Design Workshop Lecture Notes
ISI, Kolkata, November, 25-29, 2002, pp. 33-49

DS-Optimal Designs

Rita SahaRay
Theoretical Statistics and Mathematics Unit, Indian Statistical Institute
Kolkata, India

1 Introduction

The problem of characterization and construction of optimal designs under both discrete and continuous set up using the well known A-, D- and E- and universal optimality criteria has been extensively studied in the literature. See for example , Shah and Sinha(1989), Pukelsheim(1993). However, the study of the distance optimality criterion put forward by Sinha(1970) in certain treatment designs settings has received relatively less attention. Recently there has been a growing interest in this direction (cf. Liski, Luoma, Mandal and Sinha (1998); Liski, Luoma and Zaigraev (1999); Mandal, Shah and Sinha(2000), SahaRay and Bhandari(2001), Liski, Mandal, Shah, Sinha(2002)). Our present discussion is based on the papers cited above. In the process we first touch upon the definition of DS optimality criterion and then discuss its properties and finally characterize DS optimal designs in various settings. We also mention about the connection of this optimality criterion with the well known D- and E- criteria.

We start with a classical linear model

$$\underline{Y} \sim N(X\underline{\beta}, \sigma^2 I_N) \tag{1.1}$$

where the $N \times 1$ response vector $\underline{Y} = (Y_1, \ldots, Y_N)'$ follows a multivariate normal distribution, $X = (\underline{X}_1, \ldots, \underline{X}_N)'$ is the $N \times m$ design matrix, and $\underline{\beta} = (\beta_1, \ldots, \beta_m)'$ is the $m \times 1$ parameter vector. $E(\underline{Y}) = X\underline{\beta}$, and $D(\underline{Y}) = \sigma^2 I_N$ are respectively the expectation vector and the dispersion matrix of $\underline{Y}$.

Let $\hat{\underline{\beta}}_d$ be the least square estimator(LSE) of $\underline{\beta}$ in (1.1) using the design $d \in \mathcal{C}$ where $\mathcal{C}$ denotes the class of competing designs. We are interested in characterizing an experimental design d_* which maximises the probability

$$Pr\ [\ \|\hat{\underline{\beta}}_d - \underline{\beta}\| < \epsilon\ ] \qquad \forall \epsilon > 0 \tag{1.2}$$

over the class $\mathcal{C}$ of all competing designs d , where $\|\hat{\underline{\beta}}_d - \underline{\beta}\| = [(\hat{\underline{\beta}}_d - \underline{\beta})'(\hat{\underline{\beta}}_d - \underline{\beta})]^{1/2}$, the Euclidean norm of $\hat{\underline{\beta}}_d - \underline{\beta}$. As this criterion aims at minimizing the distance between the true parameter value and its estimate in a stochastic sense, it is abbreviated as the DS (**D**istance **S**tochastic)- optimality criterion in the literature.

Definition 1.1 *A design $d_* \in \mathcal{C}$ is said to be $DS(\epsilon)$ optimal for the LSE of $\underline{\beta}$ if for a given $\epsilon > 0$, it maximizes the probability $Pr\ [\ \|\hat{\underline{\beta}}_d - \underline{\beta}\| < \epsilon\ ]$ over the class $\mathcal{C}$ of all competing designs d .*

Definition 1.2 *A design $d_* \in \mathcal{C}$ is said to be DS optimal for the LSE of $\underline{\beta}$ if d_* is DS(ϵ) optimal for all $\epsilon > 0$, i.e.*

$$Pr\ [\ \|\hat{\underline{\beta}}_{d_*} - \underline{\beta}\| < \epsilon\ ] \geq Pr\ [\ \|\hat{\underline{\beta}}_{d} - \underline{\beta}\| < \epsilon\ ] \qquad \forall \epsilon > 0 \tag{1.3}$$

and for any competing design $d \in \mathcal{C}$.

Note that the DS optimality criterion is defined via peakedness of the distributions of $\hat{\underline{\beta}}_{d_*}$ and $\hat{\underline{\beta}}_{d}$. According to a definition proposed by Birnbaum (1948), a random variable Y_1 is more peaked about μ_1 than is a random variable Y_2 about μ_2 if

$$Pr\ [\ \|Y_1 - \mu_1\| \leq \epsilon\ ] \geq Pr\ [\ \|Y_2 - \mu_2\| \leq \epsilon\ ] \qquad \forall \epsilon > 0 \tag{1.4}$$

When $\mu_1 = \mu_2 = 0$, we simply say that Y_1 is more peaked than Y_2. Sherman(1955) generalised this definition to the multivariate case.

There is also a similarity between the DS-optimality criterion and Pitman nearness. However, Pitman nearness is a stochastic criterion for comparing estimators while DS optimality criterion is for comparison of designs.

Let us first consider a line fit model through the origin

$$Y_{ij} = \beta x_i + \epsilon_{ij}, \tag{1.5}$$

where

$$E(\epsilon_{ij}) = 0, \ \ V(\epsilon_{ij}) = \sigma^2$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n_i$. The responses Y_{ij}'s are uncorrelated normal random variables. Then $\frac{(\hat{\beta}-\beta)^2}{V(\hat{\beta})} = \chi_1^2$ follows the central χ^2 distribution with 1 d.f and

$$Pr\ [\ |\hat{\beta} - \beta| \leq \epsilon\ ] = Pr\ [\ (\hat{\beta} - \beta)^2 \leq \epsilon^2) = P(\chi_1^2 \leq \frac{\epsilon^2}{\sigma^2} \sum_{i=1}^{n} n_i {x_i}^2) \tag{1.6}$$

Let $\chi = [a, b]$ be the regression range and let $d_p = \{a, b; p\}$ with $0 \leq p \leq 1$ denote a design that assigns the weights p and $1 - p$ to the regression values b and a respectively. If $|b| > |a|$, then the unique maximum of the probability (1.6) is $P(\chi_1^2 \leq \frac{\epsilon^2 n b^2}{\sigma^2})$. Thus $d_1 = \{a, b; 1\}$ is the unique DS optimal design. Similarly, if $|b| < |a|$, then $d_0 = \{a, b; 0\}$ is the unique DS optimal design. Finally, if $\chi^2 = [-a, a]$; then every design $d_p = \{-a, a; p\}$ with any $0 \leq p \leq 1$ is DS optimal.

In fact under model (1.5) with normally distributed errors, the statements
(i) $Pr\ [\ |\hat{\beta}_{d_*} - \beta| \leq \epsilon\ ] \geq Pr\ [\ |\hat{\beta}_{d} - \beta| \leq \epsilon\ ]$ for all $\epsilon > 0$
(ii) $Pr\ [\ |\hat{\beta}_{d_*} - \beta| \leq \epsilon\ ] \geq Pr\ [\ |\hat{\beta}_{d} - \beta| \leq \epsilon\ ]$ for some $\epsilon > 0$
(iii) $V(\hat{\beta}_{d_*}) \leq V(\hat{\beta}_d)$
are equivalent (cf. Stepniak 1989).

We now consider a more general solution where $\underline{\beta}$ may not be estimable. Let $\underline{\eta}_{k\times 1} = L_{k\times m}\underline{\beta}_{m\times 1}$ be the vector of the linear parametric functions of interest to us. We confine only to the class $\mathcal{C}$ of the designs d (i.e. the so called design matrix X_d) under which all the components of $\underline{\eta}$ are estimable. Let the Best Linear Unbiased Estimator (BLUE) of $\underline{\eta}$ using the design d be denoted by $\hat{\underline{\eta}}_d$, where

$$\hat{\underline{\eta}}_d = L\hat{\underline{\beta}}_d,$$

Let $\underline{\hat{\eta}}_{d_1}$ and $\underline{\hat{\eta}}_{d_2}$ be the LSE's of $\underline{\hat{\eta}}$ in (1.1) under the designs d_1 and d_2 respectively. If for a given $\epsilon > 0$,

$$Pr\ [\ \|\underline{\hat{\eta}}_{d_1} - \underline{\eta}\| < \epsilon\ ] \geq Pr\ [\ \|\underline{\hat{\eta}}_{d_2} - \underline{\eta}\| < \epsilon\ ], \tag{1.7}$$

then the design d_1 is at least as good as d_2 with respect to the $DS(\epsilon)$ criterion. A design d_* is said to be $DS(\epsilon)$ optimal for the LSE of $\underline{\eta}$ if for a given $\epsilon > 0$, it maximizes the probability $Pr\ [\ \|\underline{\hat{\eta}}_d - \underline{\eta}\| < \epsilon\ ]$. When d_* is $DS(\epsilon)$ optimal for all $\epsilon > 0$, we say that d_* is DS-optimal. In the particular case $k = 1$, the DS criterion coincides with the $DS(\epsilon)$ criterion for a given $\epsilon > 0$. Note that according to the usual definition of Stochastic ordering for random variables (see Marshall & Olkin 1979, p. 481) $\|\underline{\hat{\eta}}_{d_1} - \underline{\eta}\|$ is stochastically less than $\|\underline{\hat{\eta}}_{d_2} - \underline{\eta}\|$ if (1.7) holds for all $\epsilon > 0$.

For the time being, we assume $\underline{\eta}$ to be nonsingularly estimable i.e rank(L) = k.

Let $\sigma^2\Sigma_d$ be he dispersion matrix and $I(d) = \frac{1}{\sigma^2}\Sigma_d^{-1}$ be the information matrix of $\underline{\hat{\eta}}_d$ under the given model. Let $T\Lambda_d T'$ be the spectral decomposition of $I(d)$ where T is an orthogonal $k \times k$ matrix and $\Lambda_d = Diag(\lambda_{d1}, \ldots, \lambda_{d,k})$ is the diagonal matrix of the eigenvalues of $I(d)$, arranged in decreasing order. Define

$$\underline{Z} = \frac{1}{\sigma}\Lambda_d^{1/2}T'(\underline{\hat{\eta}}_d - \underline{\eta}),$$

so that

$$\underline{Z} \sim N_k(0,\ I_k).$$

Then

$$\begin{aligned} Pr\ [\ \|\underline{\hat{\eta}}_d - \underline{\eta}\| < \epsilon\ ] &= Pr\ [(\underline{\hat{\eta}}_d - \underline{\eta})'(\underline{\hat{\eta}}_d - \underline{\eta}) < \epsilon^2\ ] && (1.8)\\ &= Pr[\ \underline{Z}'\Lambda^{-1}\underline{Z} \leq \epsilon^2/\sigma^2\ ] \\ &= Pr[\ \sum \frac{{Z_i}^2}{\lambda_{di}} \leq \delta^2\ ] && (1.9) \end{aligned}$$

for $\delta = \epsilon/\sigma$.
Thus $\forall\ \delta^2 > 0$, the DS(ϵ) optimality criterion depend on $I(d)$ only through its eigenvalues $\underline{\lambda} = (\lambda_{d1}, \ldots, \lambda_{dk})'$.

We define the criterion function ψ_ϵ or equivalently ψ_δ as

$$\psi_\epsilon(I(d)) = Pr\ [\ \|\underline{\hat{\eta}}_d - \underline{\eta}\|^2 < \epsilon^2\ ] \quad \text{and} \quad \psi_\delta(\underline{\lambda_d}) \quad = Pr[\ \sum \tfrac{{Z_i}^2}{\lambda_{di}} \leq \delta^2\ ] \tag{1.10}$$

It is clear that $\psi_\epsilon(I(d)) = \psi_\delta(\lambda_d)$ for $\delta = \frac{\epsilon}{\sigma} > 0$.

As a function of δ^2 the $DS(\epsilon)$ optimality criterion $\psi_\delta(\underline{\lambda}_d)$ is the cumulative distribution function of $\sum \frac{{Z_i}^2}{\lambda_{di}}$ for every $\underline{\lambda}_d \in \mathbb{R}_+^k$.

2 Properties of the DS optimality criterion

In the below we sometimes skip the suffix d to avoid notational complexity. Let the information matrix of the design d_i be denote by I_i.
It directly follows from (1.10) that

$$\psi_\epsilon(aI) = \psi_{\sqrt{a}\epsilon}(I) \quad \text{and} \quad \psi_\delta(a\underline{\lambda}_d) = \psi_{\sqrt{a}\delta}(\underline{\lambda})$$

for all $a > 0$. It is clear that the optimality criterion ψ_ϵ for given $\epsilon > 0$ induces an ordering among the designs and among the corresponding information matrices of designs. A design d_1 is said to be at least as good as d_2 relative to the criterion ψ_ϵ if $\psi_\epsilon(I_1) \geq \psi_\epsilon(I_2)$ which in other words mean that the informaton matrix I_1 is at least as good as I_2 with respect to ψ_ϵ .

2.1 Isotonicity and Admissibility

The function ψ_ϵ preserves the matrix ordering.

Theorem 2.1 *The DS criterion is isotonic relative to Loewner ordering i.e.*

$$I_1 \geq I_2 > 0 \;\Rightarrow\; \psi_\epsilon(I_1) \;\geq\; \psi_\epsilon(I_2) \quad \textit{for all } \epsilon > 0.$$

Proof :

$$I_1 \;\geq I_2 \;\Rightarrow \lambda_{1i} \;\geq\; \lambda_{2i} \quad \forall i = 1, \ldots, k.$$

Hence the event $E_1 : \{\underline{Z} : \sum \frac{Z_i{}^2}{\lambda_{2i}} \leq \delta^2\}$ implies the event $E_2 : \{\underline{Z} : \sum \frac{Z_i{}^2}{\lambda_{1i}} \leq \delta^2\}$ and consequently $\psi_\epsilon(I_2)$ =Pr $(\sum \frac{Z_i{}^2}{\lambda_{2i}} \leq \delta^2) \leq \Pr(\sum \frac{Z_i{}^2}{\lambda_{1i}}] \leq \delta^2) = \psi_\epsilon(I_1) \;\; \forall \epsilon > 0$.

A reasonable weakest requirement for an information matrix I is that there be no competing information matrix $\tilde{I}$ which is better than I in the Loewner ordering sense. We say that an information matrix I is admissible when every competing moment matrix $\tilde{I}$ with $\tilde{I} \geq I$ is actually equal to I (cf. Pukelsheim 1993, chapter 10). A design d is rated admissible when its information matrix $I(d)$ is admissible. The admissible designs form a competing class (Pukelsheim 1993, Lemma 10.3). Thus every inadmisible information matrix may be improved. If I is inadmissible, then there exists an admissible information matrix $\tilde{I} \neq I$ such that $\tilde{I} \geq I$. Since ψ_ϵ is isotonic relative to Loewner ordering, $DS(\epsilon)$ optimal designs as well as DS- optimal ones can be found in the set of admissible designs.

2.2 Schur Concavity

The notion of majorization proves useful in the study of the function $\psi_\delta(\lambda_d)$. Majorization concerns the diversity of the components of a vector. (cf. Marshall and Olkin 1979, p.7).

Let $\underline{a} = (a_1, \ldots, a_k)'$, and $\underline{b} = (b_1, \ldots, b_k)'$ be two $k \times 1$ vectors and $a_{(1)} \leq \ldots \leq a_{(k)}$, $b_{(1)} \leq \ldots \leq b_{(k)}$ be the ordered components.

Definition 2.2 *: For $\underline{a}, \underline{b} \in I\!R^k, \underline{a}$ is said to majorize $\underline{b}$, written* $\mathbf{a} \succ \mathbf{b}$ *if*

$$\left.\begin{array}{lcl} \sum_{i=1}^p a_{(i)} & \leq & \sum_{i=1}^p b_{(i)} \quad p = 1, \ldots, k-1, \\ \sum_{i=1}^k a_{(i)} & = & \sum_{i=1}^k b_{(i)}. \end{array}\right\} \tag{2.1}$$

Definition 2.3 *: For $\underline{a}, \underline{b} \in I\!R^k, \underline{a}$ is said to weakly supermajorize $\underline{b}$, written $\underline{a} \;{}^w\!\succ \underline{b}$ if*

$$\sum_{i=1}^p a_{(i)} \;\leq\; \sum_{i=1}^p b_{(i)} \quad p = 1, \ldots, k. \tag{2.2}$$

Majorization provides a partial ordering on $I\!R^k$. The order $\underline{a} \succ \underline{b}$ implies that the elements of $\underline{a}$ are more diverse than the elements of $\underline{b}$. Then for example, $\underline{a} \succ \bar{\underline{a}} = (\bar{a}, \ldots, \bar{a})$ for all $\underline{a}, \in I\!R^k$, where $\bar{a} = \frac{1}{k}\sum_{i=1}^k a_i$. Functions which reverse the ordering of majorization are said to be Schur Concave (cf. Marshall and Olkin 1979, p. 54).

Definition 2.4 *: A function $f(\underline{x}) : I\!R^k \to I\!R$ is said to be a Schur Concave function if for $\underline{x}, \underline{y} \in I\!R^k$ the relation $\underline{x} \succ \underline{y}$ implies $f(\underline{x}) \leq f(\underline{y})$. Thus the value of $f(\underline{x})$ becomes greater when the components of $\underline{x}$ become less diverse.*

Next we examine Schur Concavity of $DS(\epsilon)$ criterion. For that we quote below Proposition 7.4.2 of Tong 1990, p.163.

Proposition 1 *For given* $\underline{a} = (a_1, \ldots, a_k)', a_i > 0$ *consider an ellipsoid defined by*

$$A_2(\underline{a}) = \{\underline{x} : \underline{x} \in I\!R^k, , \sum (\frac{x_i}{a_i})^2 \leq \lambda\}, \quad \lambda > 0, \text{ fixed.}$$

If the density function $f(\underline{x})$ is a Schur concave function of $\underline{x}$ then $P[\underline{x} \in A_2(a)]$ is a Schur Concave function of $\underline{a}^2 = (a_1^2, \ldots, a_n^2)$.

Theorem 2.5 *The* $DS(\epsilon)$ *criterion is Schur Concave for all* $\epsilon > 0$.

Proof: The components of $\underline{Z}$ being i.i.d Normal with mean zero and variance σ^2, the joint density function $f(\underline{z})$ is Schur Concave (Tong 1990, Theorem 4.4.1). Then by the above Proposition $\psi_\delta(\underline{\lambda}_d)$ is a Schur Concave function of $\underline{\lambda}_d = (\lambda_{d_1}, \ldots, \lambda_{d_k})'$ for all $\delta > 0$.

Corollary 2.6 $\psi_\delta(\underline{\lambda}_d) \leq \psi_\delta(\bar{\underline{\lambda}}_d)$ *holds for all* $\underline{\lambda}_d \in I\!R_+^k$ *and all* $\delta > 0$ *where* $\bar{\underline{\lambda}}_d = (\bar{\lambda}_d, \ldots, \bar{\lambda}_d)$ *where* $\bar{\lambda}_d = \frac{1}{k}\sum_{i=1}^k \lambda_{di}$.

Corollary 2.7 *Let* $\underline{\lambda}_1$ *and* $\underline{\lambda}_2$ *denote the column vectors where components are the eigen values of* I_1 *and* I_2 *respectively arranged in decreasing order. If* I *is the information matrix with a vector of eigenvalues* $(1-\alpha)\underline{\lambda}_1 + \alpha\underline{\lambda}_2$ *then,* $\psi_\alpha((1-\alpha)I_1 + \alpha I_2) \geq \psi_\epsilon(I)$ *for all* $\alpha \in [0,1]$ *and all* $\epsilon > 0$.

Proof: Let $\underline{\lambda}[(1-\alpha)I_1 + \alpha I_2]$ denote the column vector of eigenvalues of $(1-\alpha)I_1 + \alpha I_2$ arranged in decreasing order. Since by Theorem G.1 (Marshall & Olkin 1979, p. 241) $(1-\alpha)\underline{\lambda}_1 + \alpha\underline{\lambda}_2 \succ \underline{\lambda}((1-\alpha)I_1 + \alpha I_2)$, the result follows.

Theorem 2.8 $\psi_\delta(\lambda)$ *is Schur concave function of* $(\log\lambda_1, \ldots, \log\lambda_k)'$ *for all* $\delta > 0$.

Corollary 2.9 $\psi_\delta(\underline{\lambda}_d) \leq \psi_\delta(\tilde{\underline{\lambda}}_d)$ *holds for all* $\underline{\lambda}_d \in I\!R_+^k$ *and all* $\delta > 0$ *where* $\tilde{\underline{\lambda}}_d = (\tilde{\lambda}, \ldots, \tilde{\lambda})$ *and* $\tilde{\lambda} = \prod_{i=1}^k \lambda_i^{1/k}$.

Theorem 2.10 *If* $\underline{\lambda}_d \overset{w}{\succ} \tilde{\underline{\lambda}}_d$ *then* $\psi_\delta(\underline{\lambda}_d) \leq \psi_\delta(\tilde{\underline{\lambda}}_d)$.

Proof: If $\underline{\lambda}_d \overset{w}{\succ} \tilde{\underline{\lambda}}_d$ then there exists (cf. Marshall and Olkin p.11) a vector $\underline{\lambda}_{d_0}$ such that

$$\underline{\lambda}_{d_0} \geq \underline{\lambda}_d \text{ and } \underline{\lambda}_{d_0} \succ \tilde{\underline{\lambda}}_d$$

Thus

$$Pr(\sum \frac{{Z_i}^2}{\lambda_{di}} \leq \delta^2) \leq Pr(\sum \frac{{Z_i}^2}{\lambda_{d_0 i}} \leq \delta^2) \leq Pr(\sum \frac{{Z_i}^2}{\tilde{\lambda}_{di}} \leq \delta^2) \quad \forall\ \delta > 0.$$

The first inequality follows from implication of events as discussed earlier in the proof of Theorem 2.1. The last inequality now follows from Theorem 2.5. Theorem 2.8 is a version of Okamoto Lemma (1960). There are other generalisations of this useful result. We reproduce them below from Sinha(1970) and Liski et.al (1998). (We omit the proofs altoghether). To describe the results along similar fashion, Okamoto's Lemma is stated in a conventional form. Below Z_i 's are i.i.d $N(0,1)$. Note further that in Lemma 2.11 and generalisations thereafter, $\frac{1}{\lambda_{di}}$'s are replaced by μ_{di}'s (omitting suffix d) and used in the numerator instead of the denominator.

Lemma 2.11 *(Okamoto)*

$$Pr(\sum_{i=1}^{k} \mu_i {Z_i}^2 \leq \epsilon^2) \leq Pr(\mu {\chi_k}^2 \leq \epsilon^2)$$

where ${\chi_k}^2$ *refers to a central* χ^2 *variate with* k *degrees of freedom and* $\mu = (\prod \mu_i)^{1/k}$.

Generalisation 1:

$$Pr(\sum_{i=1}^{k} \mu_i {Z_i}^2 \leq \epsilon^2) \quad \leq \quad Pr(\mu' {\chi_{k'}}^2 + \mu'' {\chi_{k''}}^2 \leq \epsilon^2)$$

where

$$k = k' + k'', \;\; \mu' = (\prod_{i=1}^{k'} \mu_i)^{1/k} \text{ and } \mu'' = (\prod_{i=k'+1}^{k} \mu_i)^{1/k''}$$

Generalisation 2:

$$P(\mu_1 Z_1^2 + \mu_2 Z_2^2 \leq \epsilon^2) \leq P(\gamma_1 Z_1^2 + \gamma_2 Z_2^2 \leq \epsilon^2)$$

provided

$$\mu_1 \mu_2 \geq \gamma_1 \gamma_2 \quad \text{and} \quad \max\{\mu_1, \mu_2\} \geq \max\{\gamma_1, \gamma_2\}.$$

Generalisation 3:

$$P(\mu_1^* \chi_{v_1}^2 + \mu_2^* \chi_{v_2}^2 \leq \epsilon^2) \leq P(\mu_1^{**} \chi_{v_1}^2 + \mu_2^{**} \chi_{v_2}^2 \leq \epsilon^2)$$

provided

$$(\mu_1^*)^{v_1} (\mu_2^*)^{v_2} > (\mu_1^{**})^{v_1} (\mu_2^{**})^{v_2}$$

and

$$\max\{\mu_1^*, \mu_2^*\} > \max\{\mu_1^{**}, \mu_2^{**}\} > \min\{\mu_1^{**}, \mu_2^{**}\} > \min\{\mu_1^*, \mu_2^*\}.$$

2.3 Concavity

Concavity is often regarded as a compelling property of an optimality criterion (cf. Pukelsheim 1993, p. 115), but $DS(\epsilon)$ optimality criterion is not, in general, concave.

Theorem 2.12 *The function* $\psi_\delta(\underline{\lambda})$ *is concave on* $\mathbb{R}_+^k$ *for every fixed* $\delta > 0$ *if and only if* $k \geq 2$.

Remark 1 In the below, while characterising DS optimal Designs under various settings, we use formulations (1.9) or the one stated in Lemma 2.11 depending on the tool employed to arrive at the optimal design.

3 Discrete DS optimal Designs

In this section we derive DS optimal designs for the LSE of $\underline{\eta}$ for different choices of the parametric vector of interest.

3.1 Mean vector and the set of all elementary contrasts in a CRD Model

Sinha(1970) introduced DS-optimality criterion for optimal allocation of observations with a given total in a CRD model:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, \ldots, v; \quad j = 1, \ldots, n_i, \tag{3.1}$$

where $\underline{\tau} = (\tau_1, \ldots, \tau_v)'$ is the vector of treatment effects. The parametric vector of interest is the mean vector $\underline{\eta} = (\mu + \tau_1, \ldots, \mu + \tau_v)'$. For a design $d \in \mathcal{C}$, let the ith treatment be allocated n_{di} times, $n_{di} \geq 1, \ \ 1 \leq i \leq v, \ \sum_{i=1}^{v} n_{di} = n$. Referring to (1.9) we note that $\lambda_{di} = n_{di}$ for each i. When n is divisible by v it turns out that a 'symmetrical allocation' with $n_{di} = n/v \ \ \forall i$, is uniquely DS- optimal. The general case, when n is not divisible by v is quite involved and Sinha (1978) established the partial result when v is even and $n = \frac{v}{2}(mod v)$. Very recently this problem is resolved by Liski et. al.(1998). A most symmetrical allocation d^* with $|n_{d^*i} - n_{d^*j}| \leq 1, \ \ \forall i \neq j$, turns out to be DS- optimal as is expected. We outline the proof below.

For any design $d(\neq d^*) \in \mathcal{C}$ there exists at least a pair of treatments (i, j) such that $n_{di} - n_{dj} > 1$. It is easy to check that

$$\begin{aligned} \underline{n}_d &= (n_{d1}, \ldots, n_{d(i-1)}, n_{di}, n_{d(i+1)}, \ldots, n_{d(j-1)}, n_{dj}, n_{d(j+1)} \ldots, n_{dv}) \\ &\succ \underline{n}_{d0} = (n_{d1}, \ldots, n_{d(i-1)}, n_{di} - 1, n_{d(i+1)}, \ldots, n_{d(j-1)}, n_{dj} + 1, n_{d(j+1)} \ldots, n_{dv}). \end{aligned}$$

Then using Theorem 2.5

$$Pr(\sum \frac{{Z_i}^2}{n_{di}} \leq \delta^2) \quad \leq \quad Pr(\sum \frac{{Z_i}^2}{n_{d_0 i}} \leq \delta^2) \tag{3.2}$$

By repeated application of (3.2) it can be shown that whenever allocation numbers for a pair of treatments differ by more than 1, successively reducing thier difference by 2, but keeping the total fixed, a better design can be obtained and finally, a most symmetrical allocation with $|n_{di} - n_{dj}| \leq 1, \ \ \forall i \neq j$, turns out to be DS- optimal .

Recently, SahaRay and Bhandari(2001) examined the nature of DS -optimal designs in a CRD set up for inference on the set of all elementary contrasts of the form $\tau_i - \tau_j, \ i < j$ *viz.*

$$\underline{\eta}' = (\tau_1 - \tau_2, \ldots, \tau_1 - \tau_v, \ldots, \tau_{v-1} - \tau_v).$$

Writing $\underline{\eta} = L\underline{\tau}$, we note that $R(L) = v - 1$. Thus it corresponds to a singularly estimable full rank problem. Here weak majorization plays an important role to establish DS-optimality of symmetrical or most symmetrical allocations depending on the divisibility of the total number of observations by the number of treatments or not. We sketch the proof below.

Let P be a $(v-1) \times v$ submatrix of an orthogonal $v \times v$ matrix such that

$$P_{v-1 \times v} \, P'_{v \times v-1} = I_{v-1}, \ \ P'P = (I - J/v), \tag{3.3}$$

$$\begin{aligned} D &= Diag(1/n_{d1}, \ldots, 1/n_{dv}), \\ \text{and} \quad D^{1/2} &= Diag(1/\sqrt{n_{d1}}, \ldots, 1/\sqrt{n_{dv}}). \end{aligned} \tag{3.4}$$

Writing $\hat{\underline{\eta}}_d = L\hat{\underline{\tau}}_d$, where $\hat{\underline{\tau}}_d = (\bar{y}_{1.}, \ldots, \bar{y}_{v.})$, we have from (1.7)

$$P_\epsilon \quad = \quad Pr[(\hat{\underline{\eta}}_d - \underline{\eta})'(\hat{\underline{\eta}}_d - \underline{\eta}) \leq \epsilon^2 \ ]$$

$$\begin{aligned} &= Pr[(\hat{\underline{\tau}}_d - \underline{\tau})' L' L (\hat{\underline{\tau}}_d - \underline{\tau}) \leq \epsilon^2] \\ &= Pr[(\hat{\underline{\tau}}_d - \underline{\tau})' (I - J/v)(\hat{\underline{\tau}}_d - \underline{\tau}) \leq \epsilon^2/v] \\ &= Pr[(\hat{\underline{\tau}}_d - \underline{\tau})' P' P (\hat{\underline{\tau}}_d - \underline{\tau}) \leq \epsilon^2/v] \\ &= Pr[(\underline{\xi}_d{}' \underline{\xi}_d) \leq \epsilon^2/v], \end{aligned} \tag{3.5}$$

where $\underline{\xi}_d = P(\hat{\underline{\tau}}_d - \underline{\tau}) \sim N_{v-1}(0, \sigma^2 \Sigma_d)$ with $\Sigma_d = PDP'$. For this specific problem, σ^2 does not play any role in the determination of DS optimal designs. So we assume $\sigma^2 = 1$ in particular.

It is not hard to verify that for any $d \in \mathcal{C}$, $\Sigma_d = PDP'$ is nonsingular. Let $\underline{\mu}_d = (\mu_{d1}, \ldots, \mu_{d,(v-1)})'$ denote the vector of ordered eigenvalues of PDP' where $\mu_{d1} \leq \mu_{d2} \leq \cdots \leq \mu_{d,(v-1)}$. Let $\tilde{T}\tilde{\Lambda}\tilde{T}' = \Sigma_d$ be the spectral decomposition of Σ_d , the dispersion matrix of ξ, where $\tilde{T}$ is an orthogonal $(v-1) \times (v-1)$ matrix and $\tilde{\Lambda} = Diag(\mu_{d1}, \ldots, \mu_{d,(v-1)})$ is the diagonal matrix of the eigenvalues of the dispersion matrix Σ_d, in other words of PDP'. Then using cannonical reduction as discussed in (1.9), we get

$$Pr\ [\ \|\hat{\underline{\eta}}_d - \underline{\eta}\| < \epsilon\ ] \quad = \quad Pr[\ \sum \mu_{di} Z_i{}^2 \leq \delta^2\ ] \tag{3.6}$$

for $\delta^2 = \epsilon^2/v$. (Note that here μ_{di}'s are eigenvalues of the corresponding dispersion matrix whereas in (1.9) λ_{di}'s are the eigenvalues of the corresponding information matrix.)

Thus $\forall\ \delta^2 > 0$, the DS(ϵ) optimality criterion $Pr[\ \|\hat{\underline{\eta}}_d - \underline{\eta}\| < \epsilon\ ]$ depend on the design d with allocation numbers $n_{di}, i = 1, \ldots, v$ only through the eigenvalues $\mu_{d1}, \ldots, \mu_{d,(v-1)}$ of the matrix (PDP'). It is worthwhile to note that μ_{di}'s are very nontrivial functions of n_{di}'s, unlike the problem of estimation of mean vector. Furthermore, μ_{di}'s do not depend on the choice of the P matrix where $P'P = I - J/v$ and $P'P = I_{v-1}$ as is clear , by defining $A = PD^{1/2}$ and $B = A'$ and noting that the positive eigenvalues of $AB = PDP'$ and $BA = D^{1/2}P'PD^{1/2} = D^{1/2}(I - J/v)D^{1/2}$ are equal.

Remark 2 Instead of $\underline{\eta}'$ if we had considered the set of all contrasts of the form $\tau_i - \tau_j, i \neq j$, the problem would have remained the same except that δ^2 in (3.6) would change to a scalar multiple of it, viz $\delta^2/2$.

Whenever n is divisible by v, using Okamoto's Lemma and well known inequality between the Arithmetic Mean(A.M) and the Geometric Mean(G.M) of a set of positive quantities, DS optimality of a symmetric design $d_*$with $n_{d_*i} = n/v$ can be shown through the following steps. First we note that, in the present context of estimation of the set of all elementary contrasts,

$$|\ PDP'\ | = \prod_{i=1}^{v-1} \mu_{di} = \frac{n}{v \prod_{i=1}^{v} n_{di}} \tag{3.7}$$

So

$$\begin{aligned} Pr\ (\ \sum_{i=1}^{v-1} \mu_{di} Z_i{}^2 \leq \delta^2\ \ ) \quad &\leq \quad Pr\ [(n/(v \prod_{i=1}^{v} n_{di}))^{1/v-1}\ \ \chi^2{}_{v-1} \leq \delta^2] \\ &\leq \quad Pr\ [(n/(v \prod_{i=1}^{v} n_{d_*i}))^{1/v-1}\ \ \chi^2{}_{v-1} \leq \delta^2] \\ &\qquad \text{(follows by implication of events)} \end{aligned}$$

$$
\begin{aligned}
&= Pr\,[\ (\prod_{i=1}^{v-1} \mu_{d_* i})^{\frac{1}{(v-1)}}\ \ \chi^2{}_{v-1}\ \ \leq\ \ \delta^2] \\
&= Pr\,(\ \sum_{i=1}^{v-1} \mu_{d_* i} Z_i{}^2 \leq \delta^2\ \ )\qquad (\text{ as } \mu_{d_* i} \text{ 's are all equal})\ .
\end{aligned}
$$

Thus whenever n is divisible by v, the design d_* with symmetrical allocation turns out to be DS- optimal.

The case when n is not divisible by v is dealt below. From now onwards we assume that $n = vk + t, t \geq 1$, where $k = [n/v]$ denotes the greatest integer less than or equal to n/v. Let $d_* \in \mathcal{C}$ be a design with $\underline{n}_{d_*} = (k, \ldots, k, k+1, \ldots, k+1)', k$ occuring $v - t$ times and corresponding

$$
D_* \ = \ Diag(\underbrace{1/k, \ldots, 1/k}_{v-t\ \ times}, \underbrace{1/(k+1), \ldots, 1/(k+1)}_{t\ \ times}).
$$

In order to establish d_* to be DS-optimal, in view of (3.6) and Theorem 2.10, it suffices to show that

$$
\underline{\mu}_d{}^{-1} \ \ {}^{w} \succ \underline{\mu}_{d_*}{}^{-1}
$$

where $\underline{\mu}_d{}^{-1}$ and $\underline{\mu}_{d_*}{}^{-1}$ denote respectively the vectors of eigenvalues of $(PDP')^{-1}$ and $(PD_*P')^{-1}$, the information matrices.

Step I:

We will establish that

$$
\underline{\mu}(D^{-1}(A + \epsilon I) \succ \underline{\mu}(D_*{}^{-1}(A + \epsilon I)), \quad \forall\ \epsilon^* \in I\!R, \epsilon^* \neq 0 \text{ and } -1.
$$

We first note that, for any $\epsilon^* \in I\!R, \epsilon^* \neq 0$ and -1, $A + \epsilon^* I$ is a nonsingular matrix and the eigenvalues of $D^{-1/2}(A + \epsilon^* I)D^{-1/2}$ and $D^{-1}(A + \epsilon^* I)$ are identical. It is clear that for any design $d(\neq d_*) \in \mathcal{C}$, there exists at least one pair of treatment symbols i' and j' such that $(n_{di'} - n_{dj'}) \geq 2$ and $\sum n_{di} = n$. We permute $i'$and j' treatment symbols, keeping others fixed and obtain $\tilde{d} \in \mathcal{C}$ as

$$
\begin{aligned}
n_{\tilde{d}i} &= n_{di} \quad \forall\ i \neq i', j' \\
n_{\tilde{d}i'} &= n_{dj'} \\
n_{\tilde{d}j'} &= n_{di'}
\end{aligned}
$$

and hence $D^{-1} = QD^{-1}Q'$, where Q represents the corresponding permutation matrix.

In view of the relation $Q'Q = QQ' = I,$ and $Q(A + \epsilon^* I)Q' = A + \epsilon^* I,$

$$
\tilde{D}^{-1}(A + \epsilon I) = QD^{-1}Q'(A + \epsilon I) = QD^{-1}Q'Q(A + \epsilon I)Q' = QD^{-1}(A + \epsilon I)Q'. \qquad (3.8)
$$

Hence

$$
\underline{\mu}(\tilde{D}^{-1}(A + \epsilon^* I)) = \underline{\mu}(D^{-1}(A + \epsilon^* I))
$$

Now it is easy to see that for some $0 < \alpha < 1, (n_{di'} - 1, n_{dj'} + 1)$ can be represented as a convex combination of $(n_{di'}, n_{dj'})$ and $(n_{dj'}, n_{di'})$. Choosing this α, it follows that

$$
\begin{aligned}
\underline{\mu}(D^{-1}(A + \epsilon I)) &= \alpha\underline{\mu}(D^{-1}(A + \epsilon I)) + (1 - \alpha)\underline{\mu}(QD^{-1}(A + \epsilon I)Q') \\
&= \alpha\underline{\mu}(D^{-1}(A + \epsilon I)) + (1 - \alpha)\underline{\mu}(\tilde{D}^{-1}(A + \epsilon I))
\end{aligned}
$$

$$\begin{aligned} &\succ \underline{\mu}(\alpha D^{-1}(A+\epsilon I)+(1-\alpha)\tilde{D}^{-1}(A+\epsilon I)) && (3.9)\\ &\quad \text{(cf. proof of Corollary 2.7)} \\ &= \underline{\mu}((\alpha D^{-1}+(1-\alpha)\tilde{D}^{-1})(A+\epsilon I)) \\ &= \underline{\mu}({D_0}^{-1}(A+\epsilon I)), && (3.10) \end{aligned}$$

where

$$\begin{aligned} {D_0}^{-1} &= \alpha D^{-1}+(1-\alpha)\tilde{D}^{-1} \\ &= Diag(n_{d1},\ldots,n_{d(i'-1)},n_{di'}-1,n_{d(i'+1)},\ldots,n_{d(j'-1)},n_{dj'}+1,n_{d(j'+1)},\ldots,n_{dv}). \end{aligned}$$

Thus when the pair of allocation numbers $(n_{di'},n_{dj'})$ is transferred to a pair $(n_{di'}-1,n_{dj'}+1)$, reducing their mutual difference by two, but keeping the total fixed, we get

$$\underline{\mu}(D^{-1}(A+\epsilon^* I)) \succ \underline{\mu}({D_0}^{-1}(A+\epsilon^* I)),\ \ \forall \epsilon^* \in I\!R, \epsilon^* \neq 0 \text{and } 1.$$

Note that starting from D^{-1} successive averaging by taking convex combination of any two co-ordinates of $\underline{n}_d=(n_{d1},\ldots n_{dv})'$ in the above sense, while keeping the rest of the co-ordinates fixed, we will eventually get

$${D_*}^{-1}=diag(k,\ldots,k,k+1,\ldots,k+1)$$

and similar successive steps of majorization will yield

$$\underline{\mu}(D^{-1}(A+\epsilon I)) \succ \underline{\mu}({D_0}^{-1}(A+\epsilon I)) \succ \cdots \succ \underline{\mu}({D_*}^{-1}(A+\epsilon I)). \tag{3.11}$$

Step II:
We will now establish that

$$\underline{\mu_d}^{-1} \ \ {}^{w}\!\succ \underline{\mu_{d_*}}^{-1}$$

where $\underline{\mu_d}^{-1}$ and $\underline{\mu_{d_*}}^{-1}$ denote respectively the vectors of eigenvalues of $(PDP')^{-1}$ and $(PD_*P')^{-1}$. In Step I, (3.10) can be rewritten alternatively as

$$\underline{\mu}^{-1}\ (D(A+\epsilon^* I)^{-1}) \succ \underline{\mu}^{-1}(D_*(A+\epsilon^* I)^{-1}). \tag{3.12}$$

As $A=I-J/v$,

$$(A+\epsilon I)^{-1}=\frac{1}{1+\epsilon}(A+\frac{1+\epsilon}{\epsilon}J/v), \qquad \epsilon \in I\!R, \epsilon \neq 0 \text{ and } -1. \tag{3.13}$$

Call $\frac{1+\epsilon}{\epsilon}=\theta$. Thus for any $\theta>0$, (3.4) can be rewritten as

$$\underline{\mu}^{-1}(D(A+\theta J/v)) \succ \underline{\mu}^{-1}(D_*(A+\theta J/v)). \tag{3.14}$$

Now use the result that

$$\begin{aligned} \lim_{\theta\to 0}\mu_1(D(A+\theta J/v)) &= 0, \\ \text{and} \quad \lim_{\theta\to 0}\mu_i(D(A+\theta J/v)) &= \lambda_i(DA), \qquad i\neq 1. \end{aligned}$$

and note that $\mu_i(DA)=\mu_{i-1}(PDP')$, $\forall\ i=2,\ldots,v$. Referring to the Def 2.2, it can be seen that (3.13) will yield v inequalities of which the first $(v-1)$ after imposing limit will yield

$$\underline{\mu_d}^{-1}=(1/\mu_{d1},\ldots,1/\mu_{d(v-1)}) \ \ {}^{w}\!\succ (1/\mu_{d_*1},\ldots,1/\mu_{d_*(v-1)})=\underline{\mu_{d_*}}^{-1} \tag{3.15}$$

and hence the result.

It can be easily seen that d_* as well as any permutation of d_* is uniquely DS optimal.

3.2 A full set of orthonormal contrasts in the block design model

Now we consider a block design model. We explain the connection between DS optimal designs and universally optimal designs in the sense of Kiefer(1975). For the inferential problem of a full set of orthonormal treatment contrasts i.e $\underline{\eta}_d = P\underline{\tau}_d$ the DS-optimality criterion boils down to maximize

$$Pr\ \ [\ \|\hat{\underline{\eta}}_d - \underline{\eta}\| < \epsilon\ \ ]\ \ = Pr[\underline{\xi}'_d\underline{\xi}_d \leq \epsilon^2]$$

where $\underline{\xi}_d = P(\hat{\underline{\tau}}_d - \underline{\tau}) \sim N_{v-1}(0, \Sigma_d)$ with $\Sigma_d = (PC_dP')^{-1}$ and C_d is the usual $C-$matrix under the design d. Now application of Corollary 2.6 and arguments used through implication of events yield that if there exists a design d^* for which the C_{d^*} is completely symmetric and $tr(C_{d^*}) = \max_{d\in\mathcal{C}} tr(C_d)$ then the design d^* is DS optimal. This covers BIBDs, BBDs, Yoden square designs, in particular. (vide Kiefer 1975).

3.3 Control vs. test treatment comparisons in a CRD model

Optimality studies in the context of treatment vs.control comparisons has received considerable attention of the researchers in the recent years. See Majumder (1996)and references therein. Dealing with a CRD model and block design set up, Mandal, Shah Sinha (2000) characterizes DS optimal designs for inference on the vector of parametric contrasts involving a set of treatments and one control, i.e

$$\begin{aligned} \underline{\eta} &= (\tau_0 - \tau_1, \tau_0 - \tau_2, \ldots, \tau_0 - \tau_v)' \\ &= L\underline{\tau} \end{aligned} \tag{3.16}$$

where $L = (1|-I), \underline{\tau} = (\tau_0, \tau_1, \ldots, \tau_v)'$ and τ_0 refers to the effect of the control treatment and $\tau_1, \ldots, \tau_v$ refer to the test treatments.

Let under a CRD d in the competing class $\mathcal{C}$, $n_{d0}, , n_{d1}, \ldots, n_{dv}$ denote allocation numbers subject to a total of n observations. Setting $p_{di} = \frac{n_{di}}{n}, i = 0, 1, 2, \ldots, v$ so that $\sum_{i=0}^{v} p_{di} = 1, d \in \mathcal{C}$, the problem is to seek optimal values of p_{di} 's so as to maximize $P_\epsilon = Pr\ \ [\ \|\hat{\underline{\eta}}_d - \underline{\eta}\| < \epsilon\ \ ]\ \ \forall\epsilon > 0$.

It is not hard to verfy that variance- covariance matrix of $\hat{\underline{\eta}}_d$ turns out to be

$$D(\hat{\underline{\eta}}_d) \ = \ \sigma^2[D + \frac{J_v}{n_{d0}}]$$

where $D = diag(\frac{1}{n_{d1}}, \ldots, \frac{1}{n_{dv}})$ and $J_v = ((1))_{v\times v}$, the matrix of all ones. Let $\mu_{d1}, \ldots, \mu_{dv}$ denote the eigenvalues of $\sigma^2 D(\hat{\underline{\eta}}_d)$. Then

$$\prod_{i=1}^{v} \mu_{di} = |D + \frac{J_v}{n_{d0}}| \ = \ \frac{1}{\prod_{i=1}^{v} n_{di}}(\frac{n}{n_{d0}}) \geq (\frac{v}{n - n_{d0}})^v \frac{n}{n_{d0}} \ \ (A.M \geq G.M) \tag{3.17}$$

Further

$$\begin{aligned} \mu_{\max} &\geq \frac{\underline{1}'_v(D + \frac{J_v}{n_{d0}})\underline{1}_v}{\underline{1}'\underline{1}} = \frac{\sum_{i=1}^{v} \frac{1}{n_{di}} + \frac{v^2}{n_{d0}}}{v} \\ &\geq \frac{v^2}{n - n_{d0}} + \frac{v^2}{n_{d0}} \ \ (\text{using} A.M \geq H.M) \\ &= \frac{nv}{n_{d0}(n - n_{d0})} \end{aligned} \tag{3.18}$$

In (3.16) and (3.17) "= " holds whenever $n_{d1} = n_{d2} = \cdots = n_{dv} = (n - n_{d0})/v$ for every n_{d0}. Referring to the probability inequality in Generalisation 3, we set

$$v_1 = 1, \;\; v_2 = v - 1, \;\; \mu_1^* = \mu_{\max}, \;\; \mu_2^* = (\frac{\prod \mu_{di}}{\mu_{\max}})^{1/(v-1)}$$

$$\mu_1^{**} = \frac{nv}{n_{d0}(n - n_{d0})} \text{ and } \mu_2^{**} = \frac{v}{(n - n_{d0})}$$

Note that

$$\mu_1^* = \max(\mu_1^*, \mu_2^*) \quad \geq \quad \max(\mu_1^{**}, \mu_2^{**}) = \mu_1^{**} \tag{3.19}$$

Now we deal with the two cases $\mu_2^* \geq \mu_2^{**}$ and $\mu_2^* < \mu_2^{**}$ separately.
Suppose first that $\mu_2^* \geq \mu_2^{**}$. Then an application of Generalisation 1 gives

$$\begin{aligned} Pr(\sum_{i=1}^{k} \mu_i {Z_i}^2 \leq \epsilon^2) &\leq Pr(\mu_{\max} {\chi_1}^2 + \mu_2^* {\chi_{v-1}}^2 \leq \epsilon^2) \\ &\leq Pr(\mu_1^{**} {\chi_1}^2 + \mu_2^{**} {\chi_{v-1}}^2 \leq \epsilon^2) \end{aligned} \tag{3.20}$$

The last equation follows by implication of events.
If on the other hand, $\mu_2^* < \mu_2^{**}$ conditions reqiured in Generalisation 3 are satisfied and hence the above result follows from application of Generalisation 1 followed by that of Generalisation 3.

Thus, for every fixed $n_{d0} = n_0$ say, the allocation $(n_0, \frac{n-n_0}{v}, \ldots, \frac{n-n_0}{v})$ is uniformly beter than the allocation $(n_0, n_{d1}, \ldots, n_{dv})$ for all choices of n_{di}'s subject to $\sum n_{di} = n - n_0$. Thus for this problem,

$$\begin{aligned} Pr(\sum_{i=1}^{k} \mu_i {Z_i}^2 \leq \epsilon^2) &\leq Pr(\frac{nv}{n_0(n - n_0)} \chi_1^2 + \frac{v}{n - n_0} \chi_{v-1}^2 \leq \epsilon^2) \\ &= Pr({\chi_1}^2 + p_0 {\chi_{v-1}}^2 \leq p_0(1 - p_0)\epsilon^2) \\ &= P_\epsilon(say) \end{aligned}$$

According to the approximate design theory, p_0 is chosen optimally to maximize P_ϵ for every $\epsilon > 0$. For several choices of ϵ and v using 10,000 simulation Mandal and Shah and Sinha (2000) observe that no single p_0 is optimal for all values of ϵ. Further the DS optimal design d^* are almost the same as A- optimal allocations . Since the optimal designs are very much ϵ dependent it would be reasonable to choose a value of p_0 which maximizes P_ϵ averaged with respect to an appropriate weight function for ϵ. Assuming the p.d.f of ϵ^2 as that of χ_2^2 Mandal et al. (2000) set a value for the weighted coverage probability $\bar{P}$ and search for avalue of p_0 which will maximise v/n and thereby minimize the value of n required to attain $\bar{P}$. Results are tabulated in the paper for choices of $\bar{P} = .9$ and .95.

3.4 Control-treatment comparisons in treatment- connected designs

Referring to (3.15) for a treatment connected design d with the usual C- matrix denoted by C_d

$$D(\underline{\hat{\eta}}_d) = (LC_d^+ L')\sigma^2 = \sigma^2 \Sigma_d$$

where C_d^+ denotes the Moore-Penrose inverse of C_d.
We note that

$$\mu_{\max}(\Sigma_d) \geq \frac{\underline{x}'(LC_d^+L')\underline{x}}{\underline{x}'\underline{x}} \quad \text{for any } \underline{x} \neq 0$$

In particular, taking $\underline{x} = \underline{1}$ we get

$$\mu_{\max}(\Sigma_d) \geq C_{d00}^+ \frac{(v+1)^2}{v}$$

It is not difficult to verify that

$$C_{d00}^+ C_{d00} \geq (\frac{v}{v+1})^2$$

so that

$$\mu_{\max}(\Sigma_d) \geq \frac{(v+1)^2}{v}(\frac{v}{v+1})^2 \frac{1}{c_{d00}} = \frac{v}{C_{d00}} \tag{3.21}$$

Let d^* be a design for which C_{d^*} has the structure

$$C_{d^*} = \left[\begin{array}{c|cccc} \text{a} & \text{-a/v} & \dots & \dots & \text{-a/v} \\ \hline \text{-a/v} & \text{b/v} & \text{-c/v} & \dots & \text{-c/v} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{-a/v} & \text{-c/v} & \text{-c/v} & \dots & \text{b/v} \end{array}\right] \tag{3.22}$$

where $a = C_{d00}$ and $b = tr(C_d) - C_{d00}$.
It is easy to verify that the eigenvalues of Σ_{d^*} are $\frac{v}{a}$ with multiplicity 1 and $\frac{v(v-1)}{vb-a}$ with multiplicity $(v-1)$. Whenever $b \geq a, \mu_{\max}(\Sigma_{d^*}) = \frac{v}{a}$, and in view of (3.20) $\mu_{\max}(\Sigma_d) \geq \mu_{\max}(\Sigma_{d^*})$. Mandal et.al.(2002) arguing exactly as in the case of CRD indicated in previous subsection claim that d^* improves over d uniformly in $\epsilon > 0$ in terms of increasing the coverage probability whenever the condition

$$tr(C_d) \geq 2C_{d00} \tag{3.23}$$

is satisfied.
Thus, if we restrict our attention to the class of designs of the type d^*, the problem reduces to that of choosing C_{d00} so as to maximize

$$Pr(\frac{v}{a}{\chi_1}^2 + \frac{v(v-1)}{vb-a}{\chi_{v-1}}^2 \leq \epsilon^2)$$

where $a = C_{d00}$ and $tr(C_d) = a + b$, in situations where $b \geq a$.
The solution is very much ϵ^2 dependent and hence appropriate weighted average probability by using a suitable weight distribution for ϵ^2 can be taken up.
Let $\mathcal{C}_{R_0,b,v,k}$ denote the class of *binary* connected block designs under consideration for given values of b, v and $k(< v)$ where R_0 is the replication number for the control treatment. As in the case of CRD assuming χ_2^2 distribution for ϵ^2 Mandal et.al (2000) present optimal values of n and R_0 for fixed values of wighted coverage probability $\bar{P} = .9$ and .95 and conclude that a Balanced treatment Incomplete Block (BTIB) design defined by Bechhofer and Tamhane(1981) with above values of n and R_0 is optimal for the given coverage probability.

4 DS-optimal Regression Designs

In this section we derive DS optimal designs for the LSE of $\underline{\beta}$ for a m-factor first degree model on asymmetric experimental domain and briefly summarize certain results on symmetric polynomial designs.

4.1 First degree polynomial fit model

We consider a m- factor first degree polynomial fit model:

$$Y_{ij} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \epsilon_{ij} \tag{4.1}$$

with m regression variables and experimental conditions $\underline{x}_i = (x_{i1}, \ldots, x_{im})' \;\; i = 1, \ldots, n$. The n point design $d_n = \{\underline{x}_1, \ldots, \underline{x}_n : p_1, \ldots, p_n\}$ has n_i replications of the level $\underline{x}_i$ with relative weight $p_i = \frac{n_i}{n}$, $i = 1, \ldots, n$ We assume tha the experimental domain is an $m-$ dimensional Euclidean ball of radius $\sqrt{m}$ that is

$$\mathcal{T}_{\sqrt{m}} = \{\underline{x} \in I\!R^m : \;\; \|\underline{x}\| \leq \sqrt{m}\}.$$

If the vectors $\underline{x}_i \in \mathcal{T}_{\sqrt{m}}$ fulfill the conditions

$$1 + \underline{x}_i' \underline{x}_i = m + 1, \quad 1 + \underline{x}_i' \underline{x}_j = 0 \tag{4.2}$$

for all $1 \leq i \neq j \leq m + 1$, then the vectors span a convex body in $I\!R^m$ called a *regular simplex* (cf. Pukelsheim 1993, p. 391). The vertices $\underline{x}_i$ belong to the boundary sphere of the ball $\mathcal{T}_{\sqrt{m}}$. A design which places weights p_i, $i = 1, \ldots, m + 1$ on the vertices of a regular simplex in $I\!R^m$ is called a *simplex* design. A design with equal weights $p_1 = p_2 = \cdots = \frac{1}{m+1}$ is called a *uniform simplex design.* For $m = 1$, the support points are $\underline{x}_1 = 1$ and $\underline{x}_2 = -1$ so that $(1, 1)'$ and $(1, -1)'$ satisfy the conditions (4.2). For $m = 2$, the support points $\underline{x}_1, \underline{x}_2, \underline{x}_3$ satisfying the conditions (4.2) belong to the boundary of $\mathcal{T}_{\sqrt{2}}$ and span an *equilateral triangle* on the sphere $\mathcal{T}_{\sqrt{2}}$. For example, the support points $(1, 1)', -\frac{1}{2}(1 + \sqrt{3}, 1 - \sqrt{3})', -\frac{1}{2}(1 - \sqrt{3}, 1 + \sqrt{3})'$ and every rotation of them span an equilateral triangle. For all $m \geq 1$ it is always possible to choose vectors $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{m+1}$ such that they fulfill the condition (4.2). A regular simplex design always exists. Note that any rotation of a regular simplex design is also a regular simplex design. The smallest possible support size of a feasible desgn for the LSE of $\underline{\beta}$ in a m factor first degree model is $n = m + 1$ because $\underline{\beta}$ has $m + 1$ components. Since the DS optimality criterion is isotonic with respect to the Lowener ordering and it depends on the information matrix only through the eigen values, there exists an $(m + 1)$ point simplex design on the ball $\mathcal{T}_{\sqrt{m+1}}$ that dominates any other design with respect to DS optimality criterion. Therefore the search of DS optimal design for the LSE of $\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_m)'$ can be restricted to the calss of $(m + 1)$point simplex designs.

Theorem 4.1 *Let d be an $(m + 1)$point design for the LSE of β in (4.1) over the ball $\mathcal{T}_{\sqrt{m+1}}$. Then d is DS optimal if and only if it is an $(m + 1)$ uniform simplex design.*

Proof: Note that the rows of the design matrix X under the $(m + 1)$ point design d are the support vetors $(1, \underline{x}_1'), (1, \underline{x}_2'), \ldots, (1, \underline{x}_{m+1}')$ with information matrix

$$I(d) = X'DX = \sum p_i (1, \underline{x}_i')'(1, \underline{x}_i')$$

where $D = diag(p_1, \ldots, p_{m+1})$. As already noted above, a DS optimal design can always be found in the set of $(m+1)$ point simplex designs. If d is a simplex design, the design matrix X is square and the non-zero eigenvalues of $X'DX$ and DXX' are the same. Hence

$$|I(d)| = |X'DX| = |DXX'| = (m+1)\prod_{i=1}^{m+1} p_i$$

Now the arithmetic mean -geometric mean unequality yields

$$\prod_{i=1}^{m+1} p_i \leq (\frac{1}{m+1})^{m+1} \tag{4.3}$$

Equality holds in (4.3) if and only if $p_1 = p_2 = \cdots = p_{m+1} = \frac{1}{m+1}$ i.e. d is a uniform simplex design. By Corollary 2.9

$$\psi_\epsilon[I(d)] \leq P[\chi^2_{m+1} \leq \delta^2 (m+1)^{\frac{1}{m+1}} (\prod_{i=1}^{m+1} p_i)^{\frac{1}{m+1}}]$$

where the R.H.S is an increasing function of $\prod_{i=1}^{m+1} p_i$. Thus by 4.3 d is a DS optimal design if and only if it is a uniform simplex design. ∎.
If $m = 1$ then we have the line fit model

$$y_{ij} = \beta_0 + \beta_1 x_i + \epsilon_{ij} \tag{4.4}$$

wih experimental domail $\mathcal{T} = [-1, 1]$ it follows from Theorem 4.1 that the design $d^* = \{-1, 1; 1/2\}$ which assigns weight 1/2 to the points -1 and 1 is the unique DS optimal design for the LSE of $\underline{\beta} = (\beta_0, \beta_1)'$ in 4.4.

Corollary 4.2 *Let $\mathcal{D}_n$ be the set of designs with support size $n \geq m+1$ in th m- way first degree model (4.1) on the experimental domain $\mathcal{T}_{\sqrt{m+1}}$. Then a design $d \in \mathcal{D}$ is DS optimal if $I(d) = I_{m+1}$.*

We omit the proof as it is a simple modification of the proof of the theorem 6.1.
By Theorem 6.1 $I(d) = I_{m+1}$ holds for a uniform simplex design d. Hence a design d satisfying the condition $I(d) = I_{m+1}$ exists for the minimal feasible support size $n = m+1$. Existence of a DS optimal design in general for any given pair of positive integers m and $n \geq m+1$ seems to be an unsolved problem. Nevertheless, A DS optimal design can be found for certain values of m and $n \geq m+1$. For example, the complete factorial design 2^m is a DS optimal design with $n = 2^m$. The information matrix of the complete factorial design 2^m is I_{m+1} and consequently it is a DS optimal design for te LSE of $\underline{\beta}$.

4.2 Symmetric Polynomial Designs

We postulate now a polynomial fit model of degree $k \geq 1$.

$$Y_{ij} = \beta_0 + \beta_1 x_i + \cdots + \beta_k x_i^k + \epsilon_{ij} \tag{4.5}$$

where $E(\epsilon_{ij}) = 0$ and $V(\epsilon_{ij}) = \sigma^2$ $i = 1, ..n$ and $j = 1, .., n_i$
The responses Y_{ij}'s are uncorrelated and the experimental conitions $x_1, \ldots, x_n$ are assumed to lie in $[-1, 1]$. Let $d \in \mathcal{D}$ be a design for the LSE of $\underline{\beta} = (\beta_0, \ldots, \beta_k)'$ on $\mathcal{T} = [-1, 1]$. We now consider the reflection operation. The reflected design d^R is given by $d^R =$

$\{-x_1, \ldots, -x_n : p_1, \ldots, p_n\}$. The design d and d^R have the same even moments, while the odd moments of d^R have a reversed sign. If $I(d^R)$ denotes the $(k+1)\times(k+1)$ information matrix of d^R, then

$$I(d^R) = QI(d)Q$$

where $Q = diag(1, -1, 1, -1, \ldots, \pm 1)$.
The symmetrized design

$$\bar{d} = \frac{1}{2}(d + d^R) = \{\underline{+}x_i;\ \ \frac{p_i}{2}, \frac{p_i}{2} | i \leq n\}$$

assigns the weights $\frac{p_i}{2}$ to x_i and $-x_i$ for each i.
The information matrix of $\bar{d}$ is

$$I(\bar{d}) = \frac{1}{2}[I(d) + QI(d)Q]$$

It is easy to see that $I(d^R)$ and $I(d)$ have the same eigenvalues. The DS optimality criterion for ψ_ϵ is invariant with respect to the reflection operation and it follows from the Corollary 2.7 that the symmetrized design $\bar{d}$ is at least as good as d with respect to the DS optimality criterion. Therefore, in symmetric experimental domain it is sufficient to consider symmertrized designs only. However, it turns out that in the polynomial regression model of degree $k > 1$ there exists no DS optimal design for the LSE of $\underline{\beta}$. The illustration using a quadratic regression model is to be found in Liski et.al. (1999).

5 D-, E- and DS(ϵ) optimality

Liski et.al (1999) studied the behaviour of the DS(ϵ) criterion when ϵ approaches 0 and ∞ respectively. These limiting cases have an interesting relationship wih the traditional D- and E- optimality criteria.
We now refer to a theorem from Liski et.al.(1999) without proof.

Theorem 5.1 *Let $\underline{\lambda}$ and $\underline{\gamma} \in I\!R^k$ denote vectors whose components are eigenvalues of the information matrices $I(d_1)$ and $I(d_2)$ respectively, arranged in decreasing order. Then the following statements hold:*

a) If $\psi_\delta(\underline{\lambda}) \geq \psi_\delta(\underline{\gamma})$ for all sufficiently small $\delta > 0$, then $|I(d_1)| \geq |I(d_2)|$ if $|I(d_1)| \geq |I(d_2)|$, then $\psi_\delta(\underline{\lambda}) > \psi_\delta(\underline{\gamma})$ for all sufficiently small $\delta > 0$.

b) If $\psi_\delta(\underline{\lambda}) \geq \psi_\delta(\underline{\gamma})$ for all sufficiently large δ, then $\lambda_k \geq \gamma_k$; if $\lambda_k > \gamma_k$ then $\psi_\delta(\underline{\lambda}) > \psi_\delta(\underline{\gamma})$ for all sufficiently large δ.

This theorem shows that DS(ϵ) criterion is equivalent to the D- criterion as $\epsilon \to 0$ and to the E- criterion as $\epsilon \to \infty$. These conclusions are due to the fact that both D -and E optimal designs are unique. (cf. Hoel 1958, Pukelsheim and Studden(1993)).

References

Birnbaum, Z. W. (1948). On random variables with comparable peakedness. Annals of Mathematical Statistics.**19**: 76-81.

Hoel, P. G. (1958). Efficiency problems in polynomial estimation. Annals of Mathematical Statistics. **29** :1134-1145.

Kiefer, J.C. (1974). General equivalence theory for optimum designs. Annals of Statistics. **2**: 849-879.

Kiefer, J.C.(1975). Construction and optimality of generalised Youden designs. In a survey of statistical design and linear models. Ed,. J.N.Srivastava. North Holland, Amsterdam, 333-353.

Liski, E. P., Luoma, A., Mandal, N. K. and Sinha, Bikas K. (1998). Pitman nearness,distance criterion and optimal regression designs. Calcutta Statistical Association Bulletin **48** (191-192): 179-194.

Liski, E. P., Luoma, A. and Zaigraev, A. (1999). Distance optimality design criterion in linear models. Metrika **49**: 193-211.

Majumdar, D (1996). Optimal and efficient treatment-control designs. In Handbook of Statistics **13** (eds. S. Ghosh and C.R.Rao) : 1007-1053.

Mandal, N. K. , Shah, K. R. and Sinha, Bikas K. (2000). Comparison of test vs. control treatments using distance optimality criterion. Metrika **52**: 147-162.

Marshall, A.W. and Olkin, I. (1979). Inequalities : Theory of majorization and Its Applications. Academic press, New york.

Okamoto, M. (1960). An inequality for the weighted sum of χ^2 variates. Bulletin of Mathematical Statistics **9**: 69-70.

Pukelsheim, F. (1993). Optimal Design of Experiments. John Wiley & Sons, Inc., New York.

Pukelsheim, F. and Studden, W. J. (1993). E-optimal designs for polynomial regressions. Annals of Statistics **21** 402-415.

SahaRay, R. and Bhandari, S. (2001) DS optimal designs in one-way ANOVA. to appear in Metrika.

Shah, K.R. and Sinha, Bikas K. (1989). Theory of Optimal Designs. Springer- Verlag Lecture Notes in Statistics Series, No. 54.

Sherman, S. (1955). A theorem on convex sets with applications. Annals of Mathematical Statistics **26** 763-766.

Sinha, Bikas K. (1970). On the optimality of some designs. Calcutta Statistical Association Bulletin, **20**: 1-20

Stepniak, C. (1989). Stochastic ordering and Schur-convex functions in comparison of linear experiments. Metrika**36** 291-298.

Tong, Y. L, (1990). The Multivariate Normal Distribution, Springer- Verlag, New York.