Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 1075 – 1088
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Blocking response surface designs

P. Goos[a,*], A.N. Donev[b]

[a]Department of Mathematics, Statistics and Actuarial Sciences, Faculty of Applied Economics, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium
[b]AstraZeneca, Mereside, Alderley Park, Macclesfield, SK10 4TG,UK

Received 16 February 2005; received in revised form 10 November 2005; accepted 10 November 2005
Available online 5 December 2005

Abstract

The design of experiments involving more than one blocking factor and quantitative explanatory variables is discussed, the focus being on two key aspects of blocked response surface designs: optimality and orthogonality. First, conditions for orthogonally blocked experiments are derived. Next, an algorithmic approach to compute *D*-optimal designs is presented. Finally, the relationships between design optimality and orthogonality in the context of response surface experiments are discussed in detail.

Keywords: Exchange algorithm; *D*-optimality; Fixed blocks; Random blocks; Orthogonality

1. Introduction

In this paper, we focus on the construction of response surface designs in blocks. Blocking is usually beneficial in experimental situations where it is possible to identify groups, or blocks, of experimental units, such that within blocks the experimental units are considerably more homogeneous than the blocks themselves. The present article considers experiments in which more than one source of heterogeneity is present and extends earlier work by Atkinson and Donev (1989), Cook and Nachtsheim (1989), Khuri (1992) and Goos and Vandebroek (2001). The variation between the blocks in the experiment is accounted for by including block effects in the statistical model. In most applications, as in the present paper, the block effects are considered to be nuisance parameters and the accuracy of their estimation is not important.

Depending on what randomisation has been used to form the blocks, their effects could be regarded as *random* or *fixed*. Typically, the blocks are considered random when they are selected from a population at random. However, the implementation of such a selection is often impossible or impractical. In such cases the block effects are usually considered fixed and treated as levels of one or more qualitative variables whose effects are not of interest. Some authors, e.g. Gilmour and Trinca (2000), though mention the possibility of treating the block effects as random as soon as the block labels are randomly assigned to the blocks, even when the block effects are not a random sample from a population. The issue of choosing between random and fixed blocks is also discussed by Ganju (2000).

* Corresponding author. Tel: +32 32 204 059; fax: +32 32 204 817.
E-mail address: peter.goos@ua.ac.be (P. Goos).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2005.11.003

In any case, the nature of the blocking variables has an important impact on the data analysis and therefore on the optimal design for blocked experiments. These topics receive detailed attention in Khuri (1992, 1994), Ganju (2000), Ganju and Lucas (2000) and Gilmour and Trinca (2000), who discuss the analysis of experiments with one blocking variable, and Goos (2002).

In many practical applications the experimenter may deal with both fixed and random blocking variables. Here are two examples.

Example 1. *Valve wear experiment*: In order to establish a measurable valve wear in an internal combustion engine, the latter would need to be run for a long time, typically more than 1000 h. The study required the effect of several variables on the valve wear to be modelled. Some of these variables were parameters of the engine setup, while others were related to the valves themselves (material, dimensions, coatings, etc.).

The wear characteristics of different cylinders in the engines were known to be different in a consistent way, e.g. the end cylinders run cooler than centre cylinders. Also, the scientists believed that the wear of a valve in one cylinder had a negligible effect on the wear of the valves in the other cylinders in the tested operating conditions. In the study, 6 cylinder engines were used, so there were 6 valve positions. Running different valve parameters for each of the 6 cylinders opened the prospect of a six-fold reduction of the total number of engine-hours of testing. This brought in the valve position as a first blocking variable. Although there were 6 different valve positions, the corresponding blocking variable acted at only three levels because valves at equal distances from the centre were similar. Another way to shorten the time of experimentation was to use several engines. This introduced a second blocking variable acting at as many levels as the number of engines used.

Example 2. *Food additives experiment*: An important problem faced by the research laboratory of a food additives producer was to find out how the yield of a starch extraction process depends on the water content of the dough, the flour extraction rate (or milling rate) and the temperature. For each observation in this experiment the raw material, wheat, was milled until the desired flour extraction rate (between 70% and 80%) was reached. Then, water was added after which the gluten in the dough started to agglomerate. The water content was measured by the water/flour ratio, which lay between 0.6 and 1.2. Dough was prepared at different temperatures between 10°C and 40°C. After some time, the gluten were extracted by sieving the mixture. The yield of this process was the percentage of gluten recovered from the wheat. In order to increase the yield of the process, enzymes were added to the dough. The producer used several suppliers of enzymes and the wheat came in different batches. The variation between the enzymes received from different suppliers, as well as between batches of wheat was considerable. Thus, there were two blocking variables in the experiment: suppliers and batches.

The complexity of the design problems in the examples necessitates the use of an algorithmic approach to construct optimum experimental designs for them. Several algorithms for the construction of optimally blocked response surface designs involving one blocking variable have been described in the literature. For instance, Atkinson and Donev (1989, 1992), Cook and Nachtsheim (1989), Miller and Nguyen (1994) and Trinca and Gilmour (2000) discuss the construction of designs with a fixed blocking variable. Goos and Vandebroek (2001) present an algorithm for computing optimal designs in the presence of one random blocking variable. Goos et al. (2005) review the optimal design of blocked experiments. Despite the huge interest in blocked designs, the case when there are several blocking variables has not received much attention. Exceptions are Gilmour and Trinca (2003) who discuss the row–column arrangement of factorial and composite designs, and Ankenman et al. (2003) who study the case where random blocks are formed by two nested factors.

In this paper, we present an algorithm that can be used to generate optimum designs for situations where fixed or random blocks are generated by several crossed blocking variables and where the block structure is dictated by the experimental situation. In addition, the conditions for orthogonally blocking response surface designs are extended to cases with two or more blocking variables and the trade-off between orthogonality and D-optimality is illustrated.

In the next section, the statistical model and the notation are introduced. The conditions for orthogonal blocking are derived in Section 3. The design construction algorithm is described in detail in Section 4. Finally, the algorithm is applied to an interesting problem involving two blocking and two quantitative experimental variables.