REVSTAT – Statistical Journal
Volume 6, Number 3, November 2008, 279–307

OPTIMAL AND QUASI-OPTIMAL DESIGNS *

Authors: João Paulo Martins
– Instituto Politécnico de Leiria, Esc. Sup. de Tecnologia e Gestão,
CEAUL, Centro de Estatística e Aplicações da Univ. de Lisboa, Portugal
jpmartins@estg.ipleiria.pt

Sandra Mendonça
– Universidade da Madeira, Dept. de Matemática e Engenharias,
CEAUL, Centro de Estatística e Aplicações da Univ. de Lisboa, Portugal
smfm@uma.pt

Dinis Duarte Pestana
– Universidade de Lisboa, Faculdade de Ciências, DEIO,
CEAUL, Centro de Estatística e Aplicações da Univ. de Lisboa, Portugal
dinis.pestana@fc.ul.pt

Received: July 2008 Revised: October 2008 Accepted: October 2008

Abstract:

- Optimal design theory deals with the choice of the allocation of the observations to accomplish the estimation of some linear combination of the coefficients in a regression model in an optimal way. Canonical moments provide an elegant framework to the theory of optimal designs. An optimal design for polynomial regression of a given degree r can be fatally inappropriate in case the polynomial degree should in fact be s, and hence when r is unknown it would be preferable to consider designs that show good performance for different values of the polynomial degree. Anderson's (1962) path-breaking solution of this multidecision problem has originated many developments, as optimal discriminant designs and optimal robust designs. But once again a design devised for a specific task can be grossly inefficient for a slightly different purpose. We introduce mixed designs; tables for regression of degrees $r=2,3,4$ exhibiting the loss of efficiency when the optimal mixed design is used instead of the optimal discriminant or of the optimal robust design show that the loss of efficiency is at most 1% and 2%, respectively, while the loss of efficiency when using a discriminant design instead of a robust design or vice-versa can be as high as 10%. Using recursive relations we compute pseudo-canonical moments for measures with infinite support, showing that such pseudo-canonical moments do not share the good identifiability properties of canonical moments of measures whose support is a subset of a compact interval of the real line.

Key-Words:

- *Optimal designs; discriminant designs; robust designs; mixed designs; quasi-optimal designs; canonical and pseudo-canonical moments.*

AMS Subject Classification:

- 62J02, 62K05.

*Invited lecture presented at the IWMS'08 — 17^{th} International Workshop on Matrices and Statistics, in honour of T.W. Anderson's 90^{th} birthday.

1. INTRODUCTION

Suppose the least squares method is used to estimate some linear combination of the coefficients in a regression model $Y(x) = \theta_0 + \theta_1 x + ... + \theta_r x^r + \varepsilon$ on an interval (a, b). The optimal design theory deals with the choice of the allocation of the observations to accomplish the estimation in an optimal way.

The problem has been solved by Smith (1918) using a global optimality criterion based on the variance of the estimated regression function, and *circa* 1960 Guest (1958), Hoel (1958), Box and Draper (1959, 1963), Kiefer (1959, 1961, 1962), Kiefer and Wolfowitz (1959) brought in many new results, namely by introducing sensible optimality criteria, and Anderson (1962) and Kussmaul (1969) investigated the choice of the degree in polynomial regression. See also Stigler (1971) and references therein for the discussion of alternative optimal criteria.

The *design space* $\mathcal{X}$ is the set of all possible points where measurements Y can be taken; $\mathcal{X}$ is assumed to be a compact subset of an Euclidean space. The measurements $Y = Y(x)$, the response at $x \in \mathcal{X}$, is the sum of the deterministic mean effect $\boldsymbol{f}(x)^T\theta = \mathbb{E}[Y|_x]$ and an additive error term ε. In other words,

$$\boldsymbol{Y} = \boldsymbol{f}(x)^T\boldsymbol{\theta} + \varepsilon$$

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)^T$ is a vector of unknown parameters, $\boldsymbol{f}(x) = \big(f_1(x), ..., f_k(x)\big)^T$ is a vector of real-valued linearly independent continuous regression functions, and ε is an error term with $\mathbb{E}(\varepsilon) = 0$.

For point estimation the *moment assumptions* $\mathbb{E}[\boldsymbol{Y}|_x] = \boldsymbol{f}(x)^T\boldsymbol{\theta}$ and $\mathrm{var}[\boldsymbol{Y}|_x] = \sigma^2 > 0$ provide an adequate setting, but for intervalar estimation or hypothesis testing the usual assumption is that $Y \frown$ *Gaussian*$\big(\boldsymbol{f}(x)^T\boldsymbol{\theta}, \sigma^2\big)$.

We further assume that the experimenter can take n uncorrelated observations at experimental conditions $x_1, ..., x_n \in \mathcal{X}$

$$Y_i = \boldsymbol{f}(x_i)\,\boldsymbol{\theta} + \varepsilon_i\,, \quad \mathbb{E}[\varepsilon_i\,\varepsilon_j] = \sigma^2\delta_{ij}\,, \quad \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad i, j = 1, ..., n$$

at not necessarily distinct points x_i.

Denoting the vectors of the responses $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$ and of the errors $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)^T$, we can rewrite the univariate regression model in matrix form

$$\boldsymbol{Y} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

where $X = \big(\boldsymbol{f}(x_1), ..., \boldsymbol{f}(x_n)\big)^T$ denotes the design matrix, $\mathbb{E}(\boldsymbol{Y}) = X\boldsymbol{\theta}$ and the dispersion matrix of the random vector $\boldsymbol{Y}$ is $D(\boldsymbol{Y}) = \sigma^2\boldsymbol{I}_n$.

The estimation of the unknown parameters $\boldsymbol{\theta}, \sigma^2$ from the observed responses $\boldsymbol{Y}$ is an important problem. We shall consider only linear unbiased estimators $\tilde{\boldsymbol{\theta}}_L = L\boldsymbol{Y}$ where L is a given $k \times n$ matrix and $\mathbb{E}[\tilde{\boldsymbol{\theta}}_L] = LX\boldsymbol{\theta} = \boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \mathbb{R}^k$.

In this general setting, the comparison of linear unbiased estimators is performed in terms of the Loewner ordering of the set of symmetric matrices

$$A \geq B \quad \text{iff} \quad A - B \text{ is nonnegative definite};$$
$$A > B \quad \text{iff} \quad A - B \text{ is positive definite}.$$

It is easily proven (Dette and Studden, 1997, p. 131) that the Gauss–Markov estimator $\tilde{\boldsymbol{\theta}}^{GM} = (X^T X)^{-1} X^T \boldsymbol{Y}$ is BLUE with respect to the Loewner ordering for the regression model with moment assumptions.

Often we are interested in inference about a particular linear combination $z_j\boldsymbol{\theta}$, $z_j \in \mathbb{R}^k$, $j = 1, ..., s$, of the unknown parameters. The *parameter subsystem* $K^T\boldsymbol{\theta} \in \mathbb{R}^s$, where $K = (z_1, ..., z_s)$ denotes a $k \times s$ matrix of rank $s \leq k$ is estimable if and only if there exists a linear unbiased estimator for $K^T\boldsymbol{\theta}$.

This is so if and only if the range inclusion $\text{range}(K) \subseteq \text{range}(X^T)$ is satisfied. In that case, the BLUE for the parameter subsystem $K^T\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}^T = K^T (X^T X)^- X^T \boldsymbol{Y}$$

with minimum dispersion matrix $D(\hat{\boldsymbol{\theta}}^T) = \sigma^2 K^T (X^T X)^- K$.

In the above expression, $(X^T X)^-$ denotes a generalized inverse of $(X^T X)$, i.e. $(X^T X)(X^T X)^- (X^T X) = X^T X$; under the range inclusion condition neither $\hat{\boldsymbol{\theta}}^T$ nor $D(\hat{\boldsymbol{\theta}}^T)$ depend on the specific choice of the generalized inverse.

Under the linear model with gaussian assumption, $K \in \mathbb{R}^{k \times s}$ a given matrix of rank $s \leq k$, if the range inclusion assumption is satisfied for a parameter subsystem $K^T\theta$ and if $n > \text{rank}(X)$, the null hypothesis H_0: $K^T\boldsymbol{\theta} = 0$ is rejected for large values of the test statistic

$$\frac{n - \text{rank}(X)}{\text{rank}(K)} \; \frac{\left(\hat{\boldsymbol{\theta}}_{(K)}\right)^T \left(K^T (X^T X)^- K\right)^- \hat{\boldsymbol{\theta}}_{(K)}}{\boldsymbol{Y}^T \left(I_n - X (X^T X)^- X^T\right) \boldsymbol{Y}}$$

where $\hat{\boldsymbol{\theta}}_{(K)} = K^T (X^T X)^- X^T \boldsymbol{Y}$.

With the gaussian assumption, under the null hypothesis the sampling distribution of the test statistic is a noncentral F with $(\text{rank}(K), n - \text{rank}(X))$ degrees of freedom and noncentrality parameter

$$\frac{1}{\sigma^2} \, (k^T\boldsymbol{\theta})^T \left(K^T (X^T X)^- K\right)^- (K^T\boldsymbol{\theta}) \; .$$

It is readily established that the power function of the F-test for the hypothesis H_0: $K^T\boldsymbol{\theta} = 0$ is an increasing function of the noncentrality parameter.

2. CANONICAL MOMENTS

Under the assumption of gaussian "errors" $\varepsilon \frown Gaussian(0,\sigma^2)$, or even of a less demanding *moments assumption* involving homocedasticity, the choice of the allocation of the observations to accomplish the estimation in an optimal way amounts to dealing with the minimization of some functionals of the covariance matrix, and an elegant solution is provided using the theory of canonical moments and of closely related parameters (Dette and Studden, 1997):

Let

$$m_k(\mu) \;=\; m_k := \int_a^b x^k \,\mathrm{d}\mu(x) \;, \qquad k=1,2,...$$

denote the k-th raw moment of the probability measure μ defined on the Borel sets of $[a,b]$, let

$$\boldsymbol{m}_n(\mu) \;=\; \boldsymbol{m}_n := (m_1,...,m_n)$$

denote the vector of raw moments up to order n, and $\mathcal{P}_{\boldsymbol{m}}$ the class of all probability measures defined on the Borel sets of $[a,b]$ whose moments up to the order n are $m_1,...,m_n$.

Skibinski (1967) investigated $m^+_{n+1} := \max\limits_{\mu\in\mathcal{P}_{\boldsymbol{m}}} \{m_{n+1}(\mu)\}$ and $m^-_{n+1} := \min\limits_{\mu\in\mathcal{P}_{\boldsymbol{m}}} \{m_{n+1}(\mu)\}$; from those "extreme" moments we can define several parameters, namely the *canonical moments*

$$\chi_k := \frac{m_k - m^-_{n+1}}{m^+_{n+1} - m^-_{n+1}} \;, \qquad k=1,2,... \;,$$

and the closely associated parameters

$$\zeta_0 := 1 \;, \qquad \zeta_1 := \chi_1 \;, \qquad \zeta_k := \xi_{k-1}\,\chi_k \;, \quad k\geq 2 \;,$$

and

$$\gamma_0 := 1 \;, \qquad \gamma_1 := \eta_1 \;, \qquad \gamma_k := \chi_{k-1}\,\xi_k \;, \quad k\geq 2 \;,$$

where $\xi_k := 1-\chi_k$; they have the substantial advantage of being invariant under linear transformations of the measure μ. From this invariance property, we shall in general consider $[a,b]=[-1,1]$, or, whenever more appropriate, $[a,b]=[0,1]$. Dette and Studden (1997, p. 21) claim that the parameters ζ_k and γ_k are more basic than the canonical moments.

The above parameters can be easily expressed in terms of the *Hankel determinants*

$$\underline{H}_{2n} := \begin{vmatrix} m_0 & \cdots & m_n \\ \vdots & \ddots & \vdots \\ m_n & \cdots & m_{2n} \end{vmatrix} \qquad \overline{H}_{2n} := \begin{vmatrix} m_1 - m_2 & \cdots & m_n - m_{n+1} \\ \vdots & \ddots & \vdots \\ m_n - m_{n+1} & \cdots & m_{2n-1} - m_{2n} \end{vmatrix}$$

and

$$\underline{H}_{2n+1} := \begin{vmatrix} m_1 & \cdots & m_{n+1} \\ \vdots & \ddots & \vdots \\ m_{n+1} & \cdots & m_{2n+1} \end{vmatrix} \qquad \overline{H}_{2n+1} := \begin{vmatrix} m_0 - m_1 & \cdots & m_n - m_{n+1} \\ \vdots & \ddots & \vdots \\ m_n - m_{n+1} & \cdots & m_{2n} - m_{2n+1} \end{vmatrix}$$

provided we define $\underline{H}_{-2} = \overline{H}_{-2} = \underline{H}_{-1} = \overline{H}_{-1} = \underline{H}_0 = \overline{H}_0 := 1$:

$$\chi_n = \frac{\underline{H}_n \, \overline{H}_{n-2}}{\underline{H}_{n-1} \, \overline{H}_{n-1}}\,, \quad \xi_n = \frac{\underline{H}_{n-2} \, \overline{H}_n}{\underline{H}_{n-1} \, \overline{H}_{n-1}}\,, \quad \zeta_n = \frac{\underline{H}_n \, \underline{H}_{n-3}}{\underline{H}_{n-1} \, \underline{H}_{n-2}}\,, \quad \gamma_n = \frac{\overline{H}_n \, \overline{H}_{n-3}}{\overline{H}_{n-1} \, \overline{H}_{n-2}}\,.$$

For instance, the canonical moments of $X \frown Beta(p,q)$, $p,q>0$, are $\chi_n = \left(\frac{1-(-1)^n}{2}\,p + \left[\frac{n}{2}\right]\right) / (p+q+n-1)$, $n=1,2,...$ (as usual, $[x]$ is the greatest integer less than or equal to x); observe, in particular, thal all the canonical moments of the $Beta\left(\frac{1}{2},\frac{1}{2}\right)$ (or arcsine) measure are $\chi_n = \frac{1}{2}$ (Skibinski, 1969).

It can be readily established that:

- The random variable with support $\boldsymbol{S} \subseteq [-1,1]$ corresponding to the sequence of canonical moments $\left(\frac{1}{2}, \chi_2, \frac{1}{2}, 1\right)$ is

$$X = \begin{cases} -1 & 0 & 1 \\ \dfrac{\chi_2}{2} & \xi_2 & \dfrac{\chi_2}{2} \end{cases}.$$

- The random variable with support $\boldsymbol{S} \subseteq [-1,1]$ corresponding to the sequence of canonical moments $\left(\frac{1}{2}, \chi_2, \frac{1}{2}, \chi_4, \frac{1}{2}, 1\right)$ is

$$X = \begin{cases} -1 & -\sqrt{\chi_2\,\xi_4} & \sqrt{\chi_2\,\xi_4} & 1 \\ \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} & \dfrac{1}{2} - \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} & \dfrac{1}{2} - \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} & \dfrac{\chi_2\,\chi_4}{2\,(1-\chi_2\,\xi_4)} \end{cases}.$$

- The random variable with support $\boldsymbol{S} \subseteq [-1,1]$ corresponding to the sequence of canonical moments $\left(\frac{1}{2}, \chi_2, \frac{1}{2}, \chi_4, \frac{1}{2}, \chi_6, \frac{1}{2}, 1\right)$ is

$$X = \begin{cases} -1 & -\sqrt{\chi_2\,\xi_4 + \chi_4\,\xi_6} & 0 & \sqrt{\chi_2\,\xi_4 + \chi_4\,\xi_6} & 1 \\ \alpha_1 & \alpha_2 & 1 - 2\alpha_1 - 2\alpha_2 & \alpha_2 & \alpha_1 \end{cases},$$

where $\alpha_1 = \dfrac{\chi_2\,\chi_4\,\chi_6}{2\,(\xi_2\,\xi_4 + \chi_4\,\chi_6)}$ and $\alpha_2 = \dfrac{\chi_2\,\xi_4\,\xi_6}{2\,(\chi_2\,\xi_4 + \chi_4\,\xi_6)\,(\xi_2\,\xi_4 + \chi_4\,\chi_6)}$.

For a thorough discussion on moment spaces, moment sequences, canonical moments and their connection with Stieltjes transforms, continued fractions and orthogonal polynomials, cf. Dette and Studden (1997).

3. EXACT, APPROXIMATE AND OPTIMAL DESIGNS

In what follows, we shall assume that the unknown regression functions are sufficiently smooth over the range under investigation, so that modeling with a low degree polynomial $P_r(x) = \sum_{k=0}^{r} \theta_k x^k$ is appropriate.

In other words, $\boldsymbol{f}(x) = (1, x, ..., x^r)^T$, $k = r+1$, and if the observations are taken at the points $x_1, ..., x_n$, the design matrix is

$$X = \begin{bmatrix} 1 & x_1 & \cdots & x_1^r \\ 1 & x_2 & \cdots & x_2^r \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^r \end{bmatrix} .$$

The design matrix X has rank $r+1$ if and only if there are at least $r+1$ different points among $x_1, ..., x_n$. We define the matrix of empirical moments up to order $2r$:

$$\frac{1}{n} X^T X = \begin{bmatrix} 1 & m_1 & m_2 & \cdots & m_r \\ m_1 & m_2 & m_3 & \cdots & m_{r+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_r & m_{r+1} & m_{r+2} & \cdots & m_{2r} \end{bmatrix}$$

with $m_i = \frac{1}{n} \sum_{k=0}^{n} x_k^i$, $i = 0, ..., 2r$.

The covariance matrix of the BLUE for the parameter subsystem $K^T \boldsymbol{\theta}$, where $K \in \mathbb{R}^{(r+1)\times s}$, is given by $\sigma^2 K^T (X^T X)^- K$.

If the experimenter is interested in finding out whether a polynomial regression of degree r or $r-1$ is appropriate for describing the response variable in terms of the explanatory variable, he can perform a F test as described above:

$$H_0 : \; K^T \boldsymbol{\theta} = \theta_r = 0$$

where $K = e_r = (0, 0, ..., 1)^T \in \mathbb{R}^{r+1}$ denotes the $(r+1)$-th unit vector. Assuming that the range inclusion is verified, in other words that there are at least $r+1$ different points among the $x_1, ..., x_n$, the test statistic

$$\frac{(n-r-1)\, \hat{\theta}^2_{r(e_r)} \left(e_r^T (X^T X)^{-1} e_r\right)^{-1}}{\boldsymbol{Y}^T \left(I_n - X\, (X^T X)^{-1} X^T\right) \boldsymbol{Y}} \;,$$

where $\hat{\theta}_{r(e_r)} = e_r^T \hat{\boldsymbol{\theta}}^{GM}$, has under the null hypothesis the F distribution with $(1, n-m-1)$ degrees of freedom and noncentrality parameter $\frac{1}{\sigma^2} \theta_r^2 \left(e_r^T (X^T X)^{-1} e_r\right)^{-1}$.

As we observed above, the power function of the F-test for the null hypothesis H_0: $\theta_r = 0$ increases when $e_r^T (X^T X)^{-1} e_r$ decreases with respect to the choice of observation points — and this clearly raises the question whether there exists an optimum experimental design.

To discuss this issue, let us consider the linear regression model with the moment assumptions $\mathbb{E}[\boldsymbol{Y}] = X\boldsymbol{\theta}$ and $D(\boldsymbol{Y}) = \sigma^2 \boldsymbol{I}_n$, where the design matrix is $X = \big(\boldsymbol{f}(x_1), ..., \boldsymbol{f}(x_n)\big)^T \in \mathbb{R}^{n\times k}$.

An *exact design* for sample size n is a finite probability measure on the design space $\mathcal{X}$ with support in the distinct points $x_1, ..., x_\ell$ among the $x_1, ..., x_n$, $\ell \le n$, with masses $\dfrac{n_i}{n}$, $i = 1, ..., \ell$, that are multiples of $\dfrac{1}{n}$; n_i, $i = 1, ..., \ell$, is the number of times the particular point x_i occurs among $x_1, ..., x_n$. An exact design $\partial_{(n)}$ can therefore be represented

$$\partial_{(n)} = \begin{cases} x_1 & \cdots & x_\ell \\[2ex] \dfrac{n_1}{n} & \cdots & \dfrac{n_\ell}{n} \end{cases}$$

(Kiefer, 1959), and the matrix $X^T X$ is

$$\begin{aligned} X^T X \;&=\; \sum_{k=1}^{n} \boldsymbol{f}(x_k)\, \boldsymbol{f}^T(x_k) \;=\; n \sum_{j=1}^{\ell} \frac{n_i}{n}\, \boldsymbol{f}(x_j)\, \boldsymbol{f}^T(x_j) \\ &=\; n \int \boldsymbol{f}(x)\, \boldsymbol{f}^T(x)\, \mathrm{d}\partial_{(n)}(x) \;=:\; n\, \mathcal{M}(\partial_{(n)}) \;. \end{aligned}$$

Let $K \in \mathbb{R}^{k\times s}$ be a given matrix of rank $s \le k$, and consider the problem of estimating the estimable parameter subsystem $K^T\boldsymbol{\theta}$; as the minimum dispersion matrix $D(\hat{\boldsymbol{\theta}}_{(K)}) = \frac{\sigma^2}{n} K^T M^{-1}(\partial_{(n)})\, K$ depends on the design $\partial_{(n)}$, it is reasonable to choose an optimum exact design, whenever feasible, i.e. an exact design that for some optimality criterion minimizes the dispersion matrix.

Integer optimization raises many problems, and an approximate solution can be satisfactory. Hence it may be much more convenient to use an *approximate design*, defined as a probability measure on the design space $\mathcal{X}$ with support points $x_1, ..., x_\ell$ and weights $w_1, ..., w_\ell$ adding up to 1:

$$\partial_{(n)} = \begin{cases} x_1 & \cdots & x_\ell \\ w_1 & \cdots & w_\ell \end{cases} .$$

The interpretation is obvious, and exact designs for finite sample sizes can be found by apportionment from the optimal approximate designs (Fedorov, 1972), with the huge advantage that we can use the duality theory of convex analysis in the optimization of a concave function on a convex and compact subset of the set of nonnegative definite $s\times s$ matrices $NND(s)$ instead of integer optimization.

Pukelsheim (1993) discusses in depth several different optimality criteria or information functions — real valued, positively homogeneous, nonconstant, upper semicontinuous, isotonic and concave functions on $NND(s)$ — for determining optimum designs maximizing appropriate functions of the information matrix

$$C_K\big(M(\partial_{(n)})\big) =: \big(K^T M^{-1}(\partial_{(n)})\, K\big)^{-1}.$$

A design ∂^* is G-optimal for the parameter $\boldsymbol{\theta}$ if $|\mathcal{M}(\partial^*)| > 0$ and it minimizes $G(\partial) = \max\limits_{x\in\mathcal{X}} \boldsymbol{f}^T(x)\, \mathcal{M}^{-1}(\partial)\, \boldsymbol{f}(x)$. G-optimal designs for low order polynomials have been first worked out numerically by Smith (1918), and theoretically by Guest (1958).

Hoel (1958) introduced D-optimal designs, the case $p=0$ of Kiefer's ϕ_p-criteria we shall focus on, based on the definition of the p-th matrix mean

$$\phi_p(C) = \begin{cases} \lambda_{\min}(C) & p=-\infty \\ \big(\det\, C\big)^{\frac{1}{s}} & p=0 \\ \left(\dfrac{1}{s}\ \text{trace}\ C^p\right)^{\frac{1}{p}} & p\in(-\infty,0)\cup(0,1) \end{cases}$$

for $C\in PD(s)$, the set of positive definite $s\times s$ matrices, and

$$\phi_p(C) = \begin{cases} 0 & p\in[-\infty,0] \\ \left(\dfrac{1}{s}\ \text{trace}\ C^p\right)^{\frac{1}{p}} & p\in(0,1] \end{cases}$$

for $C\in NND(s)$.

The popular D-optimality criterion uses $p=0$:

$$\phi_0\Big(C_K\big(M(\partial_{(n)})\big)\Big) = \Big(\det\big(K^T M^{-1}(\partial_{(n)})\, K\big)\Big)^{-\frac{1}{s}}.$$

A D-optimum design $\partial^D_{(n)}$ for $K^T\boldsymbol{\theta}$ minimizes the volume of the ellipsoids of concentration for the vector $K^T\boldsymbol{\theta}$ with respect to the choice of designs $\partial_{(n)}$. In particular, if $K=I_k$, the D-optimum design $\partial^D_{(n)}$ maximizes $\det(\mathcal{M}(\partial_{(n)}))$.

Guest (1958) G-optimal designs and Hoel (1958) D-optimal designs coincide, and in 1960 Kiefer and Wolfowitz established the earliest "equivalence theorem": A design ∂^* with $\mathcal{M}(\partial^*)>0$ is G-optimal for the parameter $\boldsymbol{\theta}$ if and only if it is D-optimal.

In what concerns the univariate polynomial regression model, Guest (1958) and Hoel (1958) results can be rephrased by noting that

$$|\mathcal{M}_r(\partial)| = \left|\int_0^1 \boldsymbol{f}_r(x)\, \boldsymbol{f}_r^T(x)\ \mathrm{d}\partial(x)\right| = \underline{H}_{2r} = \prod_{j=1}^{r} (\zeta_{2j-1}\,\zeta_{2j})^{r-j+1}$$

and therefore

> *The D-optimal design ∂_r^D for the full parameter $\boldsymbol{\theta}$ in the univariate polynomial regression model of degree r on the interval $[-1,1]$ has equal masses at the $r+1$ zeros of the polynomial $(x^2-1)\,L_r^{'}(x)$, where $L_r^{'}$ denotes the derivative of the r-th Legendre polynomial.*

A D-optimal design on the interval $[a,b]$ is obviously obtained from ∂_r^D by the linear transformation $\partial_{[a,b]}(\{x\}) = \partial\left(\left\{\dfrac{2x-b-a}{b-a}\right\}\right)$. Observe also that $(x^2-1)\,L_r^{'}(x) = r\,x\,L_r(x) - r\,L_{r-1}(x)$. Hence, for low degree polynomials, the optimal observation points are:

r										
2	-1	0	1							
3	-1	-0.44721	0.44721	1						
4	-1	-0.65465	0	0.65465	1					
5	-1	-0.76506	-0.28523	0.28523	0.76506	1				
6	-1	-0.83022	-0.46885	0	0.46885	0.83022	1			
7	-1	-0.8717	-0.59170	-0.20930	0.20930	0.59170	0.8717	1		
8	-1	-0.8998	-0.67719	-0.36312	0	0.36312	0.67719	0.8998	1	

The D-efficiency of a given design in the polynomial regression of degree r is

$$\mathrm{eff}_r^D(\partial) = \left(\frac{|\mathcal{M}_r(\partial)|}{|\mathcal{M}_r(\partial^D)|}\right)^{\frac{1}{r+1}} .$$

On the other hand, the information for the parameter $K^T\boldsymbol{\theta} = \theta_r$ is given by

$$C_{e_r}\big(\mathcal{M}(\partial)\big) \;=\; \big(e_r^T\,\mathcal{M}_r^{-1}(\partial)\;e_r\big)^{-1} \;=\; \frac{|\mathcal{M}_r(\partial)|}{\big|\mathcal{M}_{r-1}(\partial)\big|} .$$

A design maximizing $C_{e_r}\big(\mathcal{M}(\partial)\big)$ is called D_1-optimal in the sense that it is optimal for the estimation of the highest coefficient θ_r:

> *The D_1-optimal design $\partial_r^{D_1}$ in the univariate polynomial regression of degree r on the interval $[-1,1]$ has equal masses $\dfrac{1}{2r}$ at the points -1 and 1, and equal masses $\dfrac{1}{r}$ at the zeros of the Chebyshev polynomial of second kind $U_{r-1}(x)$.*

An example: In order to investigate if the quadratic term is relevant in the univariate quadratic model $\mathbf{Y} = \theta_0 + \theta_1 x + \theta_2 x^2 + \boldsymbol{\varepsilon}$ on the design space $\mathcal{X} = [-1,1]$, we consider $K = \boldsymbol{e}_2 = (0,0,1)^T$.

Denoting $\partial_{(n)}$ an exact design of sample size n, and $\boldsymbol{f}(x) = (1, x, x^2)^T$ the vector of regression functions, the matrix $\boldsymbol{M}(\partial_{(n)})$ is

$$\boldsymbol{M}(\partial_{(n)}) = \int_{-1}^{1} \boldsymbol{f}(x)\, \boldsymbol{f}(x)^T \, \mathrm{d}\partial_{(n)}(x) = \begin{bmatrix} 1 & m_1 & m_2 \\ m_1 & m_2 & m_3 \\ m_2 & m_3 & m_4 \end{bmatrix} .$$

The parameter $\theta_2 = \boldsymbol{e}_2^T \boldsymbol{\theta}$ is estimable if and only if $\xi_{(n)}$ has at least three support points, and for these designs the dispersion of the Gauss–Markov estimator is proportional to

$$\Big\{C_K\big(\boldsymbol{M}(\partial_{(n)})\big)\Big\}^{-1} = \boldsymbol{e}_2^T \Big\{\boldsymbol{M}(\partial_{(n)})\Big\}^{-1} \boldsymbol{e}_2 = \frac{m_2 - m_1^2}{\big|\boldsymbol{M}(\partial_{(n)})\big|} .$$

The optimal designs, maximizing $C_K\big(\boldsymbol{M}(\partial_{(n)})\big)$ — and therefore minimizing the variance of the Gauss–Markov estimator of the parameter of interest θ_2 — in the set of all exact designs with nonsingular matrix $\boldsymbol{M}(\partial_{(n)})$ are

$$\partial^*_{(n)} = \begin{cases} \begin{cases} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{cases} & \text{if} \quad n = 4p \\[2ex] \begin{cases} -1 & 0 & 1 \\ \frac{p}{4p+1} & \frac{2p+1}{4p+1} & \frac{p}{4p+1} \end{cases} & \text{if} \quad n = 4p+1 \\[2ex] \begin{cases} -1 & \pm\, x_0(n) & 1 \\ \frac{p+1}{4p+2} & \frac{2p+1}{4p+2} & \frac{p}{4p+2} \end{cases} & \text{if} \quad n = 4p+2 \\[2ex] \begin{cases} -1 & 0 & 1 \\ \frac{p+1}{4p+3} & \frac{2p+1}{4p+3} & \frac{p+1}{4p+3} \end{cases} & \text{if} \quad n = 4p+3 \end{cases} ,$$

where in the case $n = 2p+2$ the point $x_0(n)$ is the real root of the cubic polynomial $n^2 x^3 - 3nx^2 + (n^2 - 2)x - n$ (Kraft and Schaefer, 1995).

On the other hand, an optimal approximate design to estimate θ_2 maximizes

$$C_{e_2}\big(\mathcal{M}(\partial)\big) = \frac{|\mathcal{M}(\partial)|}{m_2 - m_1^2} = \frac{\underline{H}_4(\partial)}{\underline{H}_2(\partial)} .$$

This can be reexpressed in terms of the canonical moments of the measure ∂:

$$C_{e_2}\big(\mathcal{M}(\partial)\big) = 2^4 \prod_{k=1}^{2} \gamma_{2k} = 2^4 \chi_4 \prod_{j=1}^{3} \chi_j \, \xi_j .$$

The maximization in terms of canonical moments yelds $\chi_1 = \chi_2 = \chi_3 = \frac{1}{2}$ and $\chi_4 = 1$, and the approximate optimal design for estimating θ_2 is

$$\partial^* = \begin{cases} -1 & 0 & 1 \\ \dfrac{1}{4} & \dfrac{1}{2} & \dfrac{1}{4} \end{cases} .$$

Hence, n_0 denoting the closest integer to $\frac{n}{4}$, ∂^* approximates the exact design $\tilde{\partial}_{(n)} = \begin{cases} -1 & 0 & 1 \\ \dfrac{n_0}{n} & 1-\dfrac{2n_0}{n} & \dfrac{n_0}{n} \end{cases}$.

In fact, they coincide unless $n = 4p+2$, and in this case comparing the performance of the two designs using the relative efficiency ratio $\dfrac{C_{e_2}\big(\mathcal{M}(\tilde{\partial}_{(4p+2)})\big)}{C_{e_2}\big(\mathcal{M}(\partial^*_{(4p+2)})\big)}$ we can observe that for $p \geq 5$ we get $\dfrac{C_{e_2}\big(\mathcal{M}(\tilde{\partial}_{(4p+2)})\big)}{C_{e_2}\big(\mathcal{M}(\partial^*_{(4p+2)})\big)} \geq 0.995$, as seen on Table 1.

Table 1: Relative efficiency of the approximate design.

p	1	2	3	4	5
n	6	10	14	18	22
$x_0(p)$	0.0707	0.0408	0.0289	0.0224	0.0183
relative efficiency	0.9327	0.9759	0.9877	0.9925	0.9950

4. DISCRIMINANT, ROBUST AND MIXED DESIGNS

Consider the model $Y = \sum_{k=0}^{r} \theta_{rk}\, x^k + \varepsilon$, under the gaussian assumption. The optimal design to fit a linear regression model is fatally inefficient to detect curvature, and in general an optimal design for a specific task can be inappropriate for slightly different purposes. Hence we recommend that the analysis be performed in two steps, first to try to identify the appropriate degree of the polynomial, then to build up the optimal design.

The two steps can however be merged if practical considerations on data gathering costs imply that should be so.

Anderson (1962) invented a good decision rule for this problem: For a given nondecreasing sequence of levels $(\alpha_1, ..., \alpha_r)$ the procedure he devised chooses the largest integer in $\{1, ..., r\}$ for which the F-test rejects the null hypothesis H_0: $\theta_{jj} = 0$ at the levels α_j. This method has several optimality properties, and led to the introduction of discriminant and of robust designs, discussed in what follows.

Let $\mathcal{F}_r$ be the class of all possible polynomial regression models up to degree r, and $\boldsymbol{\pi} = (\pi_1, ..., \pi_r)$ nonnegative numbers with $\pi_r > 0$ and such that $\pi_1 + \cdots + \pi_r = 1$. Those are interpreted as "priors" reflecting the experimenter belief about the adequacy of the polynomial regression of degree ℓ, $\ell = 1, ..., r$.

As discussed beforehand, H_0: $\theta_{\ell\ell} = 0$ can be tested using a test statistic with non-central F distribution, and the power function increases with the non centrality parameter which we now rewrite $\delta_\ell^2(\partial) = \dfrac{\theta_{\ell\ell}^2}{\sigma^2}\left(e_\ell^T\,(X_l^T X_\ell)^{-1}\,e_\ell\right)^{-1}$. As this should ideally be maximized for $\ell = 1, ..., r$, which would amount to jointly maximizing

$$\mathrm{eff}_\ell^{D_1}(\partial) \;=\; \frac{\delta_\ell^2(\partial)}{\sup\limits_\eta \delta_\ell^2(\eta)} \;=\; 2^{2\ell-2}\,\frac{|M_\ell\,(\partial)|}{|M_{\ell-1}\,(\partial)|}\ ,$$

a task obviously beyond what is feasible, what can be done in practice is to maximize an appropriate weighted mean of the above efficiencies, using the weights in $\boldsymbol{\pi}$ corresponding to the credibility the experimenter puts in the adequacy of using polynomal regression of each of the degrees ℓ, $\ell = 1, ..., r$.

*A design $\partial_{0,\boldsymbol{\pi}}$ with moment matrix $\mathcal{M}(\partial_{0,\boldsymbol{\pi}})$ is a Ψ_0-**optimal discriminating design** for the class $\mathcal{F}_r$ with respect to the prior $\boldsymbol{\pi}$ if and only if $\partial_{0,\boldsymbol{\pi}}$ maximizes the weighted geometric mean*

$$\Psi_0^{\boldsymbol{\pi}}(\partial) \;=\; \prod_{k=1}^{r}\left(\mathrm{eff}_k^{D_1}(\partial)\right)^{\boldsymbol{\pi}_k} \;=\; \prod_{k=1}^{r}\left(\frac{2^{4k-2}}{(b-a)^{2k}}\,\frac{|\mathcal{M}_k(\partial)|}{|\mathcal{M}_{k-1}(\partial)|}\right)^{\boldsymbol{\pi}_k}.$$

(Observe that if $\boldsymbol{\pi} = (0, ..., 0, 1)$ we obtain the D_1 optimality criterion.)

It is readily established that the Ψ_0-optimal discriminating design for the class $\mathcal{F}_r$ with respect to the prior $\boldsymbol{\pi} = (\pi_1, ..., \pi_r)$ is uniquely determined by its canonical moments

$$\chi_{2i-1} = \frac{1}{2},\quad i = 1, ..., r\ , \qquad \chi_{2i} = \frac{\Pi_i}{\Pi_i + \Pi_{i+1}},\quad i = 1, ..., r-1\ , \qquad \chi_{2r} = 1\ ,$$

where $\Pi_i = \sum\limits_{\ell=i}^{r} \pi_\ell$ (Lau and Studden, 1985). For instance, with the uniform prior $\boldsymbol{\pi} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ for the class $\mathcal{F}_4$ we have

$$\Pi_1 = 1,\quad \Pi_2 = \frac{3}{4},\quad \Pi_3 = \frac{1}{2},\quad \Pi_4 = \frac{4}{4}\ ,$$

and

$$\chi_2 = \frac{4}{7},\quad \chi_4 = \frac{3}{5},\quad \chi_6 = \frac{2}{3}\ .$$

Therefore the the Ψ_0-optimal discriminating design is

$$\partial_{0,\boldsymbol{\pi}_U} = \begin{cases} -1 & -\sqrt{\dfrac{3}{7}} & 0 & \sqrt{\dfrac{3}{7}} & 1 \\[2ex] \dfrac{1}{5} & \dfrac{1}{5} & \dfrac{1}{5} & \dfrac{1}{5} & \dfrac{1}{5} \end{cases}\ .$$

In what concerns $\Psi_0^{\boldsymbol{\pi}}$-optimal discriminant designs for the classes $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$, and with $\boldsymbol{\pi}$ giving the same prior probability $1/r$ to the values of ℓ ranging from 1 to r,

r	$\boldsymbol{\pi}$	Points	$\mathrm{eff}_1^{D_1}$ / $\mathrm{eff}_2^{D_1}$ / $\mathrm{eff}_3^{D_1}$ / $\mathrm{eff}_4^{D_1}$ (ξ)
2	$\left(\frac{1}{2},\frac{1}{2}\right)$	-1, 0, 1	0.817 / 1 / − / −
3	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$	-1, -0.4472, 0, 4472, 1	0.600 / 0.640 / 0.853 / −
4	$\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)$	-1, -0.6547, 0, 0.6547, 1	0.571 / 0.588 / 0.627 / 0.836

we can observe, when comparing with the efficiency of the D_1-optimal design for polynomial regression of degree $r=4$, that the loss of efficiency in the case of degree 4 is largely compensated by the increased efficiency when the appropriate degree is lower than 4.

An alternative strategy, inspired on the way $\Psi_0^{\boldsymbol{\pi}}$-optimal discriminant designs have been defined, is to build up designs maximizing an weighted geometric mean of D-efficiencies, up to some degree r. Those designs are christened robust designs since they are quite efficient for a set of possible polynomial regression degrees.

*For a given weights vector $\boldsymbol{\pi}=(\pi_1,...,\pi_r)$, $\sum_{l=1}^r \pi_l=1$ and $\pi_i>0$, the design $\partial_{0,\boldsymbol{\pi}}$ is a $\Xi_0^{\boldsymbol{\pi}}$**-robust design** for the class $\mathcal{F}_r$ in respect to the prior $\boldsymbol{\pi}$ if and only if $\partial_{0,\boldsymbol{\pi}}$ maximizes the weighted geometric mean*

$$\Xi_0^{\boldsymbol{\pi}}(\partial) = \prod_{\ell=1}^{r} \left(\mathrm{eff}_\ell^D(\partial)\right)^{\pi_\ell} = \prod_{\ell=1}^{r} \left(\frac{|M_\ell(\partial)|}{|M_\ell(\partial^D)|}\right)^{\pi_\ell/(\ell+1)} .$$

Dette and Studden (1995) show that the canonical moments for the above defined robust design are

$$\chi_{2i-1}=\frac{1}{2},\quad i=1,...,r\ , \qquad \chi_{2i}=\frac{\sigma_i}{\sigma_i+\sigma_{i+1}},\quad i=1,...,r-1\ , \qquad \chi_{2r}=1\ ,$$

with $\sigma_i=\displaystyle\sum_{\ell=i}^{r}\frac{\ell+1-i}{\ell+1}\,\pi_\ell$.

For $\Xi_0^{\boldsymbol{\pi}}$-robust designs for the classes $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$, and with $\boldsymbol{\pi}$ giving the same prior probability $1/r$ to the values of ℓ ranging from 1 to r,

m	Points	Weights	eff_1^D/eff_2^D/eff_3^D/eff_4^D (ξ)
2	$-1, 0, 1$	$0.389, 0.222, 0, 389$	0.881/0.968/ − /−
3	$-1, -0.401, 0.401, 1$	$0.319, 0.181, 0.181, 0.319$	0.835/0.914/0.954/−
4	$-1, -0.605, 0, -1, 0.605, 1$	$0.271, 0.152, 0.153, 0.152, 0.271$	0.809/0.883/0.927/0.949

As we shall show in Tables 5–12 below, gross loss of efficiency can be incurred into — up to 10% — when a $\Xi_0^{\boldsymbol{\pi}}$-robust design is used instead of a $\Psi_0^{\boldsymbol{\pi}}$-discriminant design, or vice-versa. This prompted us to use a mixed strategy, defining $\Theta_0^{\boldsymbol{\pi}}$-mixed designs as follows:

For a given weights vector $\boldsymbol{\pi} = (\pi_1, ..., \pi_r)$, $\sum_{l=1}^{r} \pi_l = 1$ *and* $\pi_i > 0$, *the design* $\partial_{0,\boldsymbol{\pi}}$ *is a* $\Theta_0^{\boldsymbol{\pi}}$***-mixed design*** *for the class* $\mathcal{F}_r$ *in respect to the prior* $\boldsymbol{\pi}$ *if and only if* $\partial_{0,\boldsymbol{\pi}}$ *maximizes the weighted geometric mean*

$$\Theta_0^{\boldsymbol{\pi}} = \prod_{\ell=1}^{r} \left(\mathrm{eff}_\ell^{D_1}(\partial)\right)^{\pi_\ell} \prod_{j=1}^{r} \left(\mathrm{eff}_j^{D}(\partial)\right)^{\pi_j}$$
$$= \prod_{\ell=1}^{r} \left(2^{2\ell-2} \frac{|M_\ell(\partial)|}{|M_{\ell-1}(\partial)|}\right)^{\pi_\ell} \prod_{j=1}^{r} \left(\frac{M_j(\partial)}{M_j\left(\partial_j^D\right)}\right)^{\frac{\pi_j}{j+1}}.$$

In Tables 2–4 we present mixed designs for $\mathcal{F}_r$, $r = 2, 3, 4$, and in Tables 5–12 we study the corresponding efficiencies when they are used instead of the corresponding optimal discriminant or robust designs.

Table 2: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r = 2$, $\boldsymbol{\pi} = (a, 1-a)$.

a	weight at ± 1	weight at 0
0.05	0.2835	0.4330
0.10	0.2895	0.4211
0.15	0.2958	0.4084
0.20	0.3025	0.3951
0.25	0.3095	0.3810
0.30	0.3170	0.3660
0.35	0.3249	0.3502
0.40	0.3333	0.3333
0.45	0.3423	0.3154
0.50	0.3519	0.2963
0.55	0.3621	0.2759
0.60	0.3730	0.2540
0.65	0.3848	0.2305
0.70	0.3974	0.2051
0.75	0.4111	0.1778
0.80	0.4259	0.1481
0.85	0.4420	0.1159
0.90	0.4596	0.0808
0.95	0.4788	0.0423

Table 3: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r=3$, $\boldsymbol{\pi}=(a,b,1-a-b)$.

a	b	t	weight at ± 1	weight at $\pm t$
0.1	0.1	0.4911	0.2119	0.2881
0.1	0.2	0.4748	0.2190	0.2810
0.1	0.3	0.4553	0.2267	0.2733
0.1	0.4	0.4315	0.2350	0.2650
0.1	0.5	0.4019	0.2440	0.2560
0.1	0.6	0.3635	0.2538	0.2462
0.1	0.7	0.3112	0.2646	0.2354
0.1	0.8	0.2318	0.2764	0.2236
0.2	0.1	0.5001	0.2256	0.2744
0.2	0.2	0.4808	0.2338	0.2662
0.2	0.3	0.4569	0.2428	0.2572
0.2	0.4	0.4269	0.2525	0.2475
0.2	0.5	0.3876	0.2632	0.2368
0.2	0.6	0.3333	0.2750	0.2250
0.2	0.7	0.2496	0.2880	0.2120
0.3	0.1	0.5095	0.2415	0.2585
0.3	0.2	0.4858	0.2513	0.2487
0.3	0.3	0.4556	0.2619	0.2381
0.3	0.4	0.4155	0.2736	0.2264
0.3	0.5	0.3593	0.2865	0.2135
0.3	0.6	0.2709	0.3009	0.1991
0.4	0.1	0.5190	0.2606	0.2394
0.4	0.2	0.4889	0.2722	0.2278
0.4	0.3	0.4483	0.2851	0.2149
0.4	0.4	0.3902	0.2994	0.2006
0.4	0.5	0.2967	0.3154	0.1846
0.5	0.1	0.5281	0.2836	0.2164
0.5	0.2	0.4875	0.2978	0.2022
0.5	0.3	0.4279	0.3137	0.1863
0.5	0.4	0.3289	0.3316	0.1684
0.6	0.1	0.5351	0.3121	0.1879
0.6	0.2	0.4748	0.3299	0.1701
0.6	0.3	0.3705	0.3500	0.1500
0.3	0.4	0.5353	0.3481	0.1519
0.3	0.5	0.4267	0.3710	0.1290
0.3	0.6	0.5085	0.3953	0.1047
1/3	1/3	0.4407	0.2731	0.2269

Table 4: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r=4$, $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

a	b	c	t	weight at ± 1	weight at $\pm t$	weight at 0
0.1	0.1	0.1	0.6973	0.1717	0.2177	0.2210
0.1	0.1	0.2	0.6836	0.1764	0.2176	0.2119
0.1	0.1	0.3	0.6673	0.1814	0.2184	0.2004
0.1	0.1	0.4	0.6474	0.1867	0.2206	0.1853
0.1	0.1	0.5	0.6228	0.1924	0.2252	0.1647
0.1	0.1	0.6	0.5913	0.1985	0.2341	0.1349
0.1	0.1	0.7	0.5495	0.2050	0.2513	0.0874
0.1	0.2	0.1	0.6937	0.1811	0.2036	0.2307
0.1	0.2	0.2	0.6765	0.1864	0.2036	0.2200
0.1	0.2	0.3	0.6553	0.1920	0.2051	0.2057
0.1	0.2	0.4	0.6284	0.1981	0.2091	0.1856
0.1	0.2	0.5	0.5933	0.2046	0.2178	0.1553
0.1	0.2	0.6	0.5453	0.2115	0.2363	0.1043
0.1	0.3	0.1	0.6886	0.1917	0.1873	0.2420
0.1	0.3	0.2	0.6661	0.1977	0.1879	0.2289
0.1	0.3	0.3	0.6370	0.2042	0.1908	0.2100
0.1	0.3	0.4	0.5979	0.2111	0.1988	0.1802
0.1	0.3	0.5	0.5423	0.2186	0.2182	0.1265
0.1	0.4	0.1	0.6811	0.2038	0.1686	0.2552
0.1	0.4	0.2	0.6499	0.2107	0.1701	0.2383
0.1	0.4	0.3	0.6065	0.2182	0.1766	0.2104
0.1	0.4	0.4	0.5417	0.2262	0.1959	0.1557
0.1	0.5	0.1	0.6692	0.2177	0.1469	0.2707
0.1	0.5	0.2	0.6217	0.2258	0.1509	0.2466
0.1	0.5	0.3	0.5462	0.2345	0.1683	0.1943
0.1	0.6	0.1	0.6481	0.2340	0.1216	0.2887
0.1	0.6	0.2	0.5610	0.2435	0.1342	0.2447
0.1	0.7	0.1	0.5981	0.2533	0.0935	0.3064
0.2	0.1	0.1	0.7031	0.1851	0.2128	0.2041
0.2	0.1	0.2	0.6869	0.1907	0.2124	0.1937
0.2	0.1	0.3	0.6670	0.1967	0.2132	0.1801
0.2	0.1	0.4	0.6419	0.2032	0.2162	0.1613
0.2	0.1	0.5	0.6092	0.2101	0.2231	0.1335
0.2	0.1	0.6	0.5647	0.2175	0.2384	0.0881
0.2	0.2	0.1	0.6978	0.1964	0.1965	0.2143
0.2	0.2	0.2	0.6766	0.2028	0.1964	0.2017
0.2	0.2	0.3	0.6492	0.2097	0.1985	0.1837
0.2	0.2	0.4	0.6125	0.2171	0.2049	0.1559
0.2	0.2	0.5	0.5607	0.2251	0.2212	0.1074
0.2	0.3	0.1	0.6900	0.2093	0.1775	0.2265
0.2	0.3	0.2	0.6604	0.2167	0.1783	0.2101
0.2	0.3	0.3	0.6195	0.2247	0.1835	0.1836

(continued on next page)

Table 4: $\Theta_0^{\boldsymbol{\pi}}$-optimal mixed design, $r=4$, $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

(continued from previous page)

a	b	c	t	weight at ± 1	weight at $\pm t$	weight at 0
0.2	0.3	0.4	0.5586	0.2333	0.2001	0.1332
0.2	0.4	0.1	0.6778	0.2242	0.1552	0.2411
0.2	0.4	0.2	0.6325	0.2329	0.1584	0.2175
0.2	0.4	0.3	0.5610	0.2423	0.1737	0.1681
0.2	0.5	0.1	0.6562	0.2418	0.1291	0.2582
0.2	0.5	0.2	0.5728	0.2520	0.1404	0.2151
0.2	0.6	0.1	0.6058	0.2627	0.0997	0.2753
0.3	0.1	0.1	0.7089	0.2014	0.2058	0.1857
0.3	0.1	0.2	0.6891	0.2082	0.2049	0.1737
0.3	0.1	0.3	0.6637	0.2156	0.2060	0.1569
0.3	0.1	0.4	0.6297	0.2236	0.2107	0.1316
0.3	0.1	0.5	0.5820	0.2322	0.2234	0.0888
0.3	0.2	0.1	0.7009	0.2152	0.1865	0.1966
0.3	0.2	0.2	0.6732	0.2231	0.1864	0.1809
0.3	0.2	0.3	0.6349	0.2317	0.1902	0.1562
0.3	0.2	0.4	0.5785	0.2411	0.2035	0.1109
0.3	0.3	0.1	0.6884	0.2313	0.1639	0.2098
0.3	0.3	0.2	0.6457	0.2406	0.1658	0.1872
0.3	0.3	0.3	0.5787	0.2508	0.1784	0.1416
0.3	0.4	0.1	0.6663	0.2503	0.1369	0.2256
0.3	0.4	0.2	0.5872	0.2614	0.1464	0.1844
0.3	0.5	0.1	0.6155	0.2731	0.1061	0.2417
0.4	0.1	0.1	0.7144	0.2216	0.1957	0.1656
0.4	0.1	0.2	0.6888	0.2301	0.1944	0.1510
0.4	0.1	0.3	0.6536	0.2394	0.1963	0.1287
0.4	0.1	0.4	0.6022	0.2495	0.2059	0.0892
0.4	0.2	0.1	0.7016	0.2389	0.1726	0.1771
0.4	0.2	0.2	0.6619	0.2490	0.1730	0.1561
0.4	0.2	0.3	0.6001	0.2601	0.1823	0.1152
0.4	0.3	0.1	0.6791	0.2595	0.1449	0.1912
0.4	0.3	0.2	0.6051	0.2717	0.1520	0.1527
0.4	0.4	0.1	0.6281	0.2845	0.1126	0.2057
0.5	0.1	0.1	0.7182	0.2472	0.1811	0.1434
0.5	0.1	0.2	0.6821	0.2582	0.1796	0.1245
0.5	0.1	0.3	0.6263	0.2703	0.1850	0.0894
0.5	0.2	0.1	0.6956	0.2697	0.1527	0.1552
0.5	0.2	0.2	0.6276	0.2830	0.1566	0.1207
0.5	0.3	0.1	0.6445	0.2972	0.1190	0.1676
0.6	0.1	0.1	0.7172	0.2810	0.1598	0.1184
0.6	0.1	0.2	0.6566	0.2957	0.1597	0.0892
0.6	0.2	0.1	0.6667	0.3114	0.1246	0.1280
0.7	0.1	0.1	0.6978	0.3274	0.1285	0.0882
0.25	0.25	0.25	0.6484	0.2239	0.1839	0.1845

Table 5: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=2$ and $\boldsymbol{\pi}=(a,1-a)$.

a	$\Psi_0^{\boldsymbol{\pi}}(D)$	$\Psi_0^{\boldsymbol{\pi}}(R)$	$\Psi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(R)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(M)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$
0.05	0.967	0.866	0.955	−10.076	−1.110
0.10	0.935	0.844	0.926	−9.118	−0.995
0.15	0.907	0.824	0.898	−8.233	−0.890
0.20	0.880	0.806	0.872	−7.414	−0.795
0.25	0.856	0.790	0.849	−6.657	−0.707
0.30	0.834	0.775	0.828	−5.958	−0.627
0.35	0.814	0.761	0.809	−5.312	−0.554
0.40	0.797	0.750	0.792	−4.714	−0.487
0.45	0.782	0.741	0.778	−4.162	−0.426
0.50	0.770	0.733	0.766	−3.650	−0.370
0.55	0.760	0.728	0.757	−3.177	−0.320
0.60	0.753	0.726	0.751	−2.739	−0.273
0.65	0.750	0.727	0.748	−2.331	−0.230
0.70	0.751	0.731	0.749	−1.951	−0.191
0.75	0.757	0.741	0.755	−1.596	−0.155
0.80	0.768	0.756	0.767	−1.262	−0.121
0.85	0.789	0.779	0.788	−0.943	−0.090
0.90	0.822	0.815	0.821	−0.635	−0.060
0.95	0.877	0.873	0.876	−0.328	−0.031

Table 6: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=2$ and $\boldsymbol{\pi}=(a,1-a)$.

a	$\Xi_0^{\boldsymbol{\pi}}(D)$	$\Xi_0^{\boldsymbol{\pi}}(R)$	$\Xi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(D)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(M)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$
0.05	0.939	0.990	0.967	−5.084	−2.343
0.10	0.934	0.981	0.959	−4.679	−2.161
0.15	0.929	0.972	0.952	−4.293	−1.986
0.20	0.924	0.963	0.945	−3.923	−1.819
0.25	0.919	0.955	0.938	−3.570	−1.658
0.30	0.915	0.947	0.932	−3.233	−1.505
0.35	0.911	0.941	0.927	−2.912	−1.359
0.40	0.908	0.934	0.922	−2.607	−1.219
0.45	0.906	0.929	0.918	−2.317	−1.086
0.50	0.904	0.924	0.914	−2.042	−0.959
0.55	0.902	0.920	0.912	−1.781	−0.839
0.60	0.902	0.917	0.910	−1.534	−0.724
0.65	0.903	0.916	0.910	−1.301	−0.616
0.70	0.905	0.916	0.911	−1.081	−0.513
0.75	0.909	0.917	0.913	−0.873	−0.416
0.80	0.914	0.921	0.918	−0.678	−0.324
0.85	0.923	0.928	0.926	−0.495	−0.237
0.90	0.937	0.940	0.938	−0.321	−0.154
0.95	0.957	0.959	0.958	−0.158	−0.076

Table 7: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=3$ and $\boldsymbol{\pi}=(a,b,1-a-b)$.

a	b	$\Psi_0^{\boldsymbol{\pi}}(D)$	$\Psi_0^{\boldsymbol{\pi}}(R)$	$\Psi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(R)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(M)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$
0.1	0.1	0.875	0.772	0.866	−10.305	−0.976
0.1	0.2	0.825	0.736	0.816	−8.885	−0.862
0.1	0.3	0.783	0.705	0.775	−7.788	−0.778
0.1	0.4	0.751	0.681	0.744	−6.989	−0.721
0.1	0.5	0.730	0.665	0.723	−6.478	−0.688
0.1	0.6	0.722	0.659	0.715	−6.266	−0.683
0.1	0.7	0.733	0.669	0.726	−6.405	−0.708
0.1	0.8	0.776	0.706	0.769	−7.058	−0.781
0.2	0.1	0.824	0.741	0.817	−8.285	−0.758
0.2	0.2	0.777	0.707	0.771	−7.049	−0.665
0.2	0.3	0.740	0.679	0.734	−6.133	−0.600
0.2	0.4	0.714	0.659	0.708	−5.518	−0.561
0.2	0.5	0.700	0.648	0.695	−5.206	−0.548
0.2	0.6	0.704	0.652	0.699	−5.232	−0.564
0.2	0.7	0.739	0.681	0.733	−5.724	−0.621
0.3	0.1	0.781	0.716	0.776	−6.556	−0.580
0.3	0.2	0.738	0.684	0.733	−5.485	−0.504
0.3	0.3	0.706	0.659	0.702	−4.741	−0.456
0.3	0.4	0.686	0.643	0.682	−4.314	−0.435
0.3	0.5	0.683	0.641	0.679	−4.227	−0.442
0.3	0.6	0.709	0.663	0.704	−4.577	−0.487
0.4	0.1	0.747	0.696	0.743	−5.077	−0.435
0.4	0.2	0.708	0.666	0.704	−4.157	−0.373
0.4	0.3	0.681	0.645	0.678	−3.587	−0.342
0.4	0.4	0.671	0.637	0.667	−3.374	−0.340
0.4	0.5	0.687	0.651	0.683	−3.591	−0.373
0.5	0.1	0.722	0.684	0.719	−3.809	−0.316
0.5	0.2	0.687	0.657	0.685	−3.038	−0.269
0.5	0.3	0.668	0.642	0.666	−2.667	−0.255
0.5	0.4	0.675	0.647	0.672	−2.745	−0.278
0.6	0.1	0.708	0.681	0.706	−2.718	−0.219
0.6	0.2	0.679	0.657	0.677	−2.116	−0.188
0.6	0.3	0.674	0.654	0.672	−2.024	−0.198
0.7	0.1	0.708	0.690	0.706	−1.775	−0.141
0.7	0.2	0.689	0.675	0.688	−1.424	−0.131
0.8	0.1	0.729	0.719	0.728	−0.982	−0.080
1/3	1/3	0.689	0.648	0.685	−4.174	−0.407

Table 8: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=3$ and $\boldsymbol{\pi}=(a,b,1-a-b)$.

a	b	$\Xi_0^{\pi}(D)$	$\Xi_0^{\pi}(R)$	$\Xi_0^{\pi}(M)$	$100\times$ $[\Xi_0^{\pi}(D)-\Xi_0^{\pi}(R)]$	$100\times$ $[\Xi_0^{\pi}(M)-\Xi_0^{\pi}(R)]$
0.1	0.1	0.913	0.962	0.938	−4.996	−2.433
0.1	0.2	0.905	0.951	0.929	−4.597	−2.208
0.1	0.3	0.898	0.941	0.920	−4.288	−2.029
0.1	0.4	0.892	0.932	0.914	−4.070	−1.898
0.1	0.5	0.887	0.927	0.909	−3.945	−1.817
0.1	0.6	0.885	0.924	0.907	−3.920	−1.790
0.1	0.7	0.887	0.927	0.909	−4.005	−1.823
0.1	0.8	0.897	0.939	0.920	−4.221	−1.929
0.2	0.1	0.900	0.941	0.921	−4.076	−2.005
0.2	0.2	0.893	0.930	0.912	−3.713	−1.797
0.2	0.3	0.887	0.922	0.905	−3.449	−1.641
0.2	0.4	0.882	0.915	0.900	−3.288	−1.538
0.2	0.5	0.880	0.912	0.897	−3.234	−1.492
0.2	0.6	0.881	0.914	0.898	−3.296	−1.510
0.2	0.7	0.889	0.924	0.908	−3.494	−1.603
0.3	0.1	0.890	0.922	0.906	−3.258	−1.620
0.3	0.2	0.884	0.913	0.899	−2.930	−1.429
0.3	0.3	0.879	0.906	0.893	−2.716	−1.297
0.3	0.4	0.875	0.902	0.889	−2.621	−1.227
0.3	0.5	0.876	0.902	0.890	−2.654	−1.226
0.3	0.6	0.882	0.911	0.898	−2.831	−1.305
0.4	0.1	0.881	0.906	0.894	−2.534	−1.273
0.4	0.2	0.876	0.899	0.888	−2.245	−1.101
0.4	0.3	0.873	0.894	0.884	−2.090	−0.999
0.4	0.4	0.872	0.893	0.883	−2.080	−0.973
0.4	0.5	0.878	0.900	0.890	−2.229	−1.034
0.5	0.1	0.875	0.894	0.884	−1.898	−0.963
0.5	0.2	0.872	0.888	0.880	−1.654	−0.814
0.5	0.3	0.871	0.887	0.879	−1.579	−0.752
0.5	0.4	0.875	0.892	0.884	−1.688	−0.789
0.6	0.1	0.873	0.887	0.880	−1.344	−0.687
0.6	0.2	0.872	0.884	0.878	−1.164	−0.571
0.6	0.3	0.876	0.888	0.882	−1.209	−0.572
0.7	0.1	0.877	0.886	0.882	−0.868	−0.446
0.7	0.2	0.881	0.889	0.885	−0.798	−0.385
0.8	0.1	0.892	0.897	0.895	−0.482	−0.244
1/3	1/3	0.875	0.900	0.888	−2.458	−1.166

Table 9: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

a	b	c	$\Psi_0^{\boldsymbol{\pi}}(D)$	$\Psi_0^{\boldsymbol{\pi}}(R)$	$\Psi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(R)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$	$100\times$ $\left[\Psi_0^{\boldsymbol{\pi}}(M)-\Psi_0^{\boldsymbol{\pi}}(D)\right]$
0.1	0.1	0.1	0.820	0.722	0.811	−9.740	−0.814
0.1	0.1	0.2	0.773	0.688	0.766	−8.483	−0.741
0.1	0.1	0.3	0.736	0.660	0.729	−7.586	−0.693
0.1	0.1	0.4	0.710	0.640	0.703	−7.035	−0.668
0.1	0.1	0.5	0.696	0.628	0.690	−6.839	−0.668
0.1	0.1	0.6	0.700	0.630	0.693	−7.059	−0.696
0.1	0.1	0.7	0.735	0.656	0.727	−7.876	−0.768
0.1	0.2	0.1	0.772	0.691	0.765	−8.138	−0.698
0.1	0.2	0.2	0.730	0.659	0.724	−7.068	−0.639
0.1	0.2	0.3	0.698	0.634	0.692	−6.369	−0.607
0.1	0.2	0.4	0.679	0.618	0.673	−6.043	−0.599
0.1	0.2	0.5	0.676	0.614	0.669	−6.132	−0.619
0.1	0.2	0.6	0.701	0.633	0.694	−6.787	−0.681
0.1	0.3	0.1	0.734	0.665	0.728	−6.914	−0.620
0.1	0.3	0.2	0.696	0.636	0.690	−6.013	−0.574
0.1	0.3	0.3	0.669	0.614	0.664	−5.519	−0.555
0.1	0.3	0.4	0.659	0.604	0.653	−5.459	−0.566
0.1	0.3	0.5	0.675	0.616	0.669	−5.960	−0.617
0.1	0.4	0.1	0.705	0.644	0.699	−6.039	−0.575
0.1	0.4	0.2	0.671	0.618	0.665	−5.306	−0.540
0.1	0.4	0.3	0.652	0.602	0.647	−5.053	−0.538
0.1	0.4	0.4	0.659	0.605	0.653	−5.387	−0.577
0.1	0.5	0.1	0.686	0.631	0.681	−5.500	−0.559
0.1	0.5	0.2	0.658	0.608	0.652	−4.962	−0.537
0.1	0.5	0.3	0.653	0.602	0.647	−5.079	−0.561
0.1	0.6	0.1	0.681	0.627	0.675	−5.312	−0.574
0.1	0.6	0.2	0.662	0.612	0.657	−5.088	−0.573
0.1	0.7	0.1	0.695	0.640	0.689	−5.571	−0.625
0.2	0.1	0.1	0.772	0.694	0.766	−7.742	−0.622
0.2	0.1	0.2	0.730	0.663	0.724	−6.673	−0.565
0.2	0.1	0.3	0.698	0.638	0.692	−5.970	−0.532
0.2	0.1	0.4	0.678	0.622	0.673	−5.632	−0.524
0.2	0.1	0.5	0.675	0.618	0.670	−5.698	−0.541
0.2	0.1	0.6	0.700	0.637	0.694	−6.309	−0.596
0.2	0.2	0.1	0.729	0.665	0.723	−6.330	−0.525
0.2	0.2	0.2	0.691	0.636	0.686	−5.448	−0.483
0.2	0.2	0.3	0.664	0.615	0.660	−4.959	−0.466
0.2	0.2	0.4	0.654	0.605	0.649	−4.888	−0.477
0.2	0.2	0.5	0.670	0.617	0.665	−5.352	−0.523
0.2	0.3	0.1	0.695	0.642	0.690	−5.301	−0.467
0.2	0.3	0.2	0.661	0.615	0.657	−4.598	−0.437
0.2	0.3	0.3	0.643	0.599	0.638	−4.354	−0.437

(continued on next page)

Table 9: Values of $\Psi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

(continued from previous page)

a	b	c	$\Psi_0^{\pi}(D)$	$\Psi_0^{\pi}(R)$	$\Psi_0^{\pi}(M)$	$100\times$ $\left[\Psi_0^{\pi}(R)-\Psi_0^{\pi}(D)\right]$	$100\times$ $\left[\Psi_0^{\pi}(M)-\Psi_0^{\pi}(D)\right]$
0.2	0.3	0.4	0.649	0.602	0.644	−4.663	−0.474
0.2	0.4	0.1	0.671	0.625	0.667	−4.632	−0.441
0.2	0.4	0.2	0.643	0.602	0.639	−4.131	−0.424
0.2	0.4	0.3	0.639	0.596	0.634	−4.244	−0.449
0.2	0.5	0.1	0.660	0.617	0.656	−4.330	−0.447
0.2	0.5	0.2	0.642	0.601	0.638	−4.135	−0.451
0.2	0.6	0.1	0.668	0.624	0.663	−4.468	−0.489
0.3	0.1	0.1	0.732	0.672	0.728	−6.028	−0.466
0.3	0.1	0.2	0.694	0.643	0.690	−5.135	−0.423
0.3	0.1	0.3	0.668	0.621	0.664	−4.630	−0.406
0.3	0.1	0.4	0.657	0.612	0.653	−4.536	−0.413
0.3	0.1	0.5	0.674	0.624	0.669	−4.961	−0.454
0.3	0.2	0.1	0.693	0.645	0.689	−4.790	−0.387
0.3	0.2	0.2	0.659	0.618	0.656	−4.094	−0.359
0.3	0.2	0.3	0.641	0.603	0.637	−3.844	−0.359
0.3	0.2	0.4	0.647	0.606	0.643	−4.124	−0.393
0.3	0.3	0.1	0.664	0.624	0.660	−3.952	−0.347
0.3	0.3	0.2	0.636	0.601	0.633	−3.471	−0.333
0.3	0.3	0.3	0.632	0.596	0.628	−3.573	−0.357
0.3	0.4	0.1	0.647	0.612	0.643	−3.511	−0.342
0.3	0.4	0.2	0.630	0.596	0.626	−3.336	−0.348
0.3	0.5	0.1	0.648	0.613	0.645	−3.523	−0.374
0.4	0.1	0.1	0.701	0.655	0.698	−4.557	−0.341
0.4	0.1	0.2	0.667	0.629	0.664	−3.838	−0.311
0.4	0.1	0.3	0.649	0.613	0.645	−3.560	−0.308
0.4	0.1	0.4	0.655	0.617	0.652	−3.802	−0.337
0.4	0.2	0.1	0.665	0.631	0.663	−3.484	−0.277
0.4	0.2	0.2	0.638	0.608	0.635	−3.000	−0.265
0.4	0.2	0.3	0.633	0.603	0.630	−3.080	−0.286
0.4	0.3	0.1	0.642	0.614	0.640	−2.859	−0.257
0.4	0.3	0.2	0.625	0.598	0.622	−2.690	−0.265
0.4	0.4	0.1	0.637	0.609	0.634	−2.724	−0.278
0.5	0.1	0.1	0.679	0.646	0.676	−3.289	−0.240
0.5	0.1	0.2	0.650	0.623	0.648	−2.769	−0.224
0.5	0.1	0.3	0.646	0.618	0.644	−2.809	−0.240
0.5	0.2	0.1	0.648	0.624	0.646	−2.398	−0.193
0.5	0.2	0.2	0.630	0.608	0.628	−2.215	−0.201
0.5	0.3	0.1	0.634	0.613	0.632	−2.067	−0.198
0.6	0.1	0.1	0.667	0.645	0.666	−2.200	−0.159
0.6	0.1	0.2	0.650	0.630	0.648	−1.969	−0.161
0.6	0.2	0.1	0.644	0.628	0.643	−1.570	−0.136
0.7	0.1	0.1	0.672	0.659	0.671	−1.305	−0.098
0.25	0.25	0.25	0.648	0.606	0.644	−4.137	−0.391

Table 10: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

a	b	c	$\Xi_0^{\boldsymbol{\pi}}(D)$	$\Xi_0^{\boldsymbol{\pi}}(R)$	$\Xi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(D)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(M)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$
0.1	0.1	0.1	0.901	0.947	0.924	−4.527	−2.284
0.1	0.1	0.2	0.896	0.938	0.917	−4.274	−2.114
0.1	0.1	0.3	0.890	0.931	0.911	−4.113	−1.998
0.1	0.1	0.4	0.886	0.926	0.907	−4.048	−1.939
0.1	0.1	0.5	0.883	0.923	0.904	−4.084	−1.941
0.1	0.1	0.6	0.882	0.924	0.904	−4.229	−2.010
0.1	0.1	0.7	0.887	0.932	0.910	−4.503	−2.158
0.1	0.2	0.1	0.892	0.932	0.912	−4.024	−1.998
0.1	0.2	0.2	0.887	0.925	0.906	−3.811	−1.849
0.1	0.2	0.3	0.882	0.919	0.901	−3.705	−1.762
0.1	0.2	0.4	0.878	0.916	0.898	−3.711	−1.743
0.1	0.2	0.5	0.877	0.916	0.898	−3.839	−1.797
0.1	0.2	0.6	0.881	0.922	0.903	−4.106	−1.937
0.1	0.3	0.1	0.883	0.920	0.902	−3.662	−1.779
0.1	0.3	0.2	0.878	0.913	0.897	−3.492	−1.653
0.1	0.3	0.3	0.875	0.909	0.893	−3.450	−1.603
0.1	0.3	0.4	0.873	0.909	0.892	−3.546	−1.634
0.1	0.3	0.5	0.876	0.914	0.896	−3.798	−1.762
0.1	0.4	0.1	0.875	0.910	0.893	−3.439	−1.628
0.1	0.4	0.2	0.872	0.905	0.890	−3.319	−1.531
0.1	0.4	0.3	0.870	0.903	0.888	−3.361	−1.529
0.1	0.4	0.4	0.872	0.908	0.891	−3.585	−1.636
0.1	0.5	0.1	0.869	0.903	0.887	−3.355	−1.549
0.1	0.5	0.2	0.868	0.901	0.886	−3.302	−1.491
0.1	0.5	0.3	0.869	0.904	0.888	−3.472	−1.564
0.1	0.6	0.1	0.867	0.901	0.886	−3.412	−1.546
0.1	0.6	0.2	0.869	0.904	0.888	−3.477	−1.556
0.1	0.7	0.1	0.872	0.908	0.892	−3.636	−1.635
0.2	0.1	0.1	0.888	0.924	0.906	−3.630	−1.854
0.2	0.1	0.2	0.882	0.917	0.900	−3.407	−1.702
0.2	0.1	0.3	0.878	0.911	0.895	−3.288	−1.610
0.2	0.1	0.4	0.874	0.907	0.891	−3.278	−1.583
0.2	0.1	0.5	0.873	0.907	0.891	−3.385	−1.627
0.2	0.1	0.6	0.877	0.913	0.895	−3.626	−1.754
0.2	0.2	0.1	0.879	0.911	0.895	−3.171	−1.590
0.2	0.2	0.2	0.875	0.905	0.890	−2.996	−1.462
0.2	0.2	0.3	0.871	0.900	0.886	−2.945	−1.407
0.2	0.2	0.4	0.869	0.899	0.885	−3.028	−1.432
0.2	0.2	0.5	0.872	0.904	0.889	−3.261	−1.549
0.2	0.3	0.1	0.872	0.900	0.886	−2.870	−1.400
0.2	0.3	0.2	0.868	0.896	0.883	−2.749	−1.302
0.2	0.3	0.3	0.866	0.894	0.881	−2.785	−1.296

(continued on next page)

Table 10: Values of $\Xi_0^{\boldsymbol{\pi}}$ for $r=4$ and $\boldsymbol{\pi}=(a,b,c,1-a-b-c)$.

(continued from previous page)

a	b	c	$\Xi_0^{\boldsymbol{\pi}}(D)$	$\Xi_0^{\boldsymbol{\pi}}(R)$	$\Xi_0^{\boldsymbol{\pi}}(M)$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(D)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$	$100\times$ $\left[\Xi_0^{\boldsymbol{\pi}}(M)-\Xi_0^{\boldsymbol{\pi}}(R)\right]$
0.2	0.3	0.4	0.868	0.898	0.884	−2.996	−1.395
0.2	0.4	0.1	0.866	0.893	0.880	−2.727	−1.288
0.2	0.4	0.2	0.864	0.891	0.878	−2.677	−1.230
0.2	0.4	0.3	0.865	0.893	0.880	−2.841	−1.299
0.2	0.5	0.1	0.863	0.890	0.878	−2.743	−1.259
0.2	0.5	0.2	0.864	0.892	0.879	−2.809	−1.269
0.2	0.6	0.1	0.866	0.896	0.882	−2.940	−1.330
0.3	0.1	0.1	0.876	0.904	0.890	−2.842	−1.469
0.3	0.1	0.2	0.871	0.898	0.884	−2.651	−1.336
0.3	0.1	0.3	0.867	0.893	0.881	−2.583	−1.273
0.3	0.1	0.4	0.866	0.892	0.879	−2.643	−1.287
0.3	0.1	0.5	0.868	0.897	0.883	−2.848	−1.391
0.3	0.2	0.1	0.869	0.893	0.881	−2.429	−1.227
0.3	0.2	0.2	0.865	0.888	0.877	−2.300	−1.125
0.3	0.2	0.3	0.863	0.886	0.875	−2.322	−1.112
0.3	0.2	0.4	0.864	0.889	0.877	−2.513	−1.202
0.3	0.3	0.1	0.863	0.885	0.874	−2.198	−1.071
0.3	0.3	0.2	0.861	0.882	0.872	−2.145	−1.011
0.3	0.3	0.3	0.861	0.884	0.874	−2.298	−1.074
0.3	0.4	0.1	0.860	0.881	0.871	−2.152	−1.008
0.3	0.4	0.2	0.860	0.882	0.872	−2.218	−1.016
0.3	0.5	0.1	0.862	0.885	0.874	−2.309	−1.053
0.4	0.1	0.1	0.866	0.888	0.876	−2.153	−1.125
0.4	0.1	0.2	0.862	0.882	0.872	−2.003	−1.015
0.4	0.1	0.3	0.860	0.880	0.870	−2.001	−0.991
0.4	0.1	0.4	0.862	0.883	0.873	−2.161	−1.066
0.4	0.2	0.1	0.861	0.879	0.870	−1.793	−0.909
0.4	0.2	0.2	0.859	0.876	0.867	−1.728	−0.843
0.4	0.2	0.3	0.859	0.878	0.869	−1.862	−0.896
0.4	0.3	0.1	0.858	0.874	0.866	−1.652	−0.797
0.4	0.3	0.2	0.858	0.875	0.867	−1.711	−0.802
0.4	0.4	0.1	0.859	0.877	0.869	−1.749	−0.808
0.5	0.1	0.1	0.860	0.875	0.867	−1.555	−0.819
0.5	0.1	0.2	0.857	0.872	0.865	−1.463	−0.742
0.5	0.1	0.3	0.858	0.874	0.866	−1.563	−0.779
0.5	0.2	0.1	0.857	0.870	0.864	−1.266	−0.638
0.5	0.2	0.2	0.857	0.870	0.864	−1.309	−0.634
0.5	0.3	0.1	0.859	0.871	0.865	−1.265	−0.597
0.6	0.1	0.1	0.859	0.869	0.863	−1.045	−0.551
0.6	0.1	0.2	0.859	0.869	0.864	−1.050	−0.531
0.6	0.2	0.1	0.860	0.869	0.865	−0.878	−0.430
0.7	0.1	0.1	0.865	0.872	0.868	−0.637	−0.330
0.25	0.25	0.25	0.865	0.890	0.878	−2.503	−1.190

The benefits of using the mixed optimizing strategy seem evident: the loss of efficiency is much smaller when using a quasi-optimal design of this class instead of an optimal discriminating or robust design, than when a robust design is used instead of a discriminant design, or vice-versa.

5. A NOTE ON PSEUDO-CANONICAL MOMENTS OF MEASURES WITH INFINITE SUPPORT

The canonical moments are defined only for measures whose support is a subset of a closed interval. As the canonical moments are closely related with the zeros of monic orthogonal polynomials observing the recurrence relation

$$P_{m+1}(x) \;=\; \big(x-\zeta_{2m}-\zeta_{2m+1}\big)\,P_m(x) \,-\, \zeta_{2m-1}\,\zeta_{2m}\,P_{m-1}(x) \qquad \text{for} \quad m\geq 1$$

with initial conditions $P_0(x)=1$ and $P_1(x)=x-\zeta_1$, it seems worthwhile to try to investigate some "*pseudo-canonical moments*" for measures with infinite support, using the above recurrence relation together with the recurence relation

$$P_{m+1}(x) \;=\; (A_m x + B_m)\,P_m(x) \,-\, C_m\,P_{m-1}(x)\;, \qquad m=0,1,2,...\;,$$

with $P_{-1}(x)=0$ and $A_{m-1}\,A_m\,C_m>0$, valid for any family of orthogonal polynomials.

Let us first examine the gaussian case $\mathrm{d}\mu(x)=\dfrac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x,\ x\in\mathbb{R}$.

It is well known that the Hermite polynomials $H(x)$, recursively defined by

$$H_0(x)=1\;; \qquad H_1(x)=x\;; \qquad H_{n+1}(x)=x\,H_n(x)-n\,H_{n-1}(x) \quad \text{for}\ \ n\geq 1$$

are orthogonal in what regards the measure μ.

Hence, in the gaussian case, the parameters ζ_m are

$$\zeta_1=1\,; \quad \zeta_2=1\,; \quad \zeta_{2m}=-\zeta_{2m+1}=(-1)^{m+1}\,\frac{m\times(m-2)\times\,\cdots}{(m-1)\times(m-3)\times\,\cdots} \quad \text{for}\ m\geq 2\,.$$

Using the definition $m\geq 1,\ \zeta_m=\chi^*_m\,\big(1-\chi^*_{m-1}\big)$, we get

$$\chi^*_m=\frac{\zeta_m}{1-\chi^*_{m-1}} \qquad (\text{with}\ \ \chi^*_0=0)\;.$$

When $m=2$ the denominator of the previous fraction is null and therefore the gaussian distribution has only the first pseudo-canonical moment as indicated in Table 11.

Table 11: Gaussian pseudo-canonical moments, $n \leq 5$.

i	χ_i^*
1	1
2	—
3	—
4	—
5	—

Similarly, for the gamma measure with shape parameter $\alpha > 0$,

$$\mathrm{d}\mu(x) = \frac{x^{\alpha-1}\exp(-x)}{\Gamma(\alpha)}\,\mathrm{d}x\ , \qquad x \geq 0\ ,$$

which is associated with the generalized Laguerre polynomials $L^{(\alpha)}(x)$ defined by

$$L_{n+1}^{(\alpha)}(x) = (x-2n-1-\alpha)\,L_n^{(\alpha)}(x) + (n+\alpha)\,L_{n-1}^{(\alpha)}(x) \qquad \text{for} \quad n \geq 1\ ,$$

with the initial values $L_0^{(\alpha)}(x) = 1$, $L_1^{(\alpha)}(x) = x - \alpha - 1$, we get

$$\begin{cases} \zeta_{2m} = -\dfrac{m+\alpha}{\zeta_{2m-1}} \\ \zeta_{2m+1} = 2m+1+\alpha-\zeta_{2m} \end{cases} .$$

Using the fact that $\zeta_0 = 1$ and the relation $\chi_m^* = \dfrac{\zeta_m}{1-\chi_{m-1}^*}$ (with $\chi_0^* = 0$), the pseudo-canonical moments of a gamma measure with shape $\alpha = a$ up to order n are readily computed using the script

```
zeta(1) = a
zeta(2) = (1+a)/a
zeta(3) = (3+a)-zeta(2)
for j = 2:n
    zeta(2*j) = (j+a)/zeta(2*j-1)
    zeta(2*j+1) = 2*j+1+a-zeta(2*j)
end
chi(1) = zeta(1)
for j = 2:(2*n+1)
    chi(j) = zeta(j)/(1-chi(j-1))
end
```

In the table below we exhibit, as an example, the pseudo-canonical moments up to $n = 20$ for the gamma measure with shape parameter $\alpha = 3$.

Table 12: Gamma-3 pseudo-canonical moments, $n \leq 20$.

i	χ_i^*	i	χ_i^*	i	χ_i^*	i	χ_i^*
1	3	6	0.0867	11	15.4130	16	0.0323
2	0.6667	7	11.7057	12	0.0427	17	21.2785
3	22.000	8	0.0612	13	17.3558	18	0.0287
4	0.0325	9	13.4791	14	0.0368	19	23.2510
5	8.9732	10	0.0507	15	19.3125	20	0.0259

Observe that $\chi_{2n}^* \neq \frac{1}{2}$ in the case of the gaussian (while for symmetric measures with support $S \subseteq [a,b]$ we always have $\chi_{2n} = \frac{1}{2}$); or, in the case of the gamma measure, for which χ_2^* does exist, χ_2^* isn't associated with the raw moments via $\chi_2 = \dfrac{m_2 - m_1^2}{m_1(1-m_1)}$, a relation which holds true for the canonical moments of finite support measures.

These two examples plainly show that the pseudo-canonical moments do not possess the nice properties canonical moments do satisfy in the case of measures whose support is a subset of a compact interval.

ACKNOWLEDGMENTS

Research partially supported by FCT/POCTI and POCI/FEDER.

The authors wish to thank Prof. João Tiago Mexia (Universidade Nova de Lisboa, Portugal) for his kind encouragement during the preparation of this paper, and Prof. Roman Zmyślony (University of Zielona Góra, Poland) for his perceptive and inspiring comments at the IWMS'08 — 17$^{\text{th}}$ International Workshop on Matrices and Statistics in Tomar.

REFERENCES

[1] Anderson, T.W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem, *Ann. Math. Statist.*, **33**, 255–265.

[2] Box, G.E.P. and Draper, N.R. (1959). A basis for the selection of a response surface design, *J. Amer. Statist. Assoc.*, **54**, 622–654.

[3] Box, G.E.P. and Draper, N.R. (1963). The choice of a second order rotable design, *Biometrika*, **50**, 335–352.

[4] Dette, H. and Studden, W.J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, Wiley, New York.

[5] Fedorov, V.V. (1972). *Theory of Optimal Experiments*, Academic Press, New York.

[6] Guest, P.G. (1958). The spacing of observations in polynomial regression, *Ann. Math. Statist.*, **29**, 294–299.

[7] Henrici, P. (1977). *Applied and Computational Complex Analysis*, Vol. 2: *Special Functions, Integral Transforms, Asymptotics and Continued Fractions*, Wiley, New York.

[8] Hoel, P.G. (1958). Efficiency problems in polynomial estimation, *Ann. Math. Statist.*, **29**, 1134–1145.

[9] Kiefer, J. (1959). Optimum experimental designs, *J. Roy. Statist. Soc.*, **B21**, 273–319.

[10] Kiefer, J. (1961). Optimum designs, *Ann. Math. Statist.*, **32**, 298–325.

[11] Kiefer, J. (1962). Two more criteria equivalent to D-optimality of designs, *Ann. Math. Statist.*, **33**, 792–796.

[12] Kiefer, J. and Wolfowitz, J. (1959). Optimium designs in regression problems, *Ann. Math. Statist.*, **30**, 271–292.

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

[14] Kraft, O. and Schaefer, M. (1995). Exact Elfving-minimax designs for quadratic regression, *Statistics Sinica*, **5**, 475–485.

[15] Kussmaul, K. (1969). Protection against assuming the wrong degree in polynomial regression, *Technometrics*, **11**, 677–682.

[16] Lau, T.S. and Studden, W.J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments, *Ann. Statistics*, **13**, 383–394.

[17] Pukelsheim, F. (1993). *Optimal Design of Experiments*, Wiley, New York.

[18] Skibinski, M. (1967). The range of the $(n+1)$-th moment for distributions on $[0, 1]$, *J. Applied Probab.*, **4**, 543–552.

[19] Skibinski, M. (1969). Some striking properties of Binomial and Beta moments, *Ann. Math. Statist.*, **40**, 1753–1764.

[20] Smith, K. (1918). On the standard deviations and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distributions of observations, *Biometrika*, **12**, 1–85.

[21] Spruill, M.C. (1990). Good designs for testing the degree of a polynomial mean, *Sankhya*, **B52**, 67–74.

[22] Stigler, S.M. (1971). Optimal experimental design for polynomial regression, *J. Amer. Statist. Assoc.*, **66**, 311–318.