Tue Tjur

An Algorithm for Optimization of Block Designs

Summary. This paper presents an algorithm for construction of A–optimal or nearly A–optimal block designs. An implementation for IBM personal computers is available from the author.

American Mathematical Society 1980 subject classification.
Primary 62K05; secondary 62K10.

Key words and phrases.
Optimal design, block design, A–optimality, efficiency factor.

0. Introduction.

This paper describes an algorithm for construction of optimal or nearly optimal block designs by repeated interchange of the treatments on pairs of plots. Since the algorithm is very similar to algorithms presented by Jones and Eccleston (1980) and Cook and Nachtsheim (1989), we shall begin with a description of these algorithms, emphasizing the points where our algorithm is different.

The algorithm of Jones and Eccleston (1980) has two phases. In phase 1, an initial (more or less arbitrary) design of the desired dimensions is modified by a sequence of *exchange procedures*. An exchange procedure amounts to the replacement of the treatment on a plot with another treatment. Thus, block sizes are fixed, while replicate counts are changed. In each step, the choice of the next exchange to be performed is roughly based on the idea that the efficiency factor should be increased as much as possible. However, the change in efficiency due to a given exchange is estimated rather than computed. This estimation is based on the change due to removal or addition of single plots, which can be computed from the information matrix for treatment parameters, adjusted for blocks, by relatively simple formulas. When this strategy gives no further improvement, phase 2 is started. This phase is a sequence of *interchange procedures*, i.e. operations interchanging the treatments of two plots. Again, the intention seems to be to select interchanges that induce maximal increase of the efficiency, and again the increase is estimated from the consequences of simple delete/add operations.

The algorithm suggested by Cook and Nachtsheim (1989) is similar, but their strategy for choice of next ex- or interchange is exact, in the sense that they use exact formulas for the consequences of these operations for their optimality criterion. The criterion considered is the determinant of the information matrix, i.e. they consider D–optimality rather than A–optimality. Their approach is not restricted to the case where the treatment structure is the effect of a single factor, but allows for an arbitrary design matrix determining the treatment structure.

Our algorithm can (though it was developed independently of Cook and Nachtsheim) be described as a further development of the algorithm of Jones and Eccleston, making use of computational ideas similar to those used by Cook and Nachtsheim. Our optimality criterion is A–optimality, amounting to maximization of the classical efficiency factor or minimization of the average contrast variance. Our phase 1 differs from Jones and Eccleston's phase 1 by building up the design from scratch, adding experimental units one by one, rather than exchanging treatments in a full design. Our phase 2 goes like that of Jones and Eccleston, except that

the strategy is based on an exact formula for the decrease of average contrast variance induced by an interchange procedure. Like Cook and Nachtsheim, we work with the full information matrix for treatment and block parameters, rather than the adjusted information matrix for treatments only. This is essential for the availability of the computationally very cheap updating formulas, which enable us to perform a complete search through the entire set of possible interchanges within a reasonable time.

An additional feature of our implementation of this algorithm, which seems to have no counterpart in the implementations of the two other algorithms, is a randomization device for escape from 'local optima'. Our approach here has a taste of simulated annealing (see e.g. Geman and Geman 1984), but is really just based on the simple idea of searching through a small random sample of possible switches and select the best.

A further advantage of our algorithm, which can not be ignored in practice even if it is theoretically irrelevant, is that our algorithm is available as a self–explanatory program running on IBM PC's and compatibles under DOS. Please send an empty discette to the author, formatted the way you want it.

1. The algorithm.

Phase 1. Our starting point can be viewed as a totally disconnected design consisting of B empty blocks of size k, none of the plots being used for anything yet. Given are also the number T of possible treatments, and the prescribed replicate count r. Thus, $Bk = Tr$ is the number of plots or experimental units that we intend to end up with in the final design. Our implementation of the algorithm includes the possibility of different replicate counts and/or different block sizes, but for simplicity of notation, this trivial extension is ignored in the following. Now, plots are assigned treatments one by one, according to the following principles. As long as the design is disconnected, each new block–treatment combination should be such that the number of connected components of the design network is reduced by one (see Tjur 1987; our 'empty' initial design corresponds to an entirely disconnected network, consisting of B block points and T treatment points, with no connections at all). This strategy ensures that the design will be connected after the introduction of $B + T - 1$ plots. Apart from this restriction, and some less important attempts to keep the connected compononts equally sized, the assignments during these first $B + T - 1$ steps of phase 1 are performed at random, under the ties given by the design constants. Once connectedness is obtained, it makes sense to talk about the contrast

variances (i.e. the variances on differences between estimated treatment parameters in the intra–block analysis), and from then on the strategy is to select in each step that treatment–block combination which gives maximal decrease of the average of the $\binom{T}{2}$ contrast variances.

Designs obtained in this way are usually good, but rarely optimal. It can be added that for large designs this construction may take a long time on the computer. Our implementation includes, as an optional choice, the possibility of entering a design directly as the starting point for phase 2. J. A. John (personal communication) has reported good results with the best cyclic design as a starting point for phase 2, see John (XXX).

Phase 2. In order to improve the design, we proceed as follows. Let a *switch* be an operation that interchanges the treatments t_1 and t_2 assigned to plots on the blocks b_1 and b_2. This is what Jones and Eccleston called an interchange procedure, but we prefer the term 'switch' due to the analogy with electrical networks that initiated our work on this algorithm, see Tjur (1987). Among all possible switches we select, in each step, one according to the following criteria. Switches that would make the design disconnected are not taken into account. Also, switches that would make the design 'less orthogonal' than it was before are ignored; by this we mean that only switches satisfying

$$n_{b_1t_1} > n_{b_2t_1} \text{ and } n_{b_2t_2} > n_{b_1t_2}$$

are taken into account (n_{bt} denoting the number of plots of block b with treatment t). Among the switches allowed, we select the one that gives maximal decrease of the average contrast variance.

Phase 2 continues like this until a design is reached for which no switch will decrease the average contrast variance. Such designs are called *locally optimal*, because they are better than all their 'switch neighbours'. Of course, local optimality does not guarantee (global) optimality. The choice of a reasonable starting point (e.g. the result of phase 1) may be critical here. For small design constants, the algorithm consisting of phase 1 and phase 2 turns out to work well. For larger designs, phase 2 is more frequently 'trapped' by locally optimal designs. This problem can, to some extend, be handled by the addition of a third phase.

Phase 3. We proceed as in phase 2, but in each step the search is restricted to a random sample of given (user–defined) size, drawn (with replacement, for simplicity) from the set of possible switches. In this phase, switches increasing the average contrast variance are allowed (otherwise we would not be able to start). Having 'shuffled the cards' for a while

in this way, we initiate a new phase 2 search, etc. etc. In our present implementation, phase 2 is reentered when the number of random switches performed exceeds a (user–defined) bound, and also whenever a design better than all previous designs is found.

2. Test runs.

EXAMPLE 1 (a sample of BIBDs). For given T and $k < T$, let B be the smallest integer such that both $r = kB/T$ and $\lambda = k(k-1)B/(T(T-1))$ are integer. The algorithm was applied to the 51 sets of design constants (B, T, k, r) of this form satisfying $3 \leq T \leq 12$ and $B \leq 56$. In all cases, a balanced incomplete block design is known to exist, and it is known (Kiefer 1958) that BIBDs are optimal when they exist. A BIBD was found as the result of phase 2 in 40 out of the 51 cases. The number of switches required by step 2 was less than 12 in these cases. In 10 of the remaining 11 cases, a BIBD was found by randomized search in less than 40 switches.

EXAMPLE 2 (a sample mainly consisting of regular graph designs). John and Mitchell (1977) listed 145 triples (T, k, r) for which the optimal design was found by systematic search among all possible designs. In all but 9 of these cases, a design with the same efficiency as the optimal design (reported with 3 significant digits by John and Mitchell) was found by phase 2. In the remaining 9 cases, designs with at least that efficieny minus 0.001 was found. The number of switches performed was less than 8 in all cases.

EXAMPLE 3. (simple lattices for k = 3, 4, 5, 6 and 7). The algorithm was applied to design constants of the form $B = 2k$, $T = k^2$, $r = 2$, for $k = 3$, 4, 5, 6 and 7. A simple lattice was found in phase 2 after 1, 3, 2, 5 and 6 switches, respectively.

Our experiences confirm the conjecture by John and Mitchell (and several other authors) that any optimal design (among equi–replicate designs) must be a regular graph design if such exist. Indeed, the switches chosen in phase 2 tend to shrink the distribution of the concurrence counts whenever possible. Only 10 of the 145 designs found in the test run described by example 2 were not regular graph designs, and some of these are of dimensions for which a regular graph design does not exist.

3. Computations.

The statistical model considered can be written on the form

$$y_i = \alpha_t + \beta_b + \sigma u_i,$$

where the u_i are i.i.d. normal (0,1). Or, in matrix notation,

$$\mathbf{y} = \mathbf{X}_T\alpha + \mathbf{X}_B\beta + \sigma\mathbf{u}.$$

The information matrix (for the full parameter set, $B+T$ parameters) is then

$$\mathbf{C} = \begin{pmatrix} \mathbf{X}_T^*\mathbf{X}_T & \mathbf{X}_T^*\mathbf{X}_B \\ \mathbf{X}_B^*\mathbf{X}_T & \mathbf{X}_B^*\mathbf{X}_B \end{pmatrix}.$$

Let

$$\mathrm{var}(\hat{\alpha}_{t'} - \hat{\alpha}_{t''}) = \sigma^2 R(t', t'')$$

and

$$\mathrm{var}(\hat{\alpha}_t + \hat{\beta}_b) = \sigma^2 R(t, b)$$

be the expressions for contrast variances and variances on fitted values, respectively. Let $\mathbf{C}^-$ be a reflexive generalized inverse of $\mathbf{C}$ (cfr. Rao 1973). Then, with an obvious notation for column vectors of length $B+T$ which are 1 at single entry and 0 otherwise, we have

$$R(t', t'') = (1_{t'} - 1_{t''})^*\mathbf{C}^-(1_{t'} - 1_{t''})$$

and

$$R(t, b) = (1_t + 1_b)^*\mathbf{C}^-(1_t + 1_b).$$

Let $\bar{R}$ denote the average of the $\binom{T}{2}$ quantities $R(t', t'')$ (the normalized average contrast variance), and let $\mathbf{M}$ be the matrix that comes out by averaging over (t', t'') of the matrices

$$\mathbf{C}^-(1_{t'} - 1_{t''})(1_{t'} - 1_{t''})^*\mathbf{C}^-.$$

Then, we have the following 'updating formulas', relevant for phase 1 and 2 respectively.

If the design is extended by a plot in block b with treatment t, then the change in average contrast variance is given by

$$\bar{R}_{\mathrm{new}} = \bar{R}_{\mathrm{old}} - \frac{(1_b + 1_t)^*\,\mathbf{M}\,(1_b + 1_t)}{1 + R_{\mathrm{old}}(b, t)}.$$

If the design is changed by a switch, interchanging treatments t_1 and t_2 on blocks b_1 and b_2 respectively, then the change in average contrast variance is given by

$$\bar{R}_{\text{new}} = \bar{R}_{\text{old}} + \frac{(\delta_t^* \mathbf{C}^- \delta_t)(\delta_b^* \mathbf{M} \delta_b) + (\delta_b^* \mathbf{C}^- \delta_b)(\delta_t^* \mathbf{M} \delta_t) + 2(1 - \delta_b^* \mathbf{C}^- \delta_t)(\delta_t^* \mathbf{M} \delta_b)}{(1 - \delta_b^* \mathbf{C}^- \delta_t)^2 - (\delta_t^* \mathbf{C}^- \delta_t)(\delta_b^* \mathbf{C}^- \delta_b)}$$

where

$$\delta_t = 1_{t_1} - 1_{t_2}, \qquad \delta_b = 1_{b_1} - 1_{b_2}.$$

We shall not give the proof of these formulas here (see Tjur 1987). The basic idea, noticed also by Cook and Nachtsheim (1989), is that the addition of a plot to the design corresponds to the addition of a rank 1 matrix to $\mathbf{C}$, while a switch corresponds to the addition of a matrix of rank 2. There are wellknown formulas for updating a generalized inverse under such transformations. The complexity of the 'switch formula' is due to the fact that this updating involves the explicit inversion of a 2×2–matrix. Notice, however, that once the $(B+T) \times (B+T)$ symmetric matrices $\mathbf{C}^-$ and $\mathbf{M}$ are computed and stored, the consequence of a modification of the design is given by an expression which does not even involve summations over ranges depending on the design constants. The 'matrix products' occurring in the formulas are really just sums and differences of a few elements from the matrices $\mathbf{C}^-$ and $\mathbf{M}$.

References.

Cook, R. D. and Nachtsheim, C. J. (1989)
Computer–aided blocking of factorial and response–surface designs.
Technometrics **31**, 339–346.

Geman, S. and Geman, D. (1984).
Stochastic relaxation, Gibbs distributions, and the Baysian restoration of images.
IEEE transact. on pattern analysis and machine intelligence **6**, 721–741.

John, J. A. (1990 or 91).
XXX (John's contribution at the Tampere meeting)
XXX

John, J. A. and Mitchell, T. B. (1977).
Optimal incomplete block designs.
J. R. Statist. Soc. B **39**, 39–43.

Jones, B. and Eccleston, J. A. (1980).
Exchange and interchange procedures to search for optimal designs.
J. R. Statist. Soc. B **42**, 238–243.

Kiefer, J. (1958).
On the nonrandomized optimality and the randomized nonoptimality of symmetrical designs.
Ann. Math. Statist. **29**, 675–699.

Rao, C. R. (1973).
Linear Statistical Inference and Its Applications.
Wiley, New York.

Tjur, T. (1987).
Block designs and electrical networks.
Preprint 6 1987, Inst. of Math. Stat., University of Copenhagen.
(shorted version under publication in *Ann.Stat.*)

Tue Tjur
Institute of Mathematical Statistics
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen
DENMARK