Computational Statistics and Data Analysis 71 (2014) 1134–1146

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

An efficient procedure for the avoidance of disconnected incomplete block designs

J.D. Godolphin [a,*], H.R. Warren [b]

[a] University of Surrey, Guildford, Surrey, GU2 7XH, UK

[b] William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University of London, London, EC1M6BQ, UK

ARTICLE INFO

Article history:
Received 25 July 2012
Received in revised form 2 July 2013
Accepted 24 September 2013
Available online 10 October 2013

Keywords:
Connectivity
Incomplete block design
Missing values
Observation loss
Rank reducing observation sets
Robustness
Selective partitioning

ABSTRACT

Knowledge of the cardinality and the number of minimal rank reducing observation sets in experimental design is important information which makes a useful contribution to the statistician's tool-kit to assist in the selection of incomplete block designs. Its prime function is to guard against choosing a design that is likely to be altered to a disconnected eventual design if observations are lost during the course of the experiment. A method is given for identifying these observation sets based on the concept of treatment separation, which is a natural approach to the problem and provides a vastly more efficient computational procedure than a standard search routine for rank reducing observation sets. The properties of the method are derived and the procedure is illustrated by four applications which have been discussed previously in the literature.

1. Introduction

Experimental design researchers and practitioners have been aware for many years that serious field problems arise when one or more observations are not recorded, for whatever reason, during the experimental process so that the eventual design is different from the design that was planned originally. For instance, Yates (1933) drew attention to problems in the field "when the yields of some plots are lost, or are unreliable" and Cochran and Cox (1957, Section 3.7) discussed difficulties "when certain observations are missing, through failure to record, or gross errors in recording or accidents". These experimenters were concerned primarily with the removal of orthogonality of the planned design but such concerns are mitigated in modern applications by routine use of statistical software packages. However, in experiments arranged in incomplete blocks, there are other problems with missing observations in the field that are not covered by data analysis packages, viz. (i) the inevitable loss in design efficiency and (ii) the risk that the eventual design is disconnected with respect to treatment effects, so that not all treatment contrasts are estimable. Dey et al. (2001) refer to these two problems as problems of design *robustness efficiency* and design *robustness connectivity*, respectively, and, of the two problems, it is the second which is easily the most serious. If the eventual design is disconnected the test of the usual null hypothesis that all treatment effects have the same value breaks down and many, if not most, of the pairwise treatment contrasts are inestimable. The avoidance of this unwelcome situation, through careful selection of the planned design, is clearly an important objective of both researcher and practitioner.

Computer-aided procedures for construction of incomplete block designs have been given by several authors, e.g. Mitchell (1974), Jones and Eccleston (1980), Whitaker et al. (1990), Nguyen and Miller (1992), John et al. (1993),

* Corresponding author.
E-mail address: j.godolphin@surrey.ac.uk (J.D. Godolphin).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.09.025

Nguyen (1994), Angelis (2003), Soicher (2011) and others. The algorithms given by these authors employ various criteria for design construction, including a number of interchange and/or evolutionary techniques to search for an A-optimal design, either by maximizing the average efficiency factor, or some suitable approximation to it based on powers of the trace of the concurrence matrix. Since these methods do not appear to be programmed to cater for the problem of robustness connectivity, it is sensible to output several designs with optimal or near-optimal properties and consider them for robustness before suggesting a design for experimental use. This strategy is reasonable because the most efficient design is not necessarily the best design according to robust connectivity criteria, see Godolphin (2004) and Bate et al. (2008), and it is good practice to compare robustness properties of the derived 'near-optimal' designs before making a final design selection.

It is assumed throughout this paper that it is not possible, realistically, to anticipate beforehand which observations are likely to go missing during an experiment. Godolphin (2004, 2006) refers to such observations as a rank-reducing observation set (RROS) that is Type 1 if the loss results in a disconnected eventual design and is Type 3 if the loss results in the elimination of all replicates of one or more treatments. The identification of RROSs by the approach of Godolphin (2006) makes use of the Z-matrix algebra of Theil (1965) and is a useful method if the design is not too large and the RROSs are small. However, the method is not practicable if the design is relatively large sized because of the large number of sets that require examination and it is not programmed to give information on which treatment contrasts are inestimable. This paper describes a procedure based on the concept of treatment separation, which is a two stage process that searches through a restricted class of subsets of the υ treatments and scans the blocks of the design to identify RROSs induced by these treatment subsets. This selective partitioning is a more informative approach to the identification of RROSs for incomplete block designs. Furthermore, when coupled with a method of adjusted selective partitioning it provides a procedure that is highly efficient computationally. The aim of the method is to find the smallest number of observations involved in the Type 1 RROSs and the number of such minimal sets; but the searching exercise is confined to relatively few treatment subsets so the output is obtained from far less computations than those necessary for a routine search through observation sets of increasing number and size.

The main results on selective partitioning and adjusted selective partitioning for the identification of RROSs of smallest cardinality are given in Section 2. Several properties of the method, with its implications for computation reduction, are derived. In Section 3 this procedure is illustrated by identifying the RROSs and examining the vulnerability of block designs in four experimental situations in the literature. The procedure is carried out using a program written in Matlab which can be found within the online supplementary material referred to in Appendix A.

2. Description of the procedure

2.1. Preliminaries

Consider a planned connected binary block design, D, on υ treatments arranged in b blocks of size k ($k < \upsilon$); let r_i be the number of replicates of the ith treatment and let $r_{[1]} \geq r_{[2]} \geq \cdots \geq r_{[\upsilon]}$ denote the treatment replication numbers in decreasing order. If one or more observations fail to be recorded during the course of the experiment then D is effectively replaced by an eventual design D_e. In general, D_e has $\upsilon_* \leq \upsilon$ treatments arranged in $b_* \leq b$ blocks of size varying between 1 and k. Either the ith treatment does not occur in D_e or it occurs in D_e with replication number that is at most r_i. The difference between the two designs is the set of unavailable observations removed from D to yield D_e and this is referred to, simply, as an *observation set*.

2.2. Rank reducing observation sets

The connectivity status of a block design is related directly to the rank of its design matrix: see, for example, Godolphin (2013). However, there are three kinds of *rank reducing observation set* (RROS), in which the rank of the design matrix for D_e is strictly less than the rank of the design matrix for D.

(i) When a Type 1 RROS is removed from D the eventual design D_e is disconnected. In this case the b_* blocks of D_e are partitioned into two non-empty sets S_1, S_2 such that the blocks of S_1 contain all replicates of a proper subset of the υ_* treatments and the blocks of S_2 contain all replicates of the remaining treatments in D_e: see Godolphin and Warren (2011, Section 3) for further discussion of this situation. Equivalently, D_e can be represented as a bipartite graph consisting of two sets of vertices, one set corresponding to the υ_* treatments and one set corresponding to the b_* blocks, with an edge drawn between a treatment vertex and a block vertex whenever the treatment occurs in the block; this graph is disconnected since there is no path between any vertex for a treatment contained in blocks of S_1 and any vertex for a treatment contained in blocks of S_2.

(ii) A Type 2 RROS is any observation set which contains all observations from one or more whole blocks of D. A *simple* Type 2 RROS consists solely of the observations from a single block of D. There are b simple Type 2 RROSs altogether with common cardinality k, and every Type 2 RROS contains one or more simple Type 2 RROSs. If the observation set lost from D is a Type 2 RROS then D_e has fewer blocks than D, i.e. $b_* < b$.