FRACTIONAL REPLICATION IN SIMULATION STUDIES

Walter T. Federer
Biometrics Unit, Cornell University, Ithaca, NY 14853

Key Words and Phrases: Main effect plans; saturated fractions; D-optimal; D-minimal.

ABSTRACT

It is shown how to use fractional replication in simulation studies. Examples are given. Considerable savings in number of runs required can be achieved through the use of fractional replication ideas.

1. INTRODUCTION

Most statisticians will, at one time or another, be involved with a simulation study of the behavior of an estimator or statistical procedure. The study will usually be done for several values of the parameters involved. For example, for the normal distribution, the size of the mean μ and variance σ^2 separately might be of interest to the investigator. If five levels of μ and six values of σ^2 are to be studied, a total of $5{\times}6{=}30$ combinations would be required. The investigator might be willing to assume that the interaction was zero or negligible and only main effects for μ and σ^2 were of importance. The following one-third fraction of ten observations could be used as a saturated main effect plan: $\mu_1\sigma_1^2$, $\mu_2\sigma_1^2$, $\mu_3\sigma_1^2$, $\mu_4\sigma_1^2$, $\mu_5\sigma_1^2$, $\mu_1\sigma_2^2$, $\mu_1\sigma_3^2$, $\mu_1\sigma_4^2$, $\mu_1\sigma_5^2$, and

$\mu_1\sigma_6^2$ where μ_i is the ith level of μ_i and σ_j^2 is the jth level of σ^2. Thus, only one-third of the effort needs to be used to obtain information about levels of μ and of σ^2 as they affect the procedure under study. When many parameters (factors) and several levels are involved, the total number of combinations can become large. If interactions are nonexistent or negligible, then a saturated main effect plan would be called for and a considerable savings in number of points at which simulations were run could be achieved.

To illustrate, Grimes and Federer (1984) were interested in simulating results for generalized versions of the Behrens-Fisher test. Interest centered on closeness of procedures to the nominal level α where α is the stated size of the test. They were interested in the following four factors:

(i) the value of $\sum_{i=1}^{v} c_i\mu_i$ where μ_i is a parameter value for treatment i and c_i is a contrast coefficient subject to the constraint $\sum_{i=1}^{v} c_i = 0$,

(ii) population variances, σ_i^2,

(iii) sample size, and

(iv) nominal significance levels.

If each of the above factors were at four levels, a total of $4^4 = 256$ combinations would be required. Instead of a simulation at each of the 256 combinations, an orthogonal main effect plan with 16 combinations was selected. This one-sixteen fraction may be obtained from a complete set of orthogonal latin squares of order four or from an orthogonal array with 16 runs (columns), five rows, four symbols, and of strength two. Such a set would be

	Combination(Column)			
Factor	Level of factor			
A	0000	1111	2222	3333
B	0123	0123	0123	0123
C	0123	1032	2301	3210
D	0123	3210	1032	2301
E	0123	2301	3210	1032

The above is for five factors but only four were under study. Hence, the sum of squares for factor E can be used as an error sum of squares since all levels of factor E were the same. This would be true if interactions among the four factors were zero. Otherwise, the three degrees of freedom for factor E could be used to measure lack of fit from a main effect model if an independent estimate of the error variance were available.

It should be noted that *ANY* fraction and *ANY* set of parameters in a factorial can be constructed. The D-minimal design (see Anderson and Federer, 1973 and 1975), which is one with the minimal nonzero value of the determinant of the information matrix, can be constructed for any fraction and any factorial by one version of the one-at-a-time plan. For each factor separately, the range of values is used and the levels of all other factors are held constant, say at the lowest level. The procedure is illustrated for a $2\times3\times4$ factorial where a main effect plan would yield the seven combinations 000, 100, 010, 020, 001, 002, 003. One further item to note is that if one uses single degree of freedom contrasts, each observation adds one degree of freedom as in the ANOVA table on the next page for a $2\times3\times4$ factorial given in Table 1. Any subset in any order of these 24 single degree of freedom contrasts may be used, depending upon the goal of the investigator. One could select parameters A, B_1, B_2, C_1, C_2, C_3, and B_1C_1, A B_1C_1 as the parameters and the corresponding observations as the fractional replicate plan. The ordering of the parameters is not binding on the investigator.

In simulation studies, it is not clear to the author that a D-optimal design is as important as it is in quality control and product improvement studies. Hence, it may be that a D-minimal design would suffice in simulation studies. It should be noted that there can be a considerable difference in variance between a D-minimal and a D-optimal design. For example, Anderson and Federer (1975) found that for a saturated main effect plan for 11 factors at two levels each, the ratio of D-maximal to D-minimal was 1458. However, if σ^2 is very small, this large difference may not matter. This idea needs further exploration before reaching a conclusion. It is well known that a different random seed should be used for each combination.

TABLE 1

Source of variation	Degrees of freedom		Added combination
Mean	1		000
A (first factor)	1		100
B (second factor)	2		
B_1		1	010
B_2		1	020
C (third factor)	3		
C_1		1	001
C_2		1	002
C_3		1	003
$A \times B$	2		
$A \times B_1$		1	110
$A \times B_2$		1	120
$A \times C$	3		
$A \times C_1$		1	101
$A \times C_2$		1	102
$A \times C_3$		1	103
$B \times C$	6		
$B_1 \times C_1$		1	011
$B_1 \times C_2$		1	012
$B_1 \times C_3$		1	013
$B_2 \times C_1$		1	021
$B_2 \times C_2$		1	022
$B_2 \times C_3$		1	023
$A \times B \times C$	6		
$A \times B_1 \times C_1$		1	111
$A \times B_1 \times C_2$		1	112
$A \times B_1 \times C_3$		1	113
$A \times B_2 \times C_1$		1	121
$A \times B_2 \times C_2$		1	122
$A \times B_2 \times C_3$		1	123

If a replication of the whole set is needed, it is suggested that a different main effect plan be used for each complete replicate. For example, in the above $2 \times 3 \times 4$ factorial one could use the first seven observations for replicate one. Then for replicate two, use the seven combinations 123, 023, 113, 103, 122, 121, and 120 which were obtained by interchanging 0 and 1 for the first factor, 0 and 2 for the second factor, and by interchanging 0 and 3 and 2 and 1 for factor three. This is a fold-over of the original design and results in seven different combinations than were used for replication one. There are many methods of constructing fractional replicates of a complete factorial. Some of these are discussed in Raktoe *et al.* (1981) and others are discussed in the references cited by them.

Although there is a large literature on fractional replication, most statisticians contacted by the author have little or no interest in the subject, and most have only a prefunctory knowledge of the topic at best. They apparently have failed to realize the usefulness of fractional replication in simulation studies, and consequently little use has been made of fractional replication for these studies.

BIBLIOGRAPHY

Anderson, D. A. and W. T. Federer. (1973). Representation and construction of main effect plans in terms of zero-one matrices. No. BU-499-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY. (Revised and being processed as a publication.)

Anderson, D. A. and W. T. Federer. (1975). Possible absolute determinant values for square (0,1)-matrices useful in fractional replication. *Utilitas Mathematica 7*, 135-150.

Grimes, B. A. and W. T. Federer. (1984). Comparison of means from populations with unequal variances. In *W. G. Cochran's Impact on Statistics* (editors: P.R.S. Rao and J. Sedransk), John Wiley and Sons, Inc., New York, pp. 353-374.

Raktoe, B. L., A. Hedayat, and W. T. Federer. (1981). *Factorial Designs*, John Wiley & Sons, Inc., New York, xii + 269 pages.