ON THE IDENTITY RELATIONSHIP FOR FRACTIONAL REPLICATES OF THE 2^n SERIES[1]

By R. C. Burton and W. S. Connor

National Bureau of Standards

1. Summary. The paper considers $\frac{1}{2^r}$ fractional replication designs of the factorial series with n factors each at two levels. The identity relationship for such designs is often written in terms of a symbol I and collections of letters which denote interactions among the factors. These collections may conveniently be called "words," and the number of letters in a collection, the "length" of the word. The problem considered is that of the existence of an identity relationship which contains words of specified lengths.

It is known that the words of an identity, together with the symbol I, form an Abelian group. The group contains sets of independent generators, and the products of such generators. Necessary and sufficient conditions are developed for the existence of an identity relationship for which the lengths of a set of independent generators and their products are specified. Further, it is shown how to construct such an identity relationship, and it is proved that the identity relationship is unique, apart from renaming the letters.

For the more general case in which the lengths of the words are given—but are not associated with particular generators and products—a necessary condition is developed for the existence of the identity relationship. It is shown by example that this condition is not sufficient.

2. Introduction. We shall consider the case of n factors each at two levels. Let the factors be denoted by $A, B, \cdots$ and consider a $\frac{1}{2^r}$ fraction of the 2^n factorial design ($n > r$). The identity relationship consists of the symbol I and $2^r - 1$ words, connected by equality signs, as follows:

$$I = A^{a_1}B^{b_1} \cdots = \cdots = A^{a_m}B^{b_m} \cdots,$$

where $m = 2^r - 1$; $a_i, b_j, \cdots$ $(j = 1, \cdots, m)$ take on the values 0 or 1; and $A^0 = B^0 = \cdots = 1$. We shall consider the case in which all n letters are present in the identity relationship.

The product of any two words,

$$A^{a_x}B^{b_x} \cdots \quad \text{and} \quad A^{a_y}B^{b_y} \cdots,$$

is $A^{a_x+a_y}B^{b_x+b_y} \cdots$, where $a_x + a_y$, $b_x + b_y$, $\cdots$ are reduced modulo 2; and the product of any word with I is the word itself. The words of the identity relationship, together with I, form an Abelian group, in which each word is its own inverse.

Received September 4, 1956.

[1] Work performed (in part) under the sponsorship of the Bureau of Ships, Navy Department.

For a $\frac{1}{2^r}$ fraction, the group contains r words which are independent generators. The product of any i ($i = 2, \cdots, r$) of these generators is a word different from all of the generators. To illustrate, consider $n = 5$ and $r = 3$. An identity relationship is

$$I = \underline{ABC} = \underline{CDE} = \underline{AE} = ABDE = BCE = ACD = BD,$$

where a set of independent generators has been underscored.

Before proceeding with our main argument, it is helpful to review some known necessary conditions. For this purpose we shall let $w_1, \cdots, w_p$ denote the lengths of the words. Brownlee, Kelly, and Loraine [1] state that the following conditions are necessary:

(i) $\sum_{i=1}^{m} w_i = 2^{r-1}n$.
(ii) Either the w's all are even or 2^{r-1} of them are odd.
(iii) When 2^{r-1} of the w's are odd, the words with even numbers of letters must, with the identity I, form a subgroup of order 2^{r-1}.
(iv) If some $w = n$, the remaining w's must be divisible into pairs such that the total of each pair is n.
(v) If some $w = 1$, the remaining numbers must be divisible into pairs such that the numbers in each pair differ by 1.

They then state that "no other necessary conditions appear susceptible to expression in simple general terms."

These conditions do not require that the lengths of particular generators and their products be specified. The problem which we shall solve does make this requirement. We shall state necessary and sufficient conditions for the existence of an identity relationship, for which the lengths of the generators and their products are given. In addition, we shall prove that such an identity relationship is unique, apart from the naming of letters; and we shall show how to construct the identity relationship.

For the more general case considered by Brownlee, Kelly, and Loraine, we shall derive an additional necessary condition, and shall show by example that it, together with the above conditions (i), (ii), (iii), is not sufficient to imply the existence of an identity relationship.

In addition to [1], some papers which treat fractional replicates and the related problem of block confounding are listed as references [2], $\cdots$, [7].

3. The existence of an identity relationship with generators and products of specified lengths. Let $i_1, i_2, \cdots, i_s$ be s integers such that $0 < i_1 < i_2 < \cdots < i_s < r + 1$. The ith generator will be denoted by $W(i)$ and the product of the i_1-th i_2-th, $\cdots$ and i_s-th generator by $W(i_1, i_2, \cdots, i_s)$. There are exactly $2^r - 1$ words corresponding to the $2^r - 1$ symbols $(i_1, i_2, \cdots, i_s)$. The numbers of letters in the word $W(i_1 \cdots i_s)$ will be denoted by $w = w(i_1 \cdots i_s)$.

We shall find it convenient to introduce a symbol S to denote the entire collection of $2^r - 1$ symbols $(i_1 \cdots i_s)$, a symbol $O = O(i_1 \cdots i_s)$ to denote the collection of symbols which contain an odd number of indices from $(i_1 \cdots i_s)$, and a symbol $E = E(i_1 \cdots i_s)$ to denote the collection of symbols which contain none or an even number of indices from $(i_1 \cdots i_s)$.

Let $n(O)$ and $n(E)$ denote the numbers of symbols in O and E. It is readily shown that

$$n(O) = 2^{r-1}, \qquad n(E) = 2^{r-1} - 1. \tag{3.1}$$

We note that the distribution of a letter among the words of the identity is determined by its distribution among the generators. Suppose that a letter occurs in the s generators $W(i_1), \cdots, W(i_s)$, but not in the remaining generators. Then it occurs in all products which have an odd number of indices from among $i_1, \cdots, i_s$. But there are $2^r - 1$ ways in which a letter may be distributed among the generators, and hence among the words of the identity.

We shall use the symbol $t = t(i_1 \cdots i_s)$ to denote the number of letters which occur in all of the s generators $W(i_1), \cdots, W(i_s)$ but not in the remaining generators.

For example, consider three generators $W(1) = BCD$, $W(2) = ACDEF$, $W(3) = ACF$. We may illustrate with a Venn diagram which letters are common, denoting the generators by the interiors of the circles in Fig. 1.

Since the t's are the numbers of letters in the basic disjoint sets, it is obvious that

$$\textstyle\sum_S t(i_1 \cdots i_p) = n. \tag{3.2}$$

It is also clear that any set of t's which are positive integers or zero and satisfy (3.2) corresponds to a constructible identity relationship involving n letters.

We shall now show explicitly how the t's uniquely determine the w's, and conversely the w's uniquely determine the t's.

From the definitions of t, w, and O, it follows that

$$\sum_{O(i_1 \cdots i_s)} t(j_1 \cdots j_p) = w(i_1 \cdots i_s). \tag{3.3}$$

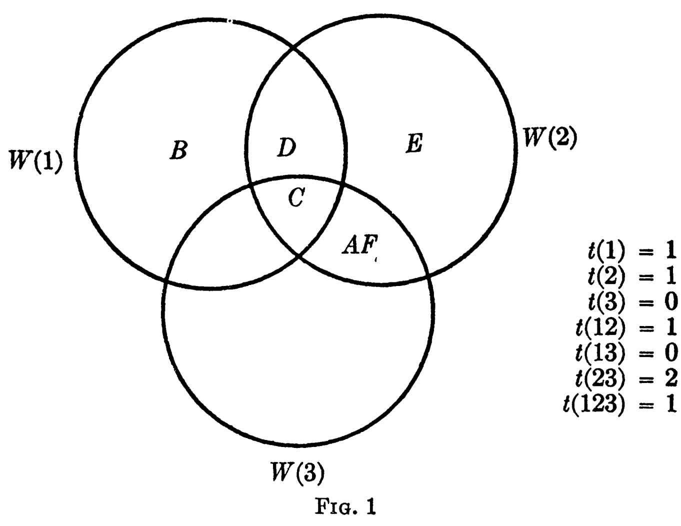

FIG. 1

There are $2^r - 1$ such equations. These equations uniquely determine the w's from the t's.

Let us now introduce a dummy variable $t(0)$, which we define to be identically zero. We may add $t(0)$ to the left-hand side of (3.2) to obtain

$$t(0) + \textstyle\sum_S t(j_1 \cdots j_p) = n. \tag{3.4}$$

Multiplying (3.3) by 2, and subtracting (3.4) from the product, we obtain

$$-t(0) + \textstyle\sum_O t(j_1 \cdots j_p) - \sum_E t(j_1 \cdots j_p) = 2w(i_1 \cdots i_s) - n, \tag{3.5}$$

where

$$O = O(i_1 \cdots i_s) \quad \text{and} \quad E = E(i_1 \cdots i_s).$$

We observe that the matrix of coefficients in (3.4) and (3.5) is a Hadamard matrix of order 2^r. Dividing these equations by $2^{r/2}$, the matrix of coefficients becomes an orthogonal matrix.

It is helpful to express the equations in matrix notation. Let t denote the column vector $[t(0), t(1), \cdots, t(12 \cdots r)]$, x the column vector $2^{-r/2}[n, 2w(1) - n, \cdots, 2w(12 \cdots r) - n]$, and C the matrix of coefficients in (3.4) and (3.5) after division by $2^{r/2}$. In this notation the equations may be written as

$$Ct = x. \tag{3.6}$$

Because C is orthogonal, $C^{-1} = C'$ and

$$C^{-1}Ct = C^{-1}x,$$

$$t = C'x. \tag{3.7}$$

The typical equation in (3.7) is

$$\textstyle\sum_O w(j_1 \cdots j_p) - \sum_E w(j_1 \cdots j_p) = 2^{r-1}t(i_1 \cdots i_s). \tag{3.8}$$

These last equations uniquely determine the t's from the w's. If the w's determine t's which are positive integers or zero, and are such that their sum is n, then the identity relationship can be constructed, and is unique, except for the renaming of letters.

We may sum up in the following theorem.

THEOREM 1. *For a $1/2^r$ fractional replicate of a 2^n factorial design, let the numbers of letters in the words of the identity relationship be $w(1), w(2), \cdots, w(r); w(12), w(13), \cdots, w(r-1, r); \cdots; w(12 \cdots r)$, where $(1), (2), \cdots, (r)$ refer to a set of independent generators and $(12), (13), \cdots (r-1, r); \cdots; (12 \cdots r)$ to their products. With respect to any one of these $2^r - 1$ symbols $(i_1 \cdots i_s)$, the remaining symbols divide into a collection $O = O(i_1 \cdots i_s)$ of 2^{r-1} symbols which have an odd number of indices in common with the given symbol, and a collection $E = E(i_1 \cdots i_s)$ of $2^{r-1} - 1$ symbols which have none or an even number of indices in common with the given symbol. If the w's satisfy the $2^r - 1$ equations*

$$\textstyle\sum_O w(j_1 \cdots j_p) - \sum_E w(j_1 \cdots j_p) = 2^{r-1}t(i_1 \cdots i_s),$$

where $\sum_O t + \sum_E t = n$ *in the sense of implying t's which are positive integers or zero, then the identity relationship exists. Conversely, if the identity relationships exists, then the equations are satisfied. Furthermore, knowledge of the t's is sufficient to construct the identity, and the identity corresponding to a set of t's is unique, apart from renaming the letters.*

4. A necessary condition. We now return to the more general problem of the existence of an identity relationship having words of given length, but without the specification of the lengths of particular generators and their products. We shall develop a necessary condition for the identity relationship to exist.

From (3.6) we have

$$t'C'Ct = x'x, \tag{4.1}$$

and because $C' = C^{-1}$,

$$\begin{aligned} t't &= x'x, \\ \sum w^2 &= 2^{r-2}(\sum t^2 + n^2). \end{aligned} \tag{4.2}$$

Thus, there must exist n or fewer positive integers—the non-zero t's—which add to n and satisfy (4.2).

We may consider the special case in which the variance of the w's is a minimum. The variance of w is defined by

$$\begin{aligned} V(w) &= [\sum w^2 - (\sum w)^2 / (2^r - 1)] / (2^r - 1) \\ &= 2^{r-2}[(2^r - 1)\sum t^2 - n^2] / (2^r - 1)^2. \end{aligned} \tag{4.3}$$

Now $V(w)$ is a minimum when $\sum t^2$ is a minimum—i.e., when $\sum t^2 = n$. In this case (4.2) becomes

$$\sum w^2 = 2^{r-2}n(n + 1). \tag{4.4}$$

We may sum up in a theorem and corollary.

THEOREM 2. *For a* $\frac{1}{2^r}$ *fractional replicate of a* 2^n *factorial design, let the numbers of letters in the words of the identity relationship be* $w_1, \cdots, w_m$, *where* $m = 2^r - 1$. *Then a necessary condition for the identity relationship to exist is that there exist n or fewer positive integers whose sum is n and whose squares add to* $2^{-r+2}\sum w^2 - n^2$.

COROLLARY. *If the variance of w is a minimum, then it is necessary that* $\sum w^2 = 2^{r-2}n(n + 1)$.

We shall show by example that this condition, together with conditions (i), (ii), (iii) are not sufficient conditions. Conditions (iv) and (v) do not apply in the example.

Let $n = 9$ and $r = 4$. Let the distribution of w's be as follows:

w	Frequency
4	7
5	6
7	2

For this distribution, $\sum w = 72$ and $\sum w^2 = 360$. This distribution has minimum variance, and satisfies the corollary. In addition, it satisfies (i), (ii), and (iii). To see that it satisfies (iii), we have only to write the following identity:

$$I = ABCD = ABEF = ACEG = CDEF = BDEG = BCFG = ADFG.$$

We shall show that the necessary conditions of Theorem 1 are not satisfied. To do this, we shall make a unique identification of the w's with the generators and their products.

Let us choose $w(1) = 7$ and $w(2) = 7$. Then because two odd lengths imply an even length, $w(12)$ must be 4. We choose $w(3) = 5$, which implies that $w(13) = w(23) = 4$, and $w(123) = 5$. Finally, we choose $w(4) = 5$, which completely determines the remaining products. The w's are as follows:

$w(1) = 7$	$w(12) = 4$	$w(123) = 5$	$w(1234) = 4$
$w(2) = 7$	$w(13) = 4$	$w(124) = 5$	
$w(3) = 5$	$w(14) = 4$	$w(134) = 5$	
$w(4) = 5$	$w(23) = 4$	$w(234) = 5$	
	$w(24) = 4$		
	$w(34) = 4$		

Now

$$\sum_{O(1)} w - \sum_{E(1)} w = 4$$

which by (3.8) implies that $t(1) = \frac{1}{2}$. Accordingly, the identity relationship does not exist.

REFERENCES

[1] K. A. Brownlee, B. K. Kelly, and P. K. Loraine, "Fractional replication arrangements for factorial experiments with factors at two levels," *Biometrika*, Vol. 35 (1948), pp. 268–276.

[2] O. Kempthorne, "A simple approach to confounding and fractional replication in factorial experiments," *Biometrika*, Vol. 34 (1947), pp. 255–272.

[3] D. J. Finney, "The fractional replication of factorial arrangements," *The Annals of Eugenics, London*, Vol. 12 (1945), pp. 291–301.

[4] R. A. Fisher, "The theory of confounding in factorial experiments in relation to the theory of groups," *Annals of Eugenics, London*, Vol. 11 (1942), pp. 341–353.

[5] C. R. Rao, "Hypercubes of strength 'd' leading to confounded designs in factorial experiments," *Bulletin Calcutta Math. Soc.*, Vol. 38 (1946), pp. 67–68.

[6] R. L. Plackett and J. P. Burman, "The design of optimum multifactorial experiments," *Biometrika*, Vol. 33 (1943–1946), pp. 305–325.

[7] R. C. Bose and K. A. Bush, "Orthogonal arrays of strength two and three," *Annals of Math. Stat.*, Vol. 23 (1952), pp. 508–524.