Available online at www.sciencedirect.com

SCIENCE DIRECT®

Reliability Engineering and System Safety 91 (2006) 1281–1289

RELIABILITY ENGINEERING & SYSTEM SAFETY

www.elsevier.com/locate/ress

Combined array experiment design

L.M. Moore*, M.D. McKay, K.S. Campbell

Statistical Sciences Group D-1, MS F600, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Available online 26 January 2006

Abstract

Experiment plans formed by combining two or more designs, such as orthogonal arrays primarily with 2- and 3-level factors, creating multi-level arrays with subsets of different strength are proposed for computer experiments to conduct sensitivity analysis. Specific illustrations are designs for 5-level factors with fewer runs than generally required for 5-level orthogonal arrays of strength 2 or more. At least 5 levels for each input are desired to allow for runs at a nominal value, 2-values either side of nominal but within a normal, anticipated range, and two, more extreme values either side of nominal. This number of levels allows for a broader range of input combinations to test the input combinations where a simulation code operates. Five-level factors also allow the possibility of up to fourth-order polynomial models for fitting simulation results, at least in one dimension. By having subsets of runs with more than strength 2, interaction effects may also be considered. The resulting designs have a "checker-board" pattern in lower-dimensional projections, in contrast to grid projection that occurs with orthogonal arrays. Space-filling properties are also considered as a basis for experiment design assessment.

Keywords: Computer experiments; Experiment design; Fractional factorial design; Orthogonal arrays; Correlation coefficient; Space-filling design; Maximin distance

1. Introduction

The context for this paper is planning runs of a non-stochastic computer code for the purpose of assessing important inputs from among p inputs. As in McKay [1], the goal of sensitivity analysis is to identify "important" input(s) and this is done based on comparison of R^2, an estimate of the correlation coefficient associated with the goodness of fit to the computer code output Y of an analysis of variance model based on a subset of inputs $\boldsymbol{X}_s$. A subset of inputs $\boldsymbol{X}_s$ is considered more "important" than another if its corresponding R^2 is larger. The following is a formula for R^2 based on a subset of inputs $\boldsymbol{X}_s$:

$$R^2(X_s) = \frac{\sum_{i\in X_s}\sum_j (y_{i\cdot} - y_{\cdot\cdot})^2}{\sum_{i\in X_s}\sum_j (y_{ij} - y_{\cdot\cdot})^2},$$

where the subscript i varies over distinct cases of values of the s inputs identified in $\boldsymbol{X}_s$, the subscript j varies over "replicate" experiments corresponding to a fixed value of the inputs $\boldsymbol{X}_s$, and the "dot" subscript indicates the standard average. The summation over j, in both the numerator and denominator, depends on the number of actual observations ("replicates") and may differ for each of the i distinct values of the s inputs identified in $\boldsymbol{X}_s$. Although the summand in the numerator of this expression does not depend on j, the notation emphasizes $R^2(X_s)$ as a relative comparison of the sum of squares replacing the data with a regression predictor (numerator) and the total sum of squares (denominator). Here the regression is non-parametric in that no particular relationship is assumed between the response and s inputs other than a different mean value for each distinct value of the s inputs that occur in X_s. This is exactly the setting of (unbalanced) analysis of variance with treatment classes defined by partitioning the responses according to the distinct values of X_s. Different subsets of the full set of p parameters may define different partitions of the response values.

Fig. 1 illustrates an assessment of relative importance of two inputs based on the R^2 value associated with each of the inputs, referred to as factors A and B in the figure. All of the response values are plotted on the center horizontal

*Corresponding author.
E-mail address: lmoore@lanl.gov (L.M. Moore).

0951-8320/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2005.11.024

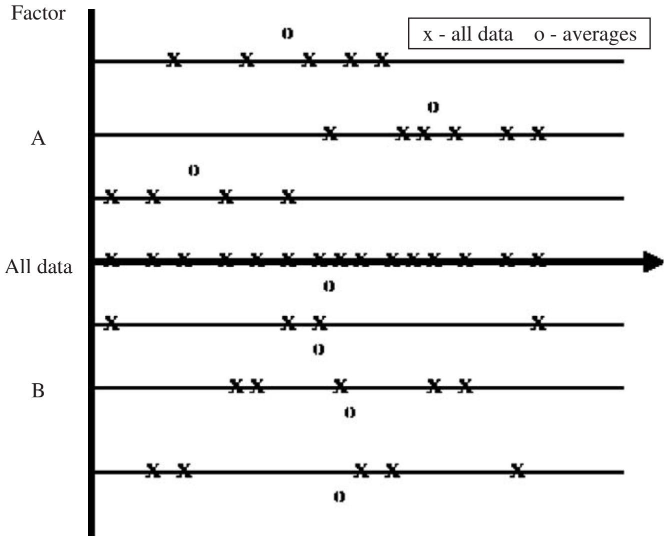

Fig. 1. A data set illustrating R^2(A) > R^2(B). All data is partitioned according to the value of Factor A above the horizontal axis, or Factor B below the horizontal axis.

axis. The sum of squares of these values adjusted by their mean value, labeled by "o" just under the axis, is the denominator of the formula for R^2. Above the horizontal axis, the response values are partitioned according to the values (3 in this case) of factor A and are marked with an "x". Within each of the resulting partitions, the average is marked above the set of partitioned values with an "o". The numerator of the formula for $R^2(A)$ is a weighted sum of the squares of the averages of the partitioned values adjusted by the grand mean for all of the data. The number of response values in each partition determines the weights. Below the axis, the response values, marked with an "x" are partitioned according to the value of factor B with the partitioned sets having means labeled "o" below the sets. The numerator of $R^2(B)$ is also calculated as the weighted sum of the squares of the averages of the partitioned values adjusted by the grand mean for all of the data. It is clear that the adjusted mean values of the factor A partitioned sets are more spread out, or there squares will be bigger than those for the factor B partitioned sets, so that $R^2(A)$ will be larger than $R^2(B)$. Factor A, as a predictor accounts for more of the total variance of the responses than factor B and would be deemed a relatively more influential input.

The experiment-planning problem fundamentally is to plan to collect measurements that will meet the needs of a planned analysis, ideally as efficiently as possible. For sensitivity assessment based on R^2, and from considering the illustration in Fig. 1, "replicate" measurements are needed for a set of values of each of the inputs. "Replicate" is in quotes since no true replicates are done. The computer simulation output is non-stochastic in that the output is fully determined by specification of the input with no variation in output for repeated runs of the code for identical input. Variation in the output is induced solely by variation in the inputs. However, the ($p-s$) inputs, identified by X_{-s} and associated with all inputs excluding those identified in X_s, may differ while X_s is fixed. These constitute "replicate" runs for a fixed value of X_s. The value $y_{i\cdot}$ is identically y_{ij} if there are no "replicate" runs. If this is the case for every value of the inputs identified by X_s, then R^2 has value identically 1 and is not very useful for identifying a subset of important inputs. Otherwise, R^2 is between 0 and 1. The desire to identify subsets of inputs that are important leads to considering experiment designs such that, for subsets of inputs of a specified size $s<p$, a sampling of values for that subset of inputs is required and "replicates" determined by a sample of values for the remaining inputs occur, for at least one of the values of the size s subset of inputs. This is a property of factorial experiment designs, or orthogonal arrays, which naturally suit this analysis approach, per Moore and McKay [2].

Factorial experiments are experiments for inputs, called factors, with a finite number of discrete values, referred to as levels, so if each input has K levels and there are p inputs then there are K^p possible distinct runs referred to as the K^p factorial design space. The K levels could be associated with K equal probability content intervals for a continuous input. If the experiment plan consisted of the entire K^p factorial design space, then for each pair of inputs (subsets of size $s=2$) there are K^2 values (levels) with K^{p-2} "replicates" for each value. This extends to subsets of inputs of size s in an obvious way. For relatively moderate K and even small sizes for p the full product space of possible experiment runs quickly becomes unmanageably large, even given the ability to run the simulation code thousands of times. In this paper, inputs with at least $K=5$ levels are desired.

Orthogonal array experiment designs are subsets of full factorial designs, also referred to as fractional factorial designs, with reduced runs obtained by relaxing the property that for any size, $s<p$, subset of inputs there are "replicate" inputs for each value of the subset. Wu and Hamada [3] and Hedayat et al. [4], are good references on orthogonal arrays, in addition to several classic texts on statistical experiment design and fractional factorial experiments by John [5] and Raktoe et al. [6]. For K levels identified by elements in the set $L=\{0,1,2,\ldots,k-1\}$, an $N\times p$ array $\boldsymbol{X}$ with entries from L is an orthogonal array with K levels, strength t ($0\leqslant t\leqslant p$) and index λ if every $N\times t$ subarray of $\boldsymbol{X}$ contains each t-tuple based on L exactly λ times as a row. An array with parameters N (number of runs), p (number of factors), k (number of levels for each factor), and t (strength) is denoted OA(N,p,k,t). From this definition, a strength t orthogonal array with index λ is a set of p-dimensional factorial design points such that if one considers any t-dimensional projection then every point in the K^t factorial design space is replicated λ times. Likewise, any projection of dimension smaller than t, say $s<t$, consist of $\lambda K^{(t-s)}$ replicates of the K^s factorial design space. A full K^P factorial design space is itself an OA(K^p, p, K, p) with index unity, that is $\lambda=1$. In a strict sense, fractional factorial designs may be any subset