Optimal Design Criteria Based on Tolerance Regions

Christine H. Müller[1] and Christos P. Kitsos[2]

[1] Carl von Ossietzky University Oldenburg, Institute for Mathematics, Postfach 2503, D-26111 Oldenburg, Germany, mueller@math.uni-oldenburg.de
[2] Department of Mathematics, Technological Educational Institute of Athens, Ag. Spyridonos, GR-12210 Egaleo, Greece, xkitsos@teiath.gr

Summary. New design criteria are derived by minimizing the tolerance regions in linear models. Since the regions usually are not completely ordered a complete ordering is provided by regarding the volume, the longest axis or the sum of the axes of the region as design criteria leading to *TD*-, *TE*- and *TA*-optimal designs. It is shown that the *TE*- and the *TA*-optimal designs coincide with the classical *E*- and *A*-optimal designs. However, the *TD*-optimal designs and *D*-optimal designs can be different which is shown by an example.

Key words: β-expectation tolerance region, prediction region, Bayesian tolerance region, linear model, optimal designs

1 Introduction

For the vector $Y = (Y_1 \ldots, Y_N)^\top$ of observations we assume a general linear model of the form

$$Y = X\gamma + \sigma\,\varepsilon \qquad \text{with } \varepsilon \sim \mathcal{N}(0_N, I_{N\times N}),$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)$ is the vector of errors, $\gamma \in \Re^r$ the unknown parameter vector, $\sigma \in \Re^+$ an unknown scale parameter, $X = (x_1, \ldots, x_N)^\top \in \Re^{N\times r}$ the known design matrix, briefly also called design, and $\mathcal{N}(0_N, I_{N\times N})$ is the normal distribution with mean vector 0_N of zeros and covariance matrix $I_{N\times N}$, the identity matrix. Given a realization y of Y, the aim is to construct a region $R(X, y) \subset \Re^M$ in which the vector V of future observations $V_1, \ldots, V_M$ will fall with a high probability. For that we assume that the vector $V = (V_1, \ldots, V_M)^\top$ of future observations will follow the same probability law as Y, namely

$$V = W\gamma + \sigma\,\varepsilon_0 \qquad \text{with } \varepsilon_0 \sim \mathcal{N}(0_M, I_{M\times M}),$$

where W is the design matrix of the future observations. Hence, the distribution of V is given by the same parameter $\theta = (\gamma, \sigma) \in \Omega := \Re^r \times \Re^+$ as the distribution of Y. Given θ, Y and V shall be independent.

It is impossible to construct a region $R(X, y)$ so that V lies in $R(X, y)$ with high probability for all θ and all realizations y of Y, i.e., $P_{V|\theta}(R(X, y))$ is large for all θ and y. One compromise are β-expectation tolerance regions or mean coverage tolerance regions where $R(X, y)$ shall satisfy

$$\int P_{V|\theta}(R(X,y))\ f_{Y|\theta}(y)\, dy = \beta \tag{1}$$

for all $\theta \in \Re^r \times \Re^+$ (see e.g. [AD75], [Gut81], [Jil81], [JA89]). But this means that $R(X, y)$ is also a prediction region (see [Pau43]) since

$$\int P_{V|\theta}(R(X,y))\ f_{Y|\theta}(y)\, dy = P_{(Y,V)|\theta}(\{(y,v);\ v \in R(X,y)\}).$$

By standard arguments used in linear models, it is easy to see that $R(X, y)$ is given by

$$R(X,y) = \Bigg\{ v \in \Re^M;\ \frac{1}{\hat{\sigma}(X,y)^2} \tag{2}$$
$$\cdot\, (v - W\hat{\gamma}(X,y))^\top\, S(X)^{-1}\, (v - W\hat{\gamma}(X,y)) \le \frac{M}{N-r}\, F_{M,N-r,\beta} \Bigg\},$$

where

$$S(X) := I_{M\times M} + W(X^\top X)^{-1} W^\top,$$
$$\hat{\gamma}(X,y) := (X^\top X)^{-1} X^\top y,$$
$$\hat{\sigma}(X,y)^2 := (y - X\hat{\gamma}(X,y))^\top (y - X\hat{\gamma}(X,y)),$$

and $F_{M,N-r,\beta}$ is the β quantile of the F distribution with M and $N-r$ degrees of freedom. But $R(X, y)$ given by (2) is also a β-expectation tolerance region with respect to θ in a Bayesian sense as studied in [HR76], i.e., it satisfies

$$\beta = P_{V|Y=y}(R(X,y)) = \int_{R(X,y)} f_{V|Y=y}(v)\, dv \tag{3}$$
$$= \int_{\Omega} \int_{R(X,y)} f_{V|\Theta=\theta}(v)\, dv\ f_{\Theta|Y=y}(\theta)\, d\theta = \int_{\Omega} P_{V|\Theta=\theta}\,(R(X,y))\ P_{\Theta|Y=y}(d\theta)$$

for all $y \in \Re^N$. This has the advantage that the mean coverage property holds for all y. Property (3) can be seen by the structural approach of [Fra68]. A modern derivation bases on invariant measures and is given in Section 2.

In Section 3 we investigate designs X that minimize the β-expectation tolerance regions of form (2). In particular, we study design criteria for minimizing $R(X, y)$ and compare the new design criteria with the classical criteria based on minimizing confidence regions for $W\gamma$.

2 β-expectation Bayesian tolerance regions

At first note the following representation of $S(X)^{-1}$.

Lemma 1. $S(X)^{-1} = I_{M\times M} - W\,(X^\top X + W^\top W)^{-1} W^\top$.

Proof. Set $L := X^\top X$. Then we have

$$(L + W^\top W)^{-1} = L^{-1} - L^{-1}W^\top (I_{M\times M} + W\,L^{-1}W^\top)^{-1} W\,L^{-1}.$$

This implies

$$\begin{aligned}
&I_{M\times M} - W(L + W^\top W)^{-1}W^\top \\
&= I_{M\times M} - W\,L^{-1}W^\top \\
&\quad + W\,L^{-1}W^\top (I_{M\times M} + W\,L^{-1}W^\top)^{-1} W\,L^{-1}W^\top \\
&= I_{M\times M} - W\,L^{-1}W^\top (I_{M\times M} + W\,L^{-1}W^\top)^{-1}\,(I_{M\times M} + W\,L^{-1}W^\top) \\
&\quad + W\,L^{-1}W^\top (I_{M\times M} + W\,L^{-1}W^\top)^{-1} W\,L^{-1}W^\top \\
&= I_{M\times M} - W\,L^{-1}W^\top (I_{M\times M} + W\,L^{-1}W^\top)^{-1} \\
&= (I_{M\times M} + W\,L^{-1}W^\top)\,(I_{M\times M} + W\,L^{-1}W^\top)^{-1} \\
&\quad - W\,L^{-1}W^\top (I_{M\times M} + W\,L^{-1}W^\top)^{-1} \\
&= (I_{M\times M} + W\,L^{-1}W^\top)^{-1} = S(X)^{-1}. \square
\end{aligned}$$

The structural inference, as described by [Fra68], bases on the underlying group structures of the parameter space $\Omega = \Re^r \times \Re^+$ and the set of transformations of the observations. If we define the operator $\circ$ on Ω as $\theta \circ \theta_* = (\gamma, \sigma) \circ (\gamma_*, \sigma_*) := (\gamma + \sigma\,\gamma_*, \sigma\,\sigma_*)$ then $(\Omega, \circ)$ is a non-abelian group. Let

$$A_\theta = A_{(\gamma,\sigma)} := \begin{pmatrix} I_{r\times r} & \gamma \\ 0_{1\times r} & \sigma \end{pmatrix}$$

and $g_\theta : \Re^{N\times r} \times \Re^N \to \Re^{N\times r} \times \Re^N$ be defined by

$$g_\theta(X, y) := (X, y)\,A_\theta = (X, X\gamma + \sigma y).$$

Then $\mathcal{G} := \{g_\theta;\ \theta \in \Omega\}$ with the composition $\circ$ of functions is also a group. Since $X\gamma + \sigma Y \sim \mathcal{N}(X(\gamma + \sigma\,\gamma_*), \sigma^2\,\sigma_*^2\,I_{N\times N})$ if $Y \sim \mathcal{N}(X\gamma_*, \sigma_*^2\,I_{N\times N})$ the linear model is invariant with respect to $(\mathcal{G}, \circ)$ and $(\Omega, \circ)$. Then, taking the right invariant measure (right Haar measure) on Ω as prior distribution the posterior distribution has a simple form (see [Fra68], p. 127). Note, that the right Haar measure ρ on Ω is given by $\frac{\rho(d(\gamma,\sigma))}{d(\gamma,\sigma)} = \frac{1}{\sigma} 1_{(0,\infty)}(\sigma)$.

Lemma 2. *Let the right Haar measure ρ the prior distribution of Θ. Then the posterior distribution of Θ given $Y = y$ has a Lebesgue density of the form*

$$\begin{aligned} & f_{\Theta|Y=y}(\theta) \qquad (4) \\ & \propto \exp\left\{-\frac{1}{2\sigma^2}\left[(\hat\gamma(X,y)-\gamma)^\top X^\top X(\hat\gamma(X,y)-\gamma) - \hat\sigma(X,y)^2\right]\right\} \\ & \quad \cdot \hat\sigma(X,y)^{N-r}\left(\frac{1}{\sigma}\right)^{N+1} 1_{(0,\infty)}(\sigma), \end{aligned}$$

where $\propto$ means equality with the exception of constants.

Proof. Set

$$\begin{aligned} T(X,y) &:= (\hat\gamma(X,y), \hat\sigma(X,y)) \in \Re^r \times \Re^+, \\ U(X,y) &:= \hat\sigma(X,y)^{-1}\left(Y - X\hat\gamma(X,y)\right) \in \mathcal{U} := \Re^N, \\ W(X,y) &:= (T(X,y), U(X,y)), \end{aligned}$$

where the mapping W maps the observation space $\Re^{N\times r}\times\Re^N$ into $\Omega\times\mathcal{U}$. The mapping $T : \Re^{N\times r}\times\Re^N \to \Omega$ is equivariant, i.e. $T(g_\theta(X,y)) = \theta \circ T(X,y)$, and the mapping $U : \Re^{N\times r}\times\Re^N \to \mathcal{U}$ is invariant, i.e. $U(g_\theta(X,y)) = U(X,y)$. The transformations T and U together with $\varepsilon \sim \mathcal{N}(0_N, I_{N\times N})$ provide the structural linear model (see [Fra68]).

Moreover, $T(X,Y)$ and $U(X,Y)$ are independent according to the theorem of Basu since T is a complete sufficient statistic, and also $\hat\gamma(X,Y)$ and $\hat\sigma(X,Y)$ are independent (see e.g. [Sch95]). If $\Theta = \theta = (\gamma,\sigma)$ then $\hat\gamma(X,Y)$ has a $\mathcal{N}(\gamma, \sigma^2 X^\top X)$ distribution and $\frac{1}{\sigma^2}\hat\sigma(X,Y)^2$ has a χ^2 distribution with $N-r$ degrees of freedom. Hence $P_{W(X,Y)|\Theta=\theta}$ has a Lebesgue density of the form $f_{W(X,Y)|\Theta=\theta}(\hat\gamma,\hat\sigma,\hat u) = f_{\hat\gamma(X,Y)|\Theta=\theta}(\hat\gamma)\cdot f_{\hat\sigma(X,Y)|\Theta=\theta}(\hat\sigma)\cdot f_{U(X,Y)}(\hat u)$, where

$$f_{\hat\gamma(X,Y)|\Theta=\theta}(\hat\gamma) \propto \left(\frac{1}{\sigma}\right)^r \exp\left\{-\frac{1}{2\sigma^2}(\hat\gamma-\gamma)^\top X^\top X(\hat\gamma-\gamma)\right\}$$

and

$$f_{\hat\sigma(X,Y)|\Theta=\theta}(\hat\sigma) \propto \left(\frac{\hat\sigma}{\sigma}\right)^{N-r-1} \exp\left\{-\frac{1}{2}\left(\frac{\hat\sigma}{\sigma}\right)^2\right\}\frac{1}{\sigma}\, 1_{(0,\infty)}(\hat\sigma).$$

Since the left Haar measure (or left invariant measure) λ on $\Omega = \Re^r\times\Re^+$ has a Lebesgue-density of the form $\frac{\lambda(d(\hat\gamma,\hat\sigma))}{d(\hat\gamma,\hat\sigma)} = \frac{1}{\hat\sigma^{r+1}} 1_{(0,\infty)}(\hat\sigma)$ we can express the Lebesgue density $f_{\hat\gamma(X,Y)|\Theta=\theta}\cdot f_{\hat\sigma(X,Y)|\Theta=\theta}$ also as a density $\tilde f_{\hat\gamma(X,Y)|\Theta=\theta}\cdot \tilde f_{\hat\sigma(X,Y)|\Theta=\theta}$ with respect to the left Haar measure λ. Let ν be the Lebesgue measure on $\Re^N$; then $\tilde f_{W(X,Y)|\Theta=\theta}(\hat\gamma,\hat\sigma,\hat u) = \tilde f_{\hat\gamma(X,Y)|\Theta=\theta}(\hat\gamma)\cdot\tilde f_{\hat\sigma(X,Y)|\Theta=\theta}(\hat\sigma)\cdot f_{U(X,Y)}(\hat u)$ is the $\lambda\otimes\nu$ density of $W(X,Y)$ given $\Theta=\theta$. The right Haar measure which is related to λ is the prior distribution ρ. Hence the conditions of Assumption 6.58 in [Sch95], p. 368/369, are satisfied so that Lemma 6.65

on p. 371 in [Sch95] holds which provides the posterior distribution of Θ given $W(X,Y) = (\hat{\gamma}, \hat{\sigma}, \hat{u})$. In particular, using the relation between λ and ρ, we have

$$\int \tilde{f}_{T(X,Y)|\Theta=\theta}(\hat{\gamma}, \hat{\sigma})\rho(d\theta) = \hat{\sigma}^r$$

so that $\tilde{f}_{W(X,Y)}(\hat{\gamma}, \hat{\sigma}, \hat{u}) = \hat{\sigma}^r f_{U(X,Y)}(\hat{u})$ is the marginal density. Then Bayes theorem provides that the posterior distribution has a density with respect to ρ of the form

$$\begin{aligned} &\tilde{f}_{\Theta|W(X,Y)=(\hat{\gamma},\hat{\sigma},\hat{u})}(\theta) \\ &= \frac{\tilde{f}_{W(X,Y)|\Theta=\theta}(\hat{\gamma}, \hat{\sigma}, \hat{u})}{\hat{\sigma}^r f_{U(X,Y)}(\hat{u})} = \frac{1}{\hat{\sigma}^r} \tilde{f}_{\hat{\gamma}(X,Y)|\Theta=\theta}(\hat{\gamma})\, \tilde{f}_{\hat{\sigma}(X,Y)|\Theta=\theta}(\hat{\sigma}). \end{aligned}$$

Using the relations of the Haar measures ρ and λ to the Lebesgue measure on $\Re^r \times \Re^+$ we see that the posterior distribution has a Lebesgue density of the form

$$\begin{aligned} f_{\Theta|W(X,Y)=(\hat{\gamma},\hat{\sigma},\hat{u})}(\theta) &= \frac{1}{\hat{\sigma}^r} f_{\hat{\gamma}(X,Y)|\Theta=\theta}(\hat{\gamma})\, f_{\hat{\sigma}(X,Y)|\Theta=\theta}(\hat{\sigma})\, \hat{\sigma}^{r+1} \frac{1}{\sigma} \\ \propto \exp\left\{-\frac{1}{2\sigma^2}\left[(\hat{\gamma}-\gamma)^\top X^\top X(\hat{\gamma}-\gamma) - \hat{\sigma}^2\right]\right\} &\hat{\sigma}^{N-r} \left(\frac{1}{\sigma}\right)^{N+1} 1_{(0,\infty)}(\sigma). \end{aligned}$$

Note that $f_{\Theta|W(X,Y)=(\hat{\gamma},\hat{\sigma},\hat{u})} = f_{\Theta|Y=y}$ if $\hat{\gamma} = \hat{\gamma}(X,y)$, $\hat{\sigma} = \hat{\sigma}(X,y)$, $\hat{u} = U(X,y)$, since W is an injective mapping. Hence the assertion is proved. □

From the posteriori density, we can deduce the β-expectation tolerance region as proposed by [HR76].

Theorem 1. *The region $R(X,y) \subset \Re^M$ given by (2) is a β-expectation tolerance region in the sense of (3), i.e., it satisfies $P_{V|Y=y}(R(X,y)) = \beta$ for all $y \in \Re^N$.*

Proof. Since V given $\Theta = \theta$ has a $\mathcal{N}(W\gamma, \sigma^2 I_{M\times M})$ distribution the Lebesgue density of V given $Y = y$ is with (4)

$$\begin{aligned} f_{V|Y=y}(v) &= \int_\Omega f_{V|\Theta=\theta}(v)\; f_{\Theta|Y=y}(\theta)\, d\theta \\ \propto \int_\Omega &\exp\left\{-\frac{1}{2\sigma^2}\left[(\hat{\gamma}-\gamma)^\top X^\top X(\hat{\gamma}-\gamma) + (v-W\gamma)^\top(v-W\gamma)\right]\right\} \\ &\cdot \exp\left\{-\frac{1}{2\sigma^2}\hat{\sigma}^2\right\} \hat{\sigma}^{N-r} \left(\frac{1}{\sigma}\right)^{N+M+1} d\theta. \end{aligned}$$

Set

$$h(X,y,v) := (X^\top X + W^\top W)^{-1}(X^\top X\,\hat{\gamma} + W^\top v).$$

Then we have with Lemma 1

$$\begin{aligned}&(\hat{\gamma}-\gamma)^\top X^\top X(\hat{\gamma}-\gamma)+(v-W\gamma)^\top(v-W\gamma)\\&=(\gamma-h(X,y,v))^\top(X^\top X+W^\top W)(\gamma-h(X,y,v))\\&\quad+(v-W\hat{\gamma})^\top S(X)^{-1}(v-W\hat{\gamma})\end{aligned}$$

which implies by integration of the density of the normal and χ^2 distribution

$$\begin{aligned}&f_{V|Y=y}(v)\\&\propto\int_{\Re^+}\left[\int_{\Re^r}\left(\frac{1}{\sigma}\right)^r\exp\left\{-\frac{1}{2\sigma^2}\left[(\gamma-h(X,y,v))^\top(X^\top X\right.\right.\right.\\&\quad\left.\left.\left.+\ W^\top W)(\gamma-h(X,y,v))\right]\right\}d\gamma\right]\cdot\\&\exp\left\{-\frac{1}{2\sigma^2}\left[(v-W\hat{\gamma})^\top S(X)^{-1}(v-W\hat{\gamma})+\hat{\sigma}^2\right]\right\}\hat{\sigma}^{N-r}\left(\frac{1}{\sigma}\right)^{N-r+M+1}d\sigma\\&\propto\left(1+\hat{\sigma}^{-2}\,(v-W\hat{\gamma})^\top S(X)^{-1}(v-W\hat{\gamma})\right)^{-(N-r+M)/2}\,\hat{\sigma}^{-M}.\end{aligned}$$

This means that $\frac{\sqrt{N-r}}{\hat{\sigma}}\ (V-W\hat{\gamma})^\top S(X)^{-1/2}$ given $Y=y$ has a multivariate T distribution with $N-r$ degrees of freedom. Thus $\frac{N-r}{M}\frac{1}{\hat{\sigma}^2}\ (V-W\hat{\gamma})^\top S(X)^{-1}(V-W\hat{\gamma})$ given $Y=y$ has an F distribution with M and $N-r$ degrees of freedom (see [TG65]). Hence, the assertion follows. □

3 Optimal designs

Optimal designs for β-expectation tolerance regions will be those designs that provide that $R(X,y)$ given by (2) is as small as possible. The best situation is that a design X_* provides a region $R(X_*,y)$ with $R(X_*,y)\subset R(X,y)$ for all other designs X and all $y\in\Re^N$. This means that $S(X_*)\leq S(X)$ for all X in the positive-semidefinite sense (i.e., $A\leq B$ iff $c^\top A\,c\leq c^\top B\,c$ for all $c\in\Re^M$). Such designs are called uniform optimal for β-expectation tolerance regions, shortly *TU*-optimal. But as in classical design theory such designs do not exist except in degenerate cases since $S(X)$ and thus $R(X,y)$ are not completely ordered. To get a complete ordering we can regard the volume of the region $R(X,y)$ which is a function of $\det(S(X))$, the determinant of $S(X)$. Since $R(X,y)$ is an ellipsoid we can also consider the sum of the axes of the ellipsoid expressed by $\mathrm{tr}(S(X))$, the trace of $S(X)$, or the longest axis expressed by $\lambda_{\max}(S(X))$, the maximum eigenvalue of $S(X)$.

Definition 1. *Let Δ be a set of competing designs. A design $X_*\in\Delta$ is called*
(i) TU-optimal if $S(X_*)\leq S(X)$ for all $X\in\Delta$,
(ii) TD-optimal if $\det(S(X_*))\leq\det(S(X))$ for all $X\in\Delta$,
(iii) TA-optimal if $\mathrm{tr}(S(X_*))\leq\mathrm{tr}(S(X))$ for all $X\in\Delta$,
(iv) TE-optimal if $\lambda_{\max}(S(X_*))\leq\lambda_{\max}(S(X))$ for all $X\in\Delta$.

We will compare the new design criteria with the classical design criteria which are motivated by minimizing the confidence region for $W\gamma$. The confidence region for $W\gamma$ is given by

$$C(X,y) = \Big\{ v \in \Re^M;\ \frac{1}{\hat\sigma(X,y)^2} \cdot \left(v - W\hat\gamma(X,y)\right)^\top \left(W(X^\top X)^{-1}W^\top\right)^{-1} \left(v - W\hat\gamma(X,y)\right) \le \frac{M}{N-r}\, F_{M,N-r,\beta} \Big\}$$

(see e.g. [Chr87]). Hence minimizing the set $C(X,y)$, the volume, the sum of the axes, or the longest axis of $C(X,y)$ leads to the classical design criteria of *U*-, *D*-, *A*-, and *E*-optimality based on $W(X^\top X)^{-1}W^\top$, $\det(W(X^\top X)^{-1}W^\top)$, $\mathrm{tr}(W(X^\top X)^{-1}W^\top)$, and $\lambda_{\max}(W(X^\top X)^{-1}W^\top)$ (see e.g. [Paz86] or [Puk93]).

From the form of $S(X)$ it is clear that the tolerance regions are always larger than the confidence regions which is due to the variability of the future observations. Nevertheless almost all new design criteria are equivalent to the classical criteria.

Theorem 2. *Let Δ be a set of competing designs and $X_* \in \Delta$. Then we have:*
(i) X_* is TU-optimal if and only if X_* is U-optimal.
(ii) X_* is TA-optimal if and only if X_* is A-optimal.
(iii) X_* is TE-optimal if and only if X_* is E-optimal.

Proof. Assertion (i) is obvious. Now let $\lambda_1, \dots, \lambda_M$ be the eigenvalues of $W\,(X^\top X)^{-1}W^\top$. Then $1+\lambda_1, \dots, 1+\lambda_M$ are the eigenvalues of $S(X)$. This implies (iii). Since the trace of a matrix is the sum of its eigenvalues also (ii) follows. □

Theorem 2 shows that the *TU*-, *TA*-, and *TE*- optimal designs can be constructed via the methods which were developed for the classical criteria (see e.g. [Paz86] or [Puk93]). The only exception is the *TD*-criterium.

Although the determinant of a matrix is the product of the eigenvalues of the matrix the proof of Theorem 2 cannot be adapted for the *D*-criterium. Namely, if $\lambda_1, \dots, \lambda_M$ are the eigenvalues of $W\,(X^\top X)^{-1}W^\top$ then $\det(W\,(X^\top X)^{-1}W^\top) = \prod_{m=1}^M \lambda_m$ and $\det(S(X)) = \prod_{m=1}^M (1+\lambda_m)$. Hence, a minimum of $\prod_{m=1}^M \lambda_m$ is not equivalent to a minimum of $\det(S(X)) = \prod_{m=1}^M (1+\lambda_m)$. The following example shows that there are really situations where the *D*- and *TD*-optimal designs are different.

Example. Consider a simple linear regression model where any observation y_n has the form $y_n = (1, t_n)\gamma + \sigma\varepsilon_n$ with $t_n \in \Re$ and $\gamma \in \Re^2$. Every future observation v_m follows the same linear regression model, i.e. $v_m = (1, u_m)\gamma + \sigma\varepsilon_{0m}$ with $u_m \in \Re$. Assume that we have $\alpha\,N$ observations at $t_n = 1$ and $(1-\alpha)\,N$ observations at $t_n = 0$ so that the design matrix depends on α and shall be denoted by X_α. The proportion α shall be chosen so that the tolerance region for two future observations V_1 and V_2, one at $u_1 = 2$ and the other at $u_2 = 3$, is as small as possible. Then we have

$$W\,(X_\alpha^\top X_\alpha)^{-1} W^\top = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} N & \alpha N \\ \alpha N & \alpha N \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}^\top$$

so that

$$\det(W\,(X_\alpha^\top X_\alpha)^{-1} W^\top) = \frac{1}{N^2}\,\frac{1}{\alpha(1-\alpha)} \tag{5}$$

and

$$\begin{aligned} \det(S(X_\alpha)) &= \det(I_{2\times 2} + W\,(X_\alpha^\top X_\alpha)^{-1} W^\top) \\ &= \frac{1}{N^2}\,\frac{1}{\alpha-\alpha^2}\,\left(N^2(\alpha-\alpha^2) + N(13-8\alpha) + 1\right). \end{aligned} \tag{6}$$

Quantity (5) is minimized by $\alpha = 0.5$ while the minimum of quantity (6) depends on N and is minimized, for example, by $\alpha = 0.616$ for $N = 10$ and by $\alpha = 0.617$ for $N = 50$. Hence, the *TD*-optimal designs are different from the classical *D*-optimal designs if Δ is the set of all designs with observations at 0 and 1, i.e. $t_n \in \{0, 1\}$. The *TD*-optimal design puts more observations at 1 than the *D*-optimal design. The same holds if the set of competing designs is Δ^*, the set of all designs with $t_n \in [0, 1]$.

References

[AD75] Aitchison, J., Dunsmore, I. Statistical Prediction Analysis. University Press, Cambridge (1975)

[Chr87] Christensen, R.: Plane Answers to Complex Questions - The Theory of Linear Models. Springer, New York (1987)

[Fra68] Fraser, D.A.S.: The Structure of Inference. Wiley, New York (1968)

[Gut81] Guttman, I.: Statistical Tolerance Regions. Classical and Bayesian. Griffin, London (1981)

[HR76] Haq, M.S, Rinco, S.: β-expectation tolerance regions for a generalized multivariate model with normal error variables. J. Mult. Anal., **6**, 414–421 (1976)

[Jil81] Jílek, M.: A bibliography of statistical tolerance regions. Math. Operationsforschung und Statistik - Series Statistics, **12**, 441–456 (1981)

[JA89] Jílek, M., Ackermann, H.: A bibliography of statistical tolerance regions II. Statistics, **20**, 165–172 (1989)

[Pau43] Paulson, E.: A note on control limits. Ann. Math. Statist., **14**, 90–93 (1943)

[Paz86] Pázman, A.: Foundations of Optimum Experimental Design. Reidel, Dordrecht (1986)

[Puk93] Pukelsheim, F.: Optimal Design of Experiments. Wiley, New York (1993)

[Sch95] Schervish, M.J.: Theory of Statistics. Springer, New York (1995)

[TG65] Tiao, G.G., Guttman, I.: The inverted Dirichlet distribution with applications. J. Amer. Statist. Assoc., **60**, 793–805 (1965)