Computational Statistics and Data Analysis 115 (2017) 79–90

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Bayesian *D*-optimal screening experiments with partial replication

Robert D. Leonard [a], David J. Edwards [b,*]

[a] Department of Information Systems and Analytics, Miami University, 800 E. High St., Oxford, OH 45056, United States
[b] Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, 1015 Floyd Avenue, Richmond, VA 23284, United States

ARTICLE INFO

Article history:
Received 23 August 2016
Received in revised form 4 April 2017
Accepted 26 May 2017
Available online 9 June 2017

Keywords:
Bayesian *DP*-optimal
Model misspecification
Partial replication
Potential terms
Primary terms
Pure error

ABSTRACT

Screening designs are frequently used in the initial stages of experimentation with the goal of identifying important main effects as well as to gain insight on potentially important two-factor interactions. Commonly utilized experimental designs for screening are unreplicated and as such, provide no unbiased estimate of experimental error. However, if statistical inference is to be performed as part of the experimental analysis, one view is that inferential procedures should be performed using a model independent error estimate instead of the residual mean square from the fitted model. As full replication of an experiment may be quite costly, partial replication offers an alternative. Gilmour and Trinca (2012) introduce criteria for constructing optimal designs for statistical inference (and hence, provide for optimal selection of replicate design points). An extension of their work is introduced by modifying the popular Bayesian *D*-optimality criterion to construct partially replicated screening designs with less dependence on an assumed model. Designs are compared using various criteria and a simulation study is conducted to investigate design performance with respect to power and false discovery rates.

1. Introduction

Screening designs are common in the early stages of experimentation with the goal of identifying factors having an effect on the system/process under study. Typical screening designs are unreplicated strength 2 or 3 orthogonal arrays as well as designs constructed using variance-based optimality criteria. One of the most popular and widely used criteria for design construction is *D*-optimality, defined as maximizing the determinant of the information matrix $\boldsymbol{X}'\boldsymbol{X}$, where $\boldsymbol{X}$ is the $n \times p$ model matrix. The formulation is attributed to the use of the determinant to calculate volumes of high dimensional confidence ellipsoids.

For the usual linear model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ (where $\boldsymbol{y}$ is the $n \times 1$ response vector, $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown model parameters, and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 I)$), it is well known that the variance–covariance matrix of the least squares estimates $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$ is $Var\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. Wald (1943) showed that maximizing $\left|\boldsymbol{X}'\boldsymbol{X}\right|$ leads to minimization of the volume of the joint confidence ellipsoid (i.e., our uncertainty) concerning $\boldsymbol{\beta}$. Thus, a *D*-optimal design, ξ^*, satisfies

$$\left|\boldsymbol{X}\left(\xi^*\right)'\boldsymbol{X}\left(\xi^*\right)\right| = \max_{\xi \in \boldsymbol{\chi}} \left|\boldsymbol{X}(\xi)'\boldsymbol{X}\left(\xi\right)\right|,$$

over all possible designs, ξ, spanning some design space $\boldsymbol{\chi}$.

* Corresponding author.
E-mail addresses: leonarrd@muohio.edu (R.D. Leonard), dedwards7@vcu.edu (D.J. Edwards).

http://dx.doi.org/10.1016/j.csda.2017.05.014
0167-9473/

As commonly utilized designs for screening are unreplicated, they provide no unbiased estimate of experimental error. However, if statistical inference is considered an integral part of the experimental analysis (e.g., testing treatment parameters, checking for lack-of-fit), one view is that inferential procedures should be performed using a pure error estimate (obtained via replicated design points) versus the residual mean square from the fitted model. Gilmour and Trinca (2012) (henceforth, GT) adopt this point of view and note that since bias in the fitted model residual mean square is unknown, inferences performed may be less informative and more challenging to interpret. While disagreement in this regard naturally persists (see, e.g., Smucker, 2012), this article is not an attempt to settle the dispute. Rather, we proceed under GT's viewpoint and seek screening designs that offer the ability to obtain a pure error estimate of σ^2.

Since full replication of an experiment may be quite costly, partial replication offers a cost-saving alternative for obtaining a model independent error estimate. The construction of partially replicated designs is not a prevalent topic in the literature; research in this area does exist, however. Lupinacci and Pigeon (2008) developed a class of partially replicated two-level designs based on Hadamard matrices. For an n-run experiment, their construction method always produces designs with $n/4$ replicates. Liao and Chai (2009) consider partial replication by constructing parallel flats designs with two identical flats. This idea was also employed by Ou et al. (2013) who utilized semifoldover techniques for design construction. Other recent work includes Tsai et al. (2012) and Tsai and Liao (2014). However, in all of these cases, the choice of the number of replicate points is determined absent optimality considerations.

A flexible strategy to construct partially replicated screening designs is to use standard exchange algorithms driven by the optimization of a specified criterion. GT use such an approach based upon Draper and Smith's (1998, pg. 144) formulation for the volume of a confidence interval region for $\boldsymbol{\beta}$ which is proportional to

$$\left(F_{p,d;1-\alpha}\right)^{p/2}\left|\boldsymbol{X}'\boldsymbol{X}\right|^{-1/2},$$

where p is the number of parameters in the assumed model, d is the pure error degrees of freedom, and $F_{p,d;1-\alpha}$ is the $(1-\alpha)$-quantile of the F-distribution. The so-called *DP*-optimal design is then defined as that which maximizes

$$\left|\boldsymbol{X}'\boldsymbol{X}\right| / \left(F_{p,d;1-\alpha}\right)^{p}. \tag{1}$$

As inference is rarely performed on the intercept, it can be treated as a nuisance parameter and hence, the *DP*-optimal design maximizes

$$\left|\boldsymbol{X}_s'\boldsymbol{Q}\boldsymbol{X}_s\right| / \left(F_{p-1,d;1-\alpha}\right)^{p-1}, \tag{2}$$

where $\boldsymbol{X}_s$ is the model matrix without a column of ones, $\boldsymbol{Q} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{J}$, $\boldsymbol{I}$ is the $n \times n$ identity matrix, and $\boldsymbol{J}$ is the $n \times n$ matrix of ones. Throughout this article, we assume the typical $\alpha = 0.05$, although other choices could be considered.

GT construct optimal designs using a standard candidate exchange algorithm which permits designs of any given run size as long as the assumed model is estimable. However, designs based upon (1) or (2) tend to have an excessive number of replicate runs, and are thus inefficient regarding the use of available degrees of freedom in screening situations. To overcome this potential deficiency and encourage the availability of degrees of freedom for testing lack-of-fit, GT propose to incorporate degree of freedom efficiency (i.e., the proportion of experimental runs used to estimate factorial effects, $df_{eff} = (n-d)/n$) into the formulation of a compound criterion. For example, combining the *DP*-criterion (2) with degrees of freedom efficiency, we maximize

$$\frac{\left|\boldsymbol{X}_s'\boldsymbol{Q}\boldsymbol{X}_s\right|^{\frac{\kappa_1}{p-1}}(n-d)^{\kappa_2}}{F_{p-1,d;1-\alpha}^{\kappa_1}}, \tag{3}$$

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$ are weights (specified by the experimenter) to reflect the relative importance of the individual criteria composing (3). For instance, $\boldsymbol{\kappa} = (1, 0)$ reduces the compound criterion to (2) whereas $\boldsymbol{\kappa} = (0.5, 0.5)$ specifies equal weight to both *DP*-efficiency and availability of degrees of freedom for lack-of-fit. Despite this modification, design construction still critically depends on the choice of an *a priori* model.

The remainder of this article proceeds as follows. In Section 2, we propose a Bayesian modification to GT's *DP*-optimality criteria for constructing partially replicated screening designs. In Section 3, we compare designs of various run sizes using measures of efficiency and model misspecification. Consideration is given to multiple criteria for design selection in Section 4. Section 5 details the results of a small simulation study and Section 6 provides concluding remarks.

2. Bayesian *D*-optimal partially replicated screening designs

A known drawback to the traditional optimal design approach is the required assumption that the model is known. To help *D*-optimal designs overcome dependency on an assumed model, DuMouchel and Jones (1994) introduced a simple Bayesian modification which involves classifying factorial effects of interest as either *primary* or *potential*. Primary terms are those effects assumed to be active while potential terms may or may not be active. In this way, the design can be chosen such that the primary terms are precisely estimated while still providing some information on potential terms. Furthermore, due to the information provided by the prior distribution on the model parameters, the Bayesian *D*-optimal approach circumvents the singularity problem of $\boldsymbol{X}'\boldsymbol{X}$ in situations where $n < p$.