DISCUSSION ON THE MATHEMATICAL THEORY OF THE DESIGN OF EXPERIMENTS

By R. C. BOSE

Statistical Laboratory, Calcutta

The interest of mathematicians in combinatorial problems, involving the arrangement of a finite number of things, in sets or patterns, satisfying given conditions can be traced back to at least as far as Euler (1872), who interested himself in the construction of Latin and Graeco-Latin squares. Many other later mathematicians including Steiner also interested themselves in combinatorial problems. Nevertheless these questions have till recently remained only side-shows, as the major developments of Mathematics have been in other directions, dictated by the necessity of answering problems raised by other sciences, chiefly Physics and Astronomy.

But as has often happened in the history of Science, combinatorial problems which were hitherto considered only to be of academic interest, have suddenly revealed themselves to be of the greatest interest for the proper designing of biological experiments. This development has been mainly due to the work of Professor R. A. Fisher and his associates. Their work has however been held up by the absence of suitable mathematical methods for solving the combinatorial problems that arise. In to-day's discussion I wish to show to you, how methods of Finite Algebra and Finite Geometry, which had initially been developed by mathematicians for other purposes, can be made applicable to just those types of combinatorial problems which have arisen during the course of work of Fisher, Yates and others. For this purpose I shall select two special problems.

The first problem about which I shall speak to you is the problem of construction of Balanced Incomplete Block designs, which were first introduced in agricultural experimentation by Yates. Here we have to distribute v varieties in b blocks, each consisting of k different varieties, such that each variety occurs in r blocks, and each pair of varieties occur in λ blocks. Combinatorially it is equivalent to the following dinner party problem:—Persons v in number are to be invited to b dinners, k persons taking part in each dinner. It is required to arrange the invitations for dinner in such a manner that each person is invited r times, and two persons meet at a dinner just λ times. Clearly

$$bk = vr, \quad \lambda(v-1) = r(k-1)$$

The question is, given λ, b, k, v, r satisfying the above equations we have to construct the combinatorial solution. The combinatorial soution is not always possible when the five integers λ, b, k, v, r satisfying these conditions are given. Fisher has shown that $b \geqq v$, but even with this restriction there may not exist a solution, *e.g.*, it is known that the cases

$$v=36,\ b=42,\ r=7,\ k=6,\ \lambda=1\ ;$$

$$v=43,\ b=43,\ r=7,\ k=7,\ \lambda=1.$$

are impossible of solution.

Some interesting cases can be solved by using Finite Geometries. If $GF(p^n)$ is the Galois field with $s = p^n$ elements, then we can form the corresponding Euclidean and Projective Finite Geometries EG (m, s), PG (m, s) of m dimensions. If we take $m = 2$ the points as varieties and the lines as blocks we obtain the series of designs

$$v = s^2,\ b = s^2 + s,\ r = s + 1,\ k = s,\ \lambda = 1$$

$$v = s^2 + s + 1,\ b = s^2 + s + 1,\ r = s + 1,\ k = s + 1,\ \lambda = 1$$

These are the orthogonal series of Yates, and it is clear that the process by which one series is derived from the other, is equivalent to the adjunction of the line at infinity. Taking $m = 3$ the points as varieties and the planes as blocks we get the series

$$v = s^3,\ b = s^3 + s^2 + s,\ r = s^2 + s + 1,\ k = s^2,\ \lambda = s + 1$$

$$v = s^3 + s^2 + s + 1,\ b = s^3 + s^2 + s + 1,\ r = s^2 + s + 1,\ k = s^2 + s + 1,\ \lambda = s + 1$$

Only the case $s = 2$ is of practical interest, since otherwise the number of replications becomes too high. Similarly we get other designs by taking $m = 3$, the points as varieties and the lines as blocks. In particular when $s = 2$ we get the design $v = 15$, $b = 35$, $r = 7$, $k = 3$, $\lambda = 1$ from PG (3, 2).

Not all designs, however, can be obtained by geometrical methods. In fact, geometrical designs form a minority among the class of all designs. Hence, other methods are necessary. One such method is what I have called "the method of symmetrically repeated differences."

Consider a finite modul M (*i.e.*, a finite algebraical system where addition and subtraction is possible, the usual laws being obeyed) consisting of m elements, $x^{(0)}$, $x^{(1)}$ . . $x^{(n-1)}$ and to each element $x^{(i)}$ let there correspond m varieties $x_1^{(i)}, x_2^{(i)} \ldots\ldots x_m^{(i)}$ If $x_u^{(i)}$ and $x_v^{(j)}$ be two varieties occuring in the same block, we say that they give rise to a difference of the type $[u, v]$ and magnitude $x^{(k)} = x^{(i)} - x^{(j)}$. The difference is said to be pure or mixed according as $u = v$ or $u \neq v$. Since the varieties are all different the pure difference of magnitude 0 of any type $[u, u]$ is impossible. Thus there are $mn\ (mn - 1)$, different differences. Given t blocks $B_1, B_2, \ldots\ldots B_t$, each containing k different varieties, r varieties belonging to the same class ($kt = mr$). If among the $kt(k-1)$ differences which the blocks give rise to, each of the $mn\ (mn - 1)$ possible differences occurs exactly λ times, we say that the differences are symmetrically repeated. From each block B_i we can generate other blocks by adding a given element of M, to every variety in B keeping the lower suffix (denoting the class) constant. The complete set of nt blocks which can thus be generated by starting with the initial blocks $B_1, B_2, \ldots\ \ldots B_t$ is then a balanced incomplete block design. Consider for example the series of designs

$$v = 6t + 3,\ b = (3t + 1)\ (2t + 1),\ r = 3t + 1,\ k = 3,\ \lambda = 1.$$

For the elements of M, take integers reduced mod $(2t + 1)$, and to each element of M let there correspond three varieties. Then a set of initial blocks is provided by the block $[0_1, 0_2, 0_3]$ taken together with the $3t$ blocks

$$[\lambda_1,\ (2t + 1 - \lambda)_1,\ 0_2],\quad [\lambda_2,\ (2t + 1 - \lambda)_2,\ 0_3],\quad [\lambda_3,\ (2t + 1 - \lambda)_3,\ 0_1],$$

$$\lambda = 1,\ 2,\ \ldots\ldots\ t.$$

The construction of initial blocks satisfying the condition of symmetrically repeated differences, is much facilitated by using the properties of Galois fields. For example if $12t+1=p^n$ where p is a prime, and x is a primitive element of the Gaolis field $GF(p^n)$, then the set of t initial blocks given by

$$[0,\ x^{2i},\ x^{4t+2i},\ x^{8t+2i}]\quad i=0,\ 1,\ 2,\ \ldots\ t-1.$$

can be proved to satisfy the condition of symmetrically repeated differences. We thus get a general solution for the series of designs

$$v=12t+1,\ b=t(12t+1),\ r=4t,\ k=4,\ \lambda=1,$$

whenever $12t+1$ is the power of a prime.

The method of forming designs by the use of 'symmetrically repeated differences' as explained above can be extended by introducing a new variety ∞, which remains unchanged by adding any element of M, and introducing appropriate modifications in the conditions for a set of initial blocks. We then get solutions for many new general series. For example, let $4t+1=p^n$ where p is a prime, and x be a primitive element of GF (p^n) and k an integer such that

$$\frac{x^k+1}{x^k-1}=x^q,\quad \text{where } q \text{ is odd.}$$

Then the set of $3t+1$ blocks consisting of $(\infty,\ 0_1\ 0_2\ 0_3)$ and

$$(x_u^{2i},\ x_u^{2t+2i},\ x_v^{k+2i},\ x_v^{k+2t+2i}),\ i=0,\ 1,\ 2,\ \ldots\ t-1;\ (u,\ v)=(1,2),\ (2,3),\ (3,1)$$

generates a solution for the series

$$v=12t+4,\ b=(4t+1)(3t+1),\ r=4t+1,\ k=4,\ \lambda=1$$

For a symmetrical design *i.e.*, when $b=v$, $r=k$, it can be proved that of the k varieties occurring in a particular block just λ occur in each of the remaining blocks. Given a symmetrical design, if we cut out one block and all the varieties belonging to it, we get another design. For example by the method of symmetrically repeated differences we can get a general solution for the series

$$v=4t+3,\ b=4t+3,\ r=2t+1,\ k=2t+1,\ \lambda=t$$

and from this derive the solution for the auxiliary series,

$$v=2t+2,\ b=4t+2,\ r=2t+1,\ k=t+1,\ \lambda=t$$

A few words now about the problem of confounding in the general symmetrical factorial design. Let us consider a factorial design s^m involving m factors each at s levels, where we shall suppose that s is a prime or a power of a prime ($s=p^n$). Any treatment combination can then be represented by a symbol of the form

$$x_1,\ x_2,\ \ldots\ldots\ x_m$$

x_i denoting the level of the i[th] factor in the treatment combination. Since x_i can assume $s=p^n$ values, we can identify these values with the elements of a Galois field, so that every element of the field represents a level. Now $x_1,\ x_2,\ \ldots\ x_m$ can be regarded as the co-ordinates of a point in EG $(m,\ s)$. So that there is a $(1,\ 1)$ correspondence between the s^m points of this geometry and the s^m treatment combinations. By adding the flat at infinity $x=0$, we can immerse EG $(m,\ s)$ in PG $(m,\ s)$, the point $x_1,\ x_2\ \ldots\ x_m$ of EG $(m,\ s)$ being now identified with the finite point $(1,\ x_1,\ x_2,\ \ldots\ x_m)$ of PG $(m,\ s)$.

Let O be the point $(1, 0, 0, \dots 0)$ of PG (m, s) and X_i the point for which $x_0 = 0$, $x_1 = 0, \dots, x_{i-1} = 0$, $x_i = 1$, $x_{i+1} = 0, \dots, x_m = 0$. Then the lines $OX_1, OX_2 \dots OX_m$ play the same part as the axes of reference in ordinary geometry, O being the origin. The points $X_i (i = 1, 2 \dots m)$ are, of course, at infinity. Then the simplex $X_1, X_2, \dots X_m$ may be termed the fundamental simplex. The points X_i $(i = 1, 2, \dots m)$ may be called its vertices or zero cells; the lines $X_i X_j$ $(i, j = 1, 2, \dots m;\ i \neq j)$ its edges or 1-cells; the triangles $X_i X_j X_k$ $(i, j, k = 1, 2, \dots m;\ i \neq j \neq k)$ its 2-cells, and, in general, the $(k-1)$ dimensional partial simplexes formed from any k of the m points $X_1, X_2, \dots X_m$ may be called its $(k-1)$-cells.

Through any $(m-2)$-flat at infinity there will pass a pencil of s parallel finite $(m-1)$-flats, each containing s^{m-1} finite points. These will divide the s^m treatments into s sets of s^{m-1} treatments each. If the treatments corresponding to the s^{m-1} points in any one of these $(m-1)$-flats are considered as belonging to the same set, the contrast between these sets represents $s-1$ degrees of freedom. We may speak of these degrees of freedom as belonging to the pencil of $(m-1)$-flats considered, this pencil being determined by the $(m-2)$-flat at infinity which may be called the vertex of this pencil.

Now the number of $(m-2)$-flats in the $(m-1)$-flat at infinity is $s^{m-1} + s^{m-2} + \dots + s^2 + s + 1$. To each of these corresponds a pencil of s finite $(m-1)$-flats with $s-1$ degrees of freedom. Thus the total number of degrees of freedom carried by these pencils is $s^m - 1$ which we know to be the total number of degrees of freedom for all treatment comparisons.

It can now be shown that the $s-1$ degrees of freedom given by the contrasts between the s sets of treatment combinations into which the totality of s^m treatments are divided by the pencil of $(m-1)$-flats represented by the equation

$$x_{i_0} + a_{u_1} x_{i_1} + a_{u_2} x_{i_2} + \dots\dots + a_{u_k} x_{i_k} = a_r \quad \left[\begin{array}{l} i_0, i_1, \dots\dots i_k, u_1, u_2 \dots\dots u_k \\ \text{fixed};\ r = 0, 1, 2 \dots s-1 \end{array} \right]$$

belong to the k^{th} order interaction. It is also readily shown that the two sets of $s-1$ degrees of freedom corresponding to any two pencils of this type are mutually orthogonal. Every $(m-1)$-flat of this pencil passes through the $(m-k-2)$-cell of the fundamental simplex, obtained by excluding the $k+1$ points $X_{i_0}, X_{i_1} \dots X_{i_k}$ from among the vertices of the simplex. Conversely every pencil whose $(m-1)$-flats pass through the same $(m-k-2)$-cell of the fundamental simplex, has an equation of the above form, and the degrees of freedom corresponding to it therefore belong to a k^{th} order interaction. As each of $u_1, u_2 \dots u_k$ can assume $s-1$ different values, the total number of pencils of the above type is $(s-1)^k$ which give $(s-1)^{k+1}$ degrees of freedom corresponding to the k^{th} order interaction between the $i_0^{th}, i_1^{th}, \dots i_k^{th}$ factors.

The principle of generalised interaction first enunciated by Barnard for the 2^m symmetrical factorial design can now be readily extended. The change of main effects is simply equivalent to a change of the fundamental simplex. The equations of transformation being once written down we can readily calculate the new designation of any interaction.

The extended principle of generalized interaction may now be utilized in enumerating the various possible types of confounding by studying the relation in which 0-flats, 1-flats, 2-flats, 3-flats, ... $(m-2)$-flats in the $(m-1)$-flat at infinity stand to the fundamental simplex.

Consider a $(m-k-1)$-flat at infinity and the degrees of freedom associated with it. It is fixed as the common $(m-k-1)$-flat of intersection of k independent $(m-2)$-flats at infinity. The pencils of $(m-1)$-flats corresponding to these, intersect in the finite portion in s^k $(m-k)$-flats which have the given flat at infinity for vertex. As these constitute the totality of $(m-k)$-flats with the given $(m-k-1)$-flat at infinity as vertex, to each of these $(m-k-1)$-flats at infinity are associated s^k-1 degrees of freedom given by the contrast between the s^k sets of s^{m-k} treatment combinations into which the s^m treatments are split up. Of these, $k(s-1)$ degrees of freedom belong to the main effects or interactions corresponding to the initial k $(m-2)$-flats at infinity, and the remaining $s^k-1-k(s-1)$ degrees of freedom to the main effects or interactions determined by the generalized interactions in their entirety of the k initial main effects or interactions. It appears, therefore, that for the formation of confounded arrangements in the case of a s^m design in s^k sub-blocks, we have to look for a particular $(m-k-1)$-flat at infinity and set down the s^{m-k} treatments occurring in each of the s^k finite $(m-k)$-flats having the given $(m-k-1)$-flat at infinity as vertex. The nature of confounding thus effected would, as above, be deducible from considering the relation in which the totality of the $(m-2)$-flats at infinity $\frac{s^k-1}{s-1}$ in number, passing through the given $(m-k-1)$-flat at infinity stand in relation to the fundamental simplex.

The totality of the number of ways of getting a s^m design arranged in s^{m-k} plot blocks may be divided up into a number of classes in accordance with the types of the $(m-k-1)$-flats at infinity in relation to the fundamental simplex, each of these different types leading to one particular type of confounding. Among these, the best sets of treatment comparisons which may profitably be confounded are those in which the main effects and first order interactions are affected as little as possible, and will correspond to the $(m-k-1)$-cells, if any, which pass clear of the fundamental simplex.