The Mathematics of Symmetrical Factorial Designs

Mausumi Bose

Mausumi Bose (nee Sen) obtained her MSc degree in Statistics from the Calcutta University and PhD degree from the Indian Statistical Institute. She is on the faculty of the Indian Statistical Institute. Her main field of research interest is design and analysis of experiments. She has held visiting positions at Purdue University, Ohio State University and University of California, San Diego. She is also interested in applied and consultancy work.

1. Introduction

R C Bose made pioneering contributions in several areas. His classic work in symmetrical factorial experiments has revolutionized the field. He initiated this work in 1940 (with Kishen) which culminated in his 1947 seminal paper on the mathematics of symmetrical factorials. He was the first to establish a link between the geometry of finite fields and combinatorial problems in the construction of designs. Since then the basic geometrical and algebraic techniques developed in this paper have been used in various contexts by researchers. A few years after his work was published, F W Levi's book on algebra came out where, with Bose's work in mind, he mentioned the utility of Galois fields in the construction of designs for agricultural experiments and remarked: "It is a startling idea that Galois fields might be helpful to provide people with more and better food."

2. Background

Experiments are carried out in almost all fields of scientific research and it is well known that 'statistical designing' of such experiments is indispensable because only data collected from a properly 'designed' experiment can lead to statistically valid conclusions about the objectives of the experiment. When a number of variables (factors) influence a character under study, it is always recommended that all the factors are investigated simultaneously. Such an experiment is called a factorial experiment and the design used for the experiment is a factorial design.

Each factor may be controlled at a number of levels and a symmetrical factorial considers the situation where

Keywords
Symmetrical factorial designs, designing confounded experiments, loss of information, balanced designs.

each factor has the same number of levels. A combination of the levels of all the factors is called a treatment combination. For example, an agricultural scientist may be interested in studying the effect of two fertilizers, potassium(K) and nitrogen(N), on the yield of a crop. Each fertilizer may be applied at three doses: high, medium or low. This is a symmetrical factorial experiment with the two fertilizers as the two factors, each factor is at three levels and there are in all 2^3 treatment combinations of the form: K-high N-high, K-high N-medium, K-low, N-low, etc.

Factorial experiments have their origin in agriculture, but they now have wide applications in many fields like biology, chemistry, animal husbandry, industry and so on.

Factorial experiments have their origin in agriculture, but they now have wide applications in many fields like biology, chemistry, animal husbandry, industry and so on. The aim of these experiments is to estimate the individual effect of each factor, called the main effect of the factor and also to study the inter-relationships of different factors, called their interaction effects. The possible presence of these interaction effects makes the factorial experiments interesting and they are more efficient than experiments where one factor is varied at a time.

In simplest factorial designs, all the treatment combinations are applied at random to experimental units and observations made. To minimize the differential effect that the units themselves may have on the response, it is imperative that all the units used in the experiment should be as similar (homogeneous) as possible.

Consider a symmetrical factorial experiment in n factors, each varied at s levels. It has s^n treatment combinations and is called an s^n experiment. For this, even if each combination is tested just once, one still needs s^n units to experiment with. Often, it may not be possible to get hold of these many homogeneous units, and in such cases, the set of available heterogeneous units is grouped into a number of subsets or blocks, in such

With an injudicious assignment of treatment combinations to blocks, important treatment effects may get confounded or mixed up with block effects and become non-estimable.

a way that, within each block the units are homogeneous. The block size, i.e. the number of units belonging to a block, is generally smaller than the number of all treatment combinations and in such cases one has to choose which treatment combinations should be assigned to which block. This choice is of utmost importance since with an injudicious assignment of treatment combinations to blocks, important treatment effects may get confounded or mixed up with block effects and become non-estimable.

Consequently, the designing of a factorial experiment becomes important and it requires a planned assignment of treatments to blocks so that all factorial effects of interest are estimable while unimportant effects may be confounded.

This problem of designing confounded experiments is a challenging mathematical problem and it is also extremely important from the point of view of application. R C Bose fully explored all the major aspects of this problem and came up with a general solution which works for all values of s which are prime or prime powers. In the next section we discuss Bose's contributions in this area and we assume a basic knowledge of finite geometry.

3. The Connection between Galois Fields and Factorial Experiments

In a factorial experiment, contrasts belonging to main effects and interactions are important, and in a s^n experiment, there are in all $(s-1)^m$ linearly independent contrasts belonging to a m-factor factorial effect, $m = 1, \ldots, n$.

Consider a Galois field (finite field) $GF(s)$ with s elements. An ordered set of n elements $(x_1, x_2, \ldots, x_n)$, where the x_i's are elements of $GF(s)$, is called a point of the finite Euclidean Geometry $EG(n, s)$. There are

s^n points in $EG(n;s)$. Any $(n-1)$-flat of $EG(n;s)$ has an equation of the form:

$$a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n = 0;\ a_i \in GF(s)$$

and it contains s^{n-1} points. By keeping $a_1; a_2; \dots; a_n$ constant and varying a_0 over the elements of $GF(s)$, we generate s parallel $(n-1)$-flats that have no common point. These s flats constitute a pencil, denoted by $P(a_1; a_2; \dots; a_n)$.

Bose considered an s^n factorial and identified the s levels of each factor with the s elements of $GF(s)$ and the s^n treatment combinations with the points in $EG(n;s)$. Using the definitions of main effects and interactions, he showed that the pencil $P(a_1; a_2; \dots; a_n)$ would represent the interaction of the i_1^{th}, i_2^{th}, $\dots$, i_n^{th} factors if and only if $a_{i_1}; a_{i_2}; \dots; a_{i_r}$ are non-zero and the other coordinates in the pencil are zero. He showed that the $(s-1)^{m-1}$ distinct pencils belonging to any m-factor effect give a representation for the treatment contrasts belonging to this effect in terms of $(s-1)^{m-1}$ mutually orthogonal sets of contrasts with $s-1$ linearly independent contrasts in each set. Then, using these pencils, he developed a method for constructing $(s^n; s^k)$ designs, i.e., designs for s^n experiments in s^k blocks of equal size.

Bose showed that the pattern of confounding in (s^n, s^k) designs can be obtained based on certain relationships in the *n*-dimensional finite projective geometry *PG*(*n*,*s*).

4. Designing $(s^n; s^k)$ Designs in Single Replicate with Total Confounding

Starting with an initial choice of k independent pencils, Bose showed how the s^n treatment combinations may be partitioned into s^k sets, each set consisting of s^{n-k} distinct combinations. Identifying each set with a block, the $(s^n; s^k)$ design is constructed and in this design, each of the initial k pencils and all their generalized interactions are totally confounded and cannot be estimated. Thus, there is complete loss of information on all these pencils.

So, the choice of the initial k pencils is of primary importance as they must be chosen in such a way that no factorial effect of interest gets confounded. The task of ascertaining whether this can be done, and if so, how, is a difficult one. Bose studied this problem in detail and for this, he exploited the connection between s^n designs and the n-dimensional finite projective geometry $PG(n; s)$ based on $GF(s)$. He showed that the pattern of confounding in $(s^n; s^k)$ designs can be obtained based on certain relationships in $PG(n; s)$. These results are very important as they allow one to enumerate all the possible confounding patterns in a given situation and then choose the best one for constructing the $(s^n; s^k)$ design for use.

5. Designing $(s^n; s^k)$ Design in Multiple Replicates with Partial Confounding

In some situations, the experimental resources allow the use of $r(> 1)$ replicates (copies) of a $(s^n; s^k)$ design instead of only a single replicate as in 3.1. The advantage of such situations is that a pencil may be confounded in some replicates, say, $r_1(< r)$ replicates and left unconfounded in the remaining$(= r - r_1)$ replicates. Such a pencil is called partially confounded as it will be estimable from the $r - r_1$ replicates where it is not confounded. The loss of information (L.I.) on this pencil will be partial, unlike the case of 3.1 where the L.I. on confounded pencils is total.

So, partially confounded designs are practically useful and Bose studied this problem in detail. He introduced the notion of balancing in $(s^n; s^k)$ designs. He called a design balanced if there is the same L.I. on all pencils belonging to effects involving the same number of factors. So, balanced designs are intuitively appealing and it can be shown that they also have good statistical properties. Using properties of projective geometry, Bose showed how designs could be constructed where all main effects have zero L.I. and balance is achieved

Suggested Reading

[1] R C Bose, Mathematical theory of the symmetrical factorial design, *Sankhya*, 8, pp.107-166, 1947.

[2] R C Bose, On some connections between the Design of Experiments and Information Theory, *Bulletin of International Statistical Institute*, 38, pp.257-271, 1961.

[3] R C Bose and K Kishen, On the problem of confounding in general symmetrical factorial design, *Sankhya*, 5, pp. 21-36, 1940.

[4] R C Bose, Autobiography of a mathematical statistician, *The Making of Statisticians*, ed. J Gani, Springer-Verlag, New York, 1982.

[5] R C Bose, and J N Srivastava, On a bound useful in the theory of factorial designs and error correcting codes. *Annals of Mathematical Statistics*, 35, pp. 408-414, 1964.

[6] A Dey and R Mukerjee, Symmetrical factorial experiments: A mathematical theory, in *Connected at infinity*, Ed. R Bhatia, Hindustan Publishing Company, New Delhi, pp.3-17, 2003.

over interactions of all orders. He also gave expressions for the L.I. on these interactions. These results are very useful for the users of factorial designs.

6. The Packing Problem

In factorial experiments, typically, the main effects and interactions involving a few factors are considered important. So, an important problem is that of finding the maximum number of factors n that may be accommodated in constructing a s^n design with a given constant block size and a given value of t, such that all effects involving t or less factors are unconfounded. This problem is known as the packing problem. Bose linked this problem with a finite projective geometry and made fundamental contributions in this area.

7. Impact of Bose's Work

Bose's work on symmetrical factorials had a tremendous impact on research on factorial designs and it also influenced other areas. His novel representation of factorial effects via pencils opened up a new direction in research in this area. His many contributions to the various aspects of the problem of confounding are mathematically elegant as well as practically useful. The geometric formulations developed by him in this connection have been found useful in other areas and his results on the packing problem led to advances in the theory of codes. Its effect was also felt in the study of orthogonal arrays which are combinatorially interesting and have wide applications in many areas. His work done in the 1940's continues to influence current research and recently, his ideas have led to the classes of fractional factorial plans called parallel flat fractions, regular fractions and minimum aberration designs. All these are important and very productive areas of current research and through them, the spirit of R C Bose lives on.

Bose's many contributions to the various aspects of the problem of confounding are mathematically elegant as well as practically useful. The geometric formulations developed by him in this connection have been found useful in other areas and his results on the packing problem led to advances in the theory of codes.

Address for Correspondence
Mausumi Bose
Applied Statistics Unit
Indian Statistical Institute
203, BT Road
Kolkata 700 108, India.