrix for the parameters $\boldsymbol{\tau}$.

Most of optimality criteria used in the literature are functionals of information matrices. We define an information matrix, say $\boldsymbol{\Lambda}_d$, for the parameters $\boldsymbol{\theta}^T = (\boldsymbol{\tau}^T, \boldsymbol{\sigma}^T)$, of a design $d \in D$ under the model (1) as a dispersion matrix of a random vector of derivatives of a likelihood function of $\mathbf{y}$ with respect to the parameters $\boldsymbol{\theta}^T = (\boldsymbol{\tau}^T, \boldsymbol{\sigma}^T)$;

that is,

$$\mathbf{\Lambda}_d = \mathrm{Cov}\left(\frac{\partial L(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right).$$

Assuming an n-dimensional normal distribution of the vector $\mathbf{y}$ we have (cf. Searle, 1970; Hocking, 1985)

$$\mathbf{\Lambda}_d = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} \end{bmatrix}, \tag{2}$$

where $\mathbf{M} = \mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$ is an information matrix for the fixed parameters $\boldsymbol{\tau}$ and $\mathbf{\Omega} = \left\{\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{V}_i\mathbf{V}^{-1}\mathbf{V}_j)\right\}_{i,j=1,\ldots,m}$ is an information matrix for the variance components $\boldsymbol{\sigma}$. Let us note that both $\mathbf{M}$ and $\mathbf{\Omega}$ depend on the dispersion matrix of the model, i.e., on the unknown variance components. This makes searching for optimum designs much more difficult than in the case of fixed models, where $\mathbf{V} = \sigma^2\mathbf{I}$, with σ^2 being a positive scalar and $\mathbf{I}$ – the identity matrix. For fixed models we have

$$\mathbf{\Lambda}_d = \begin{bmatrix} \frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X} & \mathbf{0} \\ \mathbf{0} & \frac{n}{2(\sigma^2)^2} \end{bmatrix}$$

and the matrix $\mathbf{X}^T\mathbf{X}$ is usually considered as the information matrix. The matrix (2) plays an essential role in the optimality theory as it has an important statistical interpretation, namely, its inverse is an asymptotic covariance matrix of the maximum likelihood (ML) estimator of the parameters:

$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}) \to \begin{bmatrix} \mathbf{M} & 0 \\ 0 & \mathbf{\Omega} \end{bmatrix}^{-1}, \tag{3}$$

where $(\widehat{\boldsymbol{\theta}})$ denotes the ML estimator of the vector $\boldsymbol{\theta}$. In the fixed model we are interested in $\mathrm{Cov}(\hat{\boldsymbol{\tau}}) = \mathbf{M}^{-1}$. These facts give rise to many definitions of optimality criteria, which are functionals of information matrices. The asymptotic property (3) of the inverse of the information matrix in the case of mixed models makes the criteria useful for finding optimum designs for estimation of fixed parameters in such models, too. However, in the mixed models there arise the problem of the dependence of the information matrix on the unknown variance components. Optimality criteria for the whole set of parameters $\boldsymbol{\theta}^T = (\boldsymbol{\tau}^T, \boldsymbol{\sigma}^T)$ should be defined separately, for example as a combination of some functionals of the information matrix $\mathbf{M}$ for fixed parameters and the information matrix $\mathbf{\Omega}$ for variance components (Giovagnoli and Sebastiani, 1989; Bogacka, 1995).

Throughout the paper we consider the optimality criteria for fixed parameters and the results are general, i.e., they do not depend on the type of the model of observations, and the results may be applied to fixed or mixed linear models (Bogacka and Mejza, 1996).

2. Definitions of optimality

First we introduce some notation. Let $\mathfrak{M}(d) = \{\mathbf{M}_d : d \in D\}$ denote a set of information matrices for fixed parameters of designs $d \in D$, let $\mathfrak{M} = \{\mathbf{M} \geq 0\}$ be a set of nonnegative or positive definite $(v \times v)$-dimensional matrices of rank s, whose eigenvalues are not greater than the largest eigenvalue of any $\mathbf{M}_d \in \mathfrak{M}(d)$. Furthermore, let $\boldsymbol{\lambda}(\mathbf{M}) = (\lambda_1(\mathbf{M}), ..., \lambda_s(\mathbf{M}))^T$ denote a vector of positive eigenvalues of a matrix $\mathbf{M}$ and $\boldsymbol{\lambda}(\mathfrak{M}) = \{\boldsymbol{\lambda}(\mathbf{M}) : \mathbf{M} \in \mathfrak{M}\}$ denote a set of vectors of eigenvalues of matrices $\mathbf{M} \in \mathfrak{M}$.

Given a function $\Phi : \mathfrak{M} \to \mathbb{R}_+$ we say that a design $d^* \in D$ is Φ-optimum in the class D if its information matrix $\mathbf{M}_{d^*}$ fulfills

$$\Phi(\mathbf{M}_{d^*}) \leq \Phi(\mathbf{M}_d) \qquad \text{for all } d \in D.$$

Bondar (1983), Kiefer (1975), Cheng (1978) and Shah and Sinha (1989) considered some general definitions referring to very wide sets of optimality criteria. In all of these definitions the functionals have to fulfill some conditions, usually dealing with a kind of convexity (or concavity), invariance of the functional with respect to permutations of rows or/and columns of the information matrix, a kind of monotonicity and other conditions. The differences among the definitions come from the different ways of defining some of these conditions, mainly convexity. Now, let us recall the first three of these definitions.

DEFINITION 1. (Kiefer, 1975) A design $d^* \in D$ is *universally optimal* in the class D if its information matrix $\mathbf{M}_{d^*}$ minimises every $\Phi : \mathfrak{M} \to \mathbb{R}_+$ satisfying the following conditions:

a) $\Phi(\mathbf{M}) = \Phi(\mathbf{P}^T\mathbf{M}\mathbf{P})$, where $\mathbf{P}$ is any $(v \times v)$-dimensional permutation matrix,
b) Φ is a convex function, i.e., $\Phi(\alpha\mathbf{M}_1 + (1-\alpha)\mathbf{M}_2) \leq \alpha\Phi(\mathbf{M}_1) + (1-\alpha)\Phi(\mathbf{M}_2)$,
c) $\Phi(b\mathbf{M})$ is nonincreasing in a scalar $b > 0$.

We denote the functionals defined on $\boldsymbol{\lambda}(\mathfrak{M})$ by lower case letter φ. A useful family of optimality criteria fulfilling the conditions of Definition 1, defined on eigenvalues of information matrices, is

$$\varphi_p(\boldsymbol{\lambda}(\mathbf{M})) = \left[\frac{1}{s}\sum_{i=1}^{s}\lambda_i^{-p}(\mathbf{M})\right]^{1/p}, \qquad p \in (0, \infty).$$

For $p = 1$, $p \to 0$ and $p \to \infty$ we have three functions equivalent to the known A, D and E optimality criteria, respectively:

$$\varphi_1(\boldsymbol{\lambda}(\mathbf{M})) = \frac{1}{s}\sum_{i=1}^{s}\lambda_i^{-1}(\mathbf{M}),$$

$$\lim_{p\to 0}\varphi_p(\boldsymbol{\lambda}(\mathbf{M})) = \prod_{i=1}^{s}\lambda_i^{-1}(\mathbf{M}),$$

$$\lim_{p\to 0}\varphi_p(\boldsymbol{\lambda}(\mathbf{M})) = \max_i \lambda_i^{-1}(\mathbf{M}).$$

Kiefer (1975) also took into account a functional $\varphi(\boldsymbol{\lambda}(\mathbf{M}))$, which is a combination of convex real functions. However, it was Cheng, who presented and considered formal definition of the so called general optimality.

DEFINITION 2. (Cheng, 1978) A design $d^* \in D$ is *generally optimal* in the class D if its information matrix $\mathbf{M}_{d^*}$ minimises every $\varphi : \boldsymbol{\lambda}(\mathfrak{M}) \to \mathbb{R}_+$ of the form

$$\varphi_f(\boldsymbol{\lambda}(\mathbf{M})) = \sum_{i=1}^{s} f(\lambda_i(\mathbf{M})),$$

where the function $f : [0, x_0] \to \mathbb{R}_+$, $x_0 = \max_{\mathfrak{M}(d)} tr\mathbf{M}$, satisfies the following conditions:

1) f is continuously differentiable on $(0, x_0)$,

2) first, second and third derivatives of f satisfy the following conditions on $(0, x_0) : f' < 0,\ f'' > 0,\ f''' < 0$,

3) $f(0) = \lim_{x\to 0_+} f(x) = \infty$.

The function f is strictly decreasing and strictly convex ($f' < 0$, $f'' > 0$) and its first derivative f' is strictly increasing and strictly concave ($f'' > 0$, $f''' < 0$). The third condition ensures that the designs for which the matrices $\mathbf{M}_d$ have eigenvalues near zero can not be optimal. The class of functions defined by Cheng also includes A and D optimality criteria. Namely, taking

$$f_1(x) = x^{-1} \qquad \text{and} \qquad f_2(x) = -\log x$$

we obtain

$$\varphi_{f_1}(\boldsymbol{\lambda}(\mathbf{M})) = \frac{1}{s}\sum_{i=1}^{s}\lambda_i^{-1}(\mathbf{M}) \quad \text{and} \quad \varphi_{f_2}(\boldsymbol{\lambda}(\mathbf{M})) = \log\prod_{i=1}^{s}\lambda_i^{-1}(\mathbf{M}).$$

Bondar (1983) considered another class of optimality criteria. He called the designs fulfilling conditions of his definition universally optimal designs. However, the conditions are not equivalent to the Kiefer's ones. So, we propose to term such designs

B-universally optimal (BU-optimal for short). Bondar's definition may equivalently be written as:

DEFINITION 3. A design $d^* \in D$ is *BU-optimal* in the class D if its information matrix $\mathbf{M}_{d^*}$ minimises every $\Phi : \mathfrak{M} \to \mathbb{R}_+$ satisfying the following conditions:

a') $\Phi(\mathbf{M}) \equiv \varphi(\boldsymbol{\lambda}(\mathbf{M}))$, i.e., $\Phi(\mathbf{M})$ is a function of eigenvalues of information matrices,

b') $\varphi(\boldsymbol{\lambda}(\mathbf{M}))$ is a Schur-convex function on a set $\boldsymbol{\lambda}(\mathfrak{M})$, i.e., $\boldsymbol{\lambda}(\mathbf{M}_1) \prec \boldsymbol{\lambda}(\mathbf{M}_2) \Longrightarrow \varphi(\boldsymbol{\lambda}(\mathbf{M}_1)) \leq \varphi(\boldsymbol{\lambda}(\mathbf{M}_2))$, where the symbol " $\prec$ " denotes majorization of $\boldsymbol{\lambda}(\mathbf{M}_1)$ by $\boldsymbol{\lambda}(\mathbf{M}_2)$ (cf. Marshall and Olkin, 1989, chapter 1),

c') $\lambda_{[i]}(\mathbf{M}_1) \geq \lambda_{[i]}(\mathbf{M}_2) \forall_i \Rightarrow \varphi(\boldsymbol{\lambda}(\mathbf{M}_1)) \leq \varphi(\boldsymbol{\lambda}(\mathbf{M}_2))$, where $\lambda_{[1]}(\mathbf{M}) \geq ... \geq \lambda_{[s]}(\mathbf{M})$.

Having three so wide definitions of optimality one can ask which of them should be preferably used. One of the ways to find an answer is to consider relationships between these definitions.

3. Results

3.1. Classes of optimality criteria

Comparison of the three definitions given in Section 2 is very difficult if there is no common space on which the functions Φ are defined. We restrict further consideration to functions defined on the set of vectors of positive eigenvalues of matrices $\mathbf{M}$, i.e., $\Phi(\mathbf{M}) \equiv \varphi(\boldsymbol{\lambda}(\mathbf{M}))$. Furthermore, let F_K, F_C and F_B stand for the following classes of optimality criteria:

$$\begin{aligned} F_K &= \{\varphi(\boldsymbol{\lambda}(\mathbf{M})) \quad \text{satisfying (a), (b), (c) of Def.1.}\}, \\ F_C &= \{\varphi(\boldsymbol{\lambda}(\mathbf{M})) = \sum_{i=1}^{s} f(\lambda_i(\mathbf{M})), \quad f \quad \text{satisfying (1), (2), (3) of Def.2.}\} \\ F_B &= \{\varphi(\boldsymbol{\lambda}(\mathbf{M})) \quad \text{satisfying (a'), (b'), (c') of Def.3.}\}. \end{aligned}$$

3.2. Relationships between the classes of optimality criteria

We have the following theorem:

THEOREM 1. *If a function $\varphi : \boldsymbol{\lambda}(\mathfrak{M}) \to \mathbb{R}_+$ belongs to the class F_C then the function also belongs to the class F_B, i.e.,*

$$F_C \subset F_B.$$

Proof. Let a function $\varphi(\boldsymbol{\lambda}(\mathbf{M})) = \sum_{i=1}^{s} f(\boldsymbol{\lambda}_i(\mathbf{M})) \in F_C$. The form of this function ensures that the condition (a') of the Definition 3 is fulfilled.

By the condition (2) of Definition 2 we have

$$\lambda_{[1]}(\mathbf{M}) \geq ... \geq \lambda_{[s]}(\mathbf{M}) \Rightarrow f'(\lambda_{[1]}(\mathbf{M})) \geq ... \geq f'(\lambda_{[s]}(\mathbf{M})),$$

where $f'(\lambda_{[i]}(\mathbf{M}))$, $i = 1, ..., s$, is the partial derivative of the functional φ with respect to $\lambda_{[i]}(\mathbf{M})$:

$$f'(\lambda_{[i]}(\mathbf{M})) = \frac{\partial\varphi(\boldsymbol{\lambda}(\mathbf{M}))}{\partial\lambda_{[i]}(\mathbf{M})}.$$

Hence, by the Theorem A3, p.56 in Marshall and Olkin (1989), the function φ is Schur-convex on

$$(\boldsymbol{\lambda}_{[\,]}(\mathfrak{M}) = \{\boldsymbol{\lambda}(\mathbf{M}) : \lambda_{[1]}(\mathbf{M}) \geq ... \geq \lambda_{[s]}(\mathbf{M})\}$$

and also on $\boldsymbol{\lambda}(\mathfrak{M})_{[\,]} \cap \boldsymbol{\lambda}(\mathfrak{M})$. Hence, by symmetry of $\boldsymbol{\lambda}(\mathfrak{M})$ and φ and by the Remark on p.54 in Marshall and Olkin (1989), φ is Schur-convex on $\boldsymbol{\lambda}(\mathfrak{M})$, which means that it satisfies the condition (b') of Definition 3.

Furthermore, f is a nonincreasing function, i.e.,

$$\lambda_{[i]}(\mathbf{M}_1) \geq \lambda_{[i]}(\mathbf{M}_2) \Rightarrow f(\lambda_{[i]}(\mathbf{M}_1)) \leq f(\lambda_{[i]}(\mathbf{M}_2)),$$

and φ satisfies the condition (c') of the Definition 3.

So, the function $\varphi \in F_C$ satisfies all conditions from the set F_B. □

Another inclusion is shown in Theorem 2.

THEOREM 2. *If a function* $\varphi : \boldsymbol{\lambda}(\mathfrak{M}) \to \mathbb{R}_+$ *belongs to the class* F_C *then the function also belongs to the class* F_K, *i.e.*,

$$F_C \subset F_K.$$

Proof. Let a function $\varphi \in F_C$. By the condition (2) of Definition 2 the function φ is convex and decreasing in a scalar $b > 0$. Furthermore, the form

$$\varphi(\boldsymbol{\lambda}(\mathbf{M})) = \sum_{i=1}^{s} f(\boldsymbol{\lambda}_i(\mathbf{M}))$$

ensures that this function fulfills the Kiefer's condition (a) of permutation invariance. So, the function $\varphi \in F_C$ satisfies all conditions from the set F_K. □

There is no such relation between the classes F_K and F_B. If the function $\Phi(\mathbf{M}) = \varphi(\boldsymbol{\lambda}(\mathbf{M}))$ is convex then $\varphi(\boldsymbol{\lambda}(\mathbf{M}))$ is a convex and Schur-convex function. So, we have

$$\{\varphi(\boldsymbol{\lambda}(\mathbf{M})) : (a), (c')\} \subset F_B.$$

However, condition (c') is stronger than (c), i.e., (c) does not imply (c').

From the above results follow the corollaries:

COROLLARY 1. *A universally optimal design is also generally optimal, φ_p-optimal and A-, D- and E-optimal design.*

COROLLARY 2. *A BU–optimal design is also generally optimal and A-, D-optimal design.*

4. Discussion

The relationships between classes of optimality criteria were considered also by Shah and Sinha (1989). They defined the so called *Extended Universal Optimality,* where somewhat different conditions, particularly suitable in the case of a fixed linear model, are imposed on the optimality functional (Chapter 1, p.7). A special kind of convexity, called 'weak convexity', is considered by them. The authors show inclusions of four sets of functionals (Chapter 1, p.6), however the sets are different than F_C, F_B, F_K considered here.

The criteria examined here are general and may be applied in the fixed or mixed linear model situations. But not all optimality criteria, introduced in fixed linear model case, can be directly adopted to the mixed linear model case. There are some problems, especially in mixed linear model case when dispersion structure of the model follows from a randomization model. The reader is referred to paper by Bogacka and Mejza (1996).

The paper by Bogacka and Mejza (1994) considers the experiment carried out in a block design under the mixed linear model. In particular, it presents the conditions for a generally balanced block design to be optimal with respect to the optimality criteria discussed here. In the characterisations of the optimal classes of block designs relationships between criteria considered here were used.

REFERENCES

Bogacka B. (1995). On information matrices for fixed and random parameters in Generally Balanced experimental block designs. *MODA4 - Advances in Model Oriented Data Analysis. Proc. of the 4th International Workshop.* Spetses, Greece, June 5-9, 1995, Christos P.Kitsos and Werner G. Mueller (eds.). Springer Verlag, 141-149.

Bogacka B., Mejza S. (1994). Optimality of generally balanced experimental block designs. *Proc. of the International Conference on Linear Statistical Inference, LINSTAT'93,* T. Caliński and R. Kala (eds.). Kluwer Academic Publishers, Dordrecht, 185-194.

Bogacka B., Mejza S. (1996). On optimality criteria of experimental block designs under linear mixed models. *Tatra Mountains Math. Publ.* **7**, 45-52.

Bondar J.V. (1983). Universal optimality of experimental designs: definitions and a criterion. *Can. J. Statist.* **11**, 325-331.

Cheng C.S. (1978). Optimality of certain asymmetrical experimental designs. *Ann. Statist.* **6**, 1239-1261.

Giovagnoli A., Sebastiani P. (1989). Experimental designs for mean and variance estimation in variance components models. *Computational Statistics and Data Analysis* **8**, 21-28.

Hocking R.R. (1985). *The Analysis of Linear Models.* Monterey, California: Brooks/Cole Publishing Company.

Kiefer J. (1975). Construction and optimality of generalized Youden designs. In: *A Survey of Statistical Design and Linear Models*, J.N. Srivastava (ed.), 333-353. North Holland Publishing Company, Amsterdam.

Marshall A.W., Olkin I. (1979). *Inequalities: Theory of Majorization and Its Application.* Academic Press, New York.

Searle S.R. (1970). Large sample variances of Maximum Likelihood Estimators of variance components using unbalanced data. *Biometrics* **26**, 505-524.

Shah K.R., Sinha B.K. (1989). *Theory of Optimal Designs.* Springer Verlag, Berlin.

Received 12 August 1996; revised 14 October 1997

Relacje pomiędzy pewnymi kryteriami optymalności

STRESZCZENIE

Praca przedstawia relacje pomiędzy trzema kryteriami optymalności układów doświadczalnych: Kiefera, Bondara i Chenga. Wykazano że kryterium Chenga zawiera się w kryterium Bondara oraz że kryterium Chenga zawiera się w kryterium Kiefera. Niestety, nie udało się znaleźć relacji pomiędzy kryteriami Kiefera i Bondara. Przedyskutowano możliwość stosowania powyższych kryteriów dla doświadczeń, w których obserwacje opisywane są modelem liniowym stałym lub mieszanym.

SLOWA KLUCZOWE: kryteria optymalności, macierz informacji, układy doświadczalne.