A COMPARISON OF THE DETERMINANT, MAXIMUM ROOT, AND TRACE OPTIMALITY CRITERIA

J. N. Srivastava and D. A. Anderson

Department of Statistics, Colorado State University
Fort Collins, Colorado, 80521
Department of Statistics, University of Wyoming
Laramie, Wyoming, 82071

ABSTRACT

Three basic criteria, determinant, trace and maximum root, are in common use for determining optimality of experimental designs. Here examples are presented where the three criteria give rise to different designs. The examples are balanced resolution IV^* of the 2^m series and are particularly insightful with respect to the dependence of the criteria on the correlation between estimators of the parameters.

INTRODUCTION

Consider the usual full rank linear model

$$E\{\underline{y}\} = X\underline{\beta}, \ Cov\{\underline{y}\} = \sigma^2 I_n \qquad (1)$$

where the matrix $X(n \times q)$ is determined by the design Δ with n points, and where $\underline{y}(n \times 1)$ is the vector of observations (assumed independent) on these n points. The least squares estimate of

$\underline{\beta}(q \times 1)$ is

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}, \text{ where} \tag{2}$$

$$\text{Cov}\{\hat{\underline{\beta}}\} = \sigma^2(X'X)^{-1} = V \quad \text{(say)}. \tag{3}$$

Three basic criteria are in common use for choosing an optimal design Δ:

(i) Determinant optimal: minimize $|V| = D$, say

(ii) Trace optimal: minimize $\text{tr}V = T$, say

(iii) Maximum-root optimal: minimize $\text{ch}_{\max} V = G$, say.

In terms of the characteristic roots of $(X'X)^{-1}$, these three criteria minimize, respectively, the product, sum, and maximum of the roots. The properties of these are well known. (See, for example, the papers listed at the end.) Briefly speaking, the determinant criterion leads (under normality) to a confidence region for the parameters with minimum "volume". The trace criterion minimizes "average variance", or equivalently, the expected mean square error $L^2(\hat{\underline{\beta}})$, given by

$$L^2(\hat{\underline{\beta}}) = (\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}). \tag{4}$$

The maximum root criterion finds that (normalized) linear combination of the parameters whose estimate has the largest variance and chooses Δ to minimize that variance.

Let $\lambda_1 \geq \cdots \geq \lambda_q$ denote the latent roots of $(X'X)$. Let

$$T_s = \left[\frac{1}{q} \sum_{j=1}^{q} \left(\frac{1}{\lambda_j}\right)^s\right]^{1/s} \cdots \tag{5}$$

Then, assuming that the first four moments of $\underline{y}$ are the same as under normality, it is easily checked that

$$\text{Exp}\{L^2(\hat{\underline{\beta}})\} = \sigma^2 \sum_{1}^{q} \left(\frac{1}{\lambda_j}\right) = \sigma^2 q T_1 = \sigma^2 T; \text{ and} \tag{6}$$

$$\mathrm{Var}\{L^2(\hat{\underline{\beta}})\} = 2\sigma^4 \sum_{1}^{q} \left(\frac{1}{\lambda_j^2}\right) = 2q\sigma^4 T_2^2. \qquad (7)$$

Also, it is well known that

$$\lim_{s\to\infty} T_s = \lambda_1^{-1} = \frac{1}{\sigma^2} G, \; \lim_{s\to 0} T_s = \left[\prod_{j=1}^{q} \lambda_j^{-1}\right]^{1/q} = \frac{1}{\sigma^2} \{D\}^{1/q}, \qquad (8)$$

so that the class of criteria T_s $(0 \leq s \leq \infty)$ contains the three main ones.

There are occasions for which a design Δ may satisfy all of the basic criteria. On other occasions, since the minimization is on different functions of the roots, the various criteria may give different designs and the experimenter must choose his optimality criteria. The purpose of this note is to present an example where the various criteria give rise to different designs and to compare the criteria relative to these designs. This example is particularly insightful with respect to the dependence of the criteria on the correlation between the estimator of various parameters. The example is selected from balanced resolution IV* designs of the 2^m series, that is designs permitting estimates of main effects orthogonal to general mean and two-factor interactions. For brevity, only T_0, T_1, T_2 and T_∞ will be considered.

OPTIMAL BALANCED RESOLUTION IV* DESIGNS - 2^m SERIES

Srivastava and Anderson (1970) have shown that a necessary condition a design Δ be balanced and resolution IV* is that Δ be (1,0) symmetric of strength 3, and that for any pair of factors the number of assemblies in Δ with both factors at level 1 must be a constant, say ω, independent of the pair of factors. The design is trace-optimal if

1. $N = 4\alpha$, α even, $\omega = \alpha$;
2. $N = 4\alpha + 2$, α even, $\omega = \alpha + 1$;
3. $N = 4\alpha$, α odd, $\omega = \alpha + 1$;
4. $N = 4\alpha + 2$, α odd, $\begin{cases} \omega = \alpha \text{ if } m \leq (N + 6)/3 \\ \omega = \alpha + 2 \text{ if } > (N + 6)/3. \end{cases}$

The case $N = 4\alpha + 2$ with α odd provides an interesting comparison between the optimality criteria. Four cases, $N = 14, 22, 30, 38$, will be sufficient to illustrate this comparison. Denote by design I (or Δ_1) a design with $\omega = \alpha$ and by design II (or Δ_2) a design with $\omega = \alpha + 2$. The determinant-optimal design is I for all values except $N = 30$, $m = 15$, and $N = 38$, $m = 19$ (for $N = 14, 22, 30, 38$). The maximum-root-optimal design is Δ_1 when $m < 4$, Δ_2 when $m > 4$, and when $m = 4$ the maximum roots are the same for Δ_1 and Δ_2. The T_2-optimal design is strictly between G and T_1 optimal in terms of intervals of m. Thus only when $N = 30$, $m = 15$; $N = 38$, $m = 19$, and $m \leq 4$, do the four criteria coincide.

For this example it is a simple matter to compare the designs selected by the various criteria. In Table I we present the optimality constants of designs I and II. The first two columns, headed by G, give the roots of $(X'X)^{-1}$ with multiplicates $m - 1$ and 1, respectively. The third column, headed by T_2, actually gives T_2. The column headed by T_1 gives T_1, and since the designs are balanced this equals $\mathrm{var}(\hat{\beta}_i)$. The column headed by D gives $\sigma^{-2}D^{1/q}$. Notice that this choice makes the quantities in columns 1-5 comparable. The final column gives the correlation between estimates of the β_i, $\mathrm{corr}(\hat{\beta}_i, \hat{\beta}_j)$, which is a constant for all i and j since the designs are balanced.

For purpose of discussion let us consider the case $N = 22$. The trace and determinant criteria differ when $m = 10$ and 11. One notes from Table I that the determinant optimal design has not only a larger variance in each case but also excessively large correlations between estimates. When $m = 8$ and 9, design I is T_1 optimal and design II is T_2 optimal. Design Δ_2 has only slightly larger variance than Δ_1 in these two cases and correlations which are substantially smaller (in absolute value). The G and T_2 criteria differ for $m = 5, 6, 7$ and possibly 4 as the G criterion is indifferent between Δ_1 and Δ_2. As m decreases we find the relative variance of design II to I increasing while the relative correlation is decreasing, so that for $m \leq 5$ the absolute correlations for design

TABLE I

Optimality Constants

Design I $\omega = (N-2)/4$

	m	G		T_2	T_1	D	Corr.
N=14	3	.0625	.1000	.0771	.0750	.0731	.1667
	4	.0625	.1250	.0827	.0781	.0743	.2000
	5	.0625	.1667	.0932	.0833	.0760	.2500
	6	.0625	.2500	.1170	.0938	.0787	.3333
	7	.0625	.5000	.1976	.1250	.0841	.5000
N=22	3	.0417	.0556	.0468	.0463	.0459	.1000
	4	.0417	.0625	.0477	.0469	.0461	.1111
	5	.0417	.0714	.0491	.0476	.0464	.1250
	6	.0417	.0833	.0510	.0486	.0467	.1429
	7	.0417	.1000	.0540	.0500	.0472	.1667
	8	.0417	.1250	.0589	.0521	.0478	.2000
	9	.0417	.1667	.0680	.0555	.0486	.2500
	10	.0417	.2500	.0884	.0625	.0498	.3333
	11	.0417	.5000	.1560	.0833	.0522	.5000
N=30	3	.0313	.0385	.0338	.0337	.0334	.0714
	4	.0313	.0417	.0342	.0338	.0336	.0769
	5	.0313	.0455	.0346	.0341	.0337	.0833
	6	.0313	.0500	.0341	.0344	.0338	.0909
	7	.0313	.0556	.0357	.0347	.0339	.1000
	8	.0313	.0625	.0366	.0352	.0341	.1111
	9	.0313	.0714	.0379	.0357	.0342	.1250
	10	.0313	.0833	.0397	.0365	.0345	.1429
	11	.0313	.1000	.0424	.0375	.0347	.1667
	12	.0313	.1250	.0469	.0391	.0351	.2000
	13	.0313	.1667	.0551	.0417	.0355	.2500
	14	.0313	.2500	.0733	.0469	.0363	.3333
	15	.0313	.5000	.1326	.0625	.0376	.5000
N=38	3	.0250	.0294	.0266	.0265	.0264	.0555
	4	.0250	.0313	.0267	.0266	.0264	.0588
	5	.0250	.0333	.0269	.0267	.0265	.0625
	6	.0250	.0357	.0271	.0268	.0265	.0667
	7	.0250	.0385	.0273	.0269	.0266	.0714
	8	.0250	.0417	.0276	.0271	.0266	.0769
	9	.0250	.0454	.0280	.0273	.0267	.0833
	10	.0250	.0500	.0285	.0275	.0268	.0909
	11	.0250	.0555	.0291	.0278	.0269	.1000
	12	.0250	.0625	.0300	.0281	.0270	.1111
	13	.0250	.0714	.0311	.0286	.0271	.1250
	14	.0250	.0833	.0328	.0292	.0272	.1429
	15	.0250	.1000	.0354	.0300	.0274	.1667
	16	.0250	.1250	.0395	.0313	.0276	.2000
	17	.0250	.1667	.0471	.0333	.0280	.2500
	18	.0250	.2500	.0637	.0375	.0284	.3333
	19	.0250	.5000	.1173	.0500	.0293	.5000

TABLE I

Optimality Constants

Design II $\omega = (N+6)/4$

	m	G		T_2	T_1	D	Corr.
N=14	3	.1250	.0385	.1045	.0962	.0844	-.3000
	4	.1250	.0313	.1094	.1016	.0884	-.2308
	5	.1250	.0261	.1124	.1053	.0915	-.1875
	6	.1250	.0227	.1145	.1080	.0941	-.1579
	7	.1250	.0200	.1160	.1100	.0962	-.1364
N=22	3	.0625	.0294	.0538	.0515	.0486	-.2143
	4	.0625	.0250	.0556	.0531	.0497	-.1765
	5	.0625	.0217	.0567	.0543	.0506	-.1500
	6	.0625	.0192	.0576	.0553	.0514	-.1304
	7	.0625	.0172	.0582	.0560	.0520	-.1154
	8	.0625	.0156	.0587	.0566	.0526	-.1034
	9	.0625	.0148	.0591	.0571	.0530	-.0938
	10	.0625	.0136	.0594	.0577	.0535	-.0857
	11	.0625	.0122	.0597	.0579	.0539	-.0789
N=30	3	.0417	.0238	.0367	.0357	.0346	-.1667
	4	.0417	.0208	.0376	.0365	.0350	-.1429
	5	.0417	.0185	.0382	.0370	.0354	-.1250
	6	.0417	.0167	.0386	.0375	.0358	-.1111
	7	.0417	.0152	.0390	.0379	.0361	-.1000
	8	.0417	.0139	.0393	.0382	.0363	-.0909
	9	.0417	.0128	.0395	.0385	.0366	-.0833
	10	.0417	.0190	.0397	.0387	.0368	-.0769
	11	.0417	.0111	.0399	.0389	.0369	-.0714
	12	.0417	.0142	.0400	.0391	.0371	-.0667
	13	.0417	.0098	.0401	.0392	.0373	-.0625
	14	.0417	.0093	.0402	.0394	.0374	-.0588
	15	.0417	.0088	.0403	.0395	.0376	-.0556
N=38	3	.0313	.0200	.0280	.0275	.0269	-.1364
	4	.0313	.0179	.0285	.0279	.0272	-.1200
	5	.0313	.0161	.0289	.0282	.0274	-.1071
	6	.0313	.0147	.0292	.0285	.0276	-.0968
	7	.0313	.0135	.0294	.0287	.0277	-.0882
	8	.0313	.0125	.0296	.0289	.0279	-.0811
	9	.0313	.0116	.0297	.0291	.0280	-.0750
	10	.0313	.0109	.0298	.0292	.0281	-.0698
	11	.0313	.0120	.0300	.0293	.0282	-.0652
	12	.0313	.0096	.0300	.0294	.0283	-.0612
	13	.0313	.0091	.0301	.0295	.0284	-.0577
	14	.0313	.0086	.0302	.0296	.0285	-.0545
	15	.0313	.0082	.0303	.0297	.0286	-.0517
	16	.0313	.0078	.0303	.0298	.0287	-.0492
	17	.0313	.0075	.0304	.0299	.0287	-.0469
	18	.0313	.0071	.0304	.0299	.0288	-.0448
	19	.0313	.0068	.0305	.0300	.0289	-.0429

II are greater than for design I. Thus for m = 5, the G-criterion selects a design with both larger variance and larger correlation. Only for m = 3 do all criteria agree. A similar pattern may be noted for the other values of N.

To have a closer look at the situation, notice that V is a $(q \times q)$ matrix of the form

$$V = \sigma^2[(1-\rho)I + \rho J], \tag{9}$$

when $I(q \times q)$ and $J(q \times q)$ are respectively the identity matrix, and the matrix having 1 everywhere. The roots of $\sigma^{-2}V$ are $(1-\rho)$ and $(1-\rho+\rho q)$ with multiplicates (q-1) and 1 respectively. Hence, we obtain

$$|V|^{1/q} = \sigma^2(1-\rho)^{1-1/q}[1 + \rho(q-1)]^{1/q}, \quad -1/(q-1) \leq \rho \leq 1, \tag{10}$$

$$trV = \sigma^2[(1-\rho)(q-1) + (1-\rho + \rho q)]$$

$$ch_{max} V = \sigma^2 max[1-\rho, 1-\rho+\rho q].$$

Notice that for $q > 2$, the lower bound for ρ is not (-1), but a quantity which rapidly decreases in magnitude with q. Now, let V_1, σ_1^2 and ρ_1 denote the values of V, σ^2 and ρ respectively for Δ_1, and V_2, σ_2^2 and ρ_2 the same quantities for Δ_2. From Table I, notice that we always have $\rho_1 > 0$, $\rho_2 < 0$. Hence for larger values of q

$$ch_{max} V_1 = \sigma_1^2(1-\rho_1 + \rho_1 q), \quad ch_{max} V_2 = \sigma_2^2(1-\rho_2). \tag{11}$$

Since σ_1^2 and σ_2^2 are nearly equal, Δ_2 is G-preferable to Δ_1, if we (roughly) have $\rho_1 > (q-1)^{-1}(-\rho_2)$. On the other hand, $ch_{max}(V_2)$ has multiplicity (q-1). Thus, for larger values of q, $q > 4$, Δ_1 has one root (say R_1) somewhat large and the rest small, while Δ_2 has (q-1) roots equal and a little smaller than R_1. In other words, most roots are small in case of Δ_1 and large in case of Δ_2. Hence, in this case the G-criterion has some obvious deficiencies.

Now, compare T and D, in cases where they give rise to different designs. For many of the large values of m, Δ_2 becomes T-preferable while Δ_1 still remains D-optimal. In such situations, notice that ρ_1 is rather large compared to $|-\rho_2|$. In other words, the D-optimality of Δ_1 seems to be because of high correlations. Thus, because of such features, and also because of its simple interpretive value, the trace-optimality seems to be the better criterion for the above type of designs.

BIBLIOGRAPHY

[1] Box, M. J. and Draper, N. R. [1971). Factorial designs, the $|X'X|$ criterion, and some related matters. Technometrics 13, 731-43.

[2] Kiefer, J. (1959). Optimum experimental designs. J. R. Statist. Soc. B 21, 272-304.

[3] Kiefer, J. (1960). Optimum experimental designs V, with applications to systematic and rotatable designs. Proc. 4th Berkeley Symp. Math. Statist. Prob. 1, 381-405. Univ. of California Press, Berkeley.

[4] Nalimov, V. V., Golikova, T. I. and Nikeshina, N. G. (1970). On practical use of the concept of D-optimality. Technometrics 48, 799-812.

[5] Srivastava, J. N. (1965). Optimal balanced 2^m fractional factorial designs. S. N. Roy Memorial Volume, University of North Carolina, Chapel Hill and Indian Statistical Institute, Calcutta.

[6] Srivastava, J. N. and Anderson, D. A. (1970). Optimal fractional factorial plans for main effects orthogonal to two-factor interactions: 2^m series. J. Amer. Statist. Assoc. 65, 828-43.