Commentary

R A Fisher, design theory, and the Indian connection

Design Theory, a branch of mathematics, was born out of the experimental statistics research of the population geneticist R A Fisher and of Indian mathematical statisticians in the 1930s. The field combines elements of combinatorics, finite projective geometries, Latin squares, and a variety of further mathematical structures, brought together in surprising ways. This essay will present these structures and ideas as well as how the field came together, in itself an interesting story.

1. Introduction

What do the following have in common:

• Kirkman’s School Girl Problem: “15 young ladies in a school walk out three abreast for 7 days in succession; it is required to arrange them daily, so that not two will walk twice abreast” (Beth et al. 1985).

• The puzzle-game SuDoKu, one of the greatest mathematicians, and judging effectiveness of fertilizers on potato varieties.

• R A Fisher, statistician and population geneticist, a key figure in the synthesis between Darwin and Mendel.

• Branches of mathematics called design theory and coding theory.

• Projective geometry, a subject in which unlike in Euclidean geometry, there is a duality between points and lines such that interchanging them in any theorem does not affect its validity.

• India’s pioneering statistician and early associates in the school he founded.

This essay will present the interesting mathematical structures and ideas in the above items and the human interest thread that weaves through them. Whether arranging numbers from 1 to 9 in a 9×9 array so that each numeral occurs once and only once in each row and column, arranging schoolgirls in 5×3 blocks so that no pair is repeated, or arranging plots of potato varieties and the laying of different fertilizers on them so that each variety is subjected to each type of fertilizer to gauge effectiveness, these are all problems of ‘experimental design’ and now a branch of mathematics called ‘design theory’ (Lenz 1991; Beth et al. 1985), related also to coding theory (Jungnickel 1990). These are parts of the wider fields of combinatorics as well as finite projective geometries (Hall 1967; Rao Raghava 1971; Bose and Manvel 1984). While some of the basics go back to the great mathematician Euler, it is the work of Fisher and of a school of Indian mathematical statisticians that gave birth to Design Theory (Gropp 1992). In statistics, this is also referred to as ‘Design of Experiments’ or ‘Experimental Designs’.

2. Design theory

Kirkman’s School Girl Problem, originally posed by W S B Woolhouse in 1844 (Woolhouse 1844) and solved by the Rev. Thomas Kirkman, a Lancashire clergyman and amateur mathematician (Biggs 1981), in 1847 in a charmingly named journal (Kirkman 1850), is a precursor of what have come to be known as ‘designs’ and more specifically, ‘balanced incomplete block (BIB) designs’ (Yates 1936; Bose 1939; Rao Raghava 1971) or ‘Steiner triple systems’ (Rao Raghava 1971; Witt 1938). There was also early work by

Keywords. Bose R C; design theory; finite projective geometries; Fisher R A; Latin squares; mathematical statistics; octonions and higher arithmetics
the great mathematician Euler and today, all of this is part of a branch of mathematics called design theory (Lenz 1991).

The idea is to consider two sets, members of one to be allotted to those of the other with certain specified conditions. The first set of \( v \) objects or symbols (may be anything: numbers, potatoes, ...) as with \( v = 15 \) ladies, is to be put into \( b \) blocks. Each block contains exactly \( k \) distinct symbols, as in \( k = 3 \) ladies abreast, each symbol to occur in exactly \( r \) different blocks and every pair of distinct symbols to occur together in exactly \( \lambda \) blocks. In the case of the school girls, \( r = 7 \), the number of days, and \( \lambda = 1 \) because no two should recur from one day to the next. Kirkman constructed the solution with \( b = 35 \), these being the number of rows of three, 5 for each of the 7 days.

A \( (v,b,r,k,\lambda) \) design or BIB is thus one of \( v \) objects in \( b \) blocks with each block containing exactly \( k \) distinct objects, each object occurring in exactly \( r \) different blocks and every pair (or more general \( t \)) of distinct objects occurring together in exactly \( \lambda \) blocks. Block designs with \( k = 3 \) are called triple systems. Those with \( \lambda = 1 \) are called Steiner systems \( S(t,k,\nu) \) and, if \( k = 3 \) as well, Kirkman or Steiner triple systems \( S(2,3,\nu) \) because the Berlin mathematican Jakob Steiner, proposed their existence in 1853, conjecturing that the number \( \nu \) had to be such that it would leave a remainder of 1 or 3 upon dividing by 6 (Steiner 1853). This was proved by Reiss (Reiss 1859) six years later but they were unaware of Kirkman’s work (Kirkman 1847).

The following relationships define a BIB: \( vr = bk, \lambda(v-1) = r(k-1) \). For triple systems with \( k = 3 \), these reduce to \( r = \lambda \) \((v-1)/2 \), \( b = \lambda v(v-1)/6 \). Another notation used for BIBs is \( t-(v,k,\lambda) \) so that a Steiner triple system is \( 2-(v,3,1) \), the Kirkman problem being a \( 2-(15, 3, 1) \) design. An even smaller one is \( 2-(7, 3, 1) \) or \( S(2,3,7) \) which we will encounter in Section 4 in a geometrical context of placing 7 points on 7 lines such that each line has three points on it and each point lies on three lines with no pair of points on more than one line. The terminology of symbols and blocks is replaced by the geometrical ones of points and lines, respectively. With \( (v=b, r=k) \), such a BIB is said to be symmetrical. The result of Steiner and Reiss allows a parametrization of Steiner triple systems in terms of a single integer \( n \). One family has \( (v=6n+3, b=(3n+1)(2n+1), r=3n+1) \) and a second \( (v=6n+1, b=n(6n+1), r=3n) \). With increasing \( v \), establishing the exact number, 80 for \( v=15 \) and over two million for \( v=19 \), and classifying Steiner triple systems becomes complicated. For these and the long history of establishing the result of two non-equivalent designs for \( v=13 \), (see Gropp 1991).

Much development of the subject comes from the work of R A Fisher who formulated the principles of statistical designs in 1925 in the context of agricultural research/statistics, and from Yates who introduced the use of BIB designs in 1936 (Yates 1936). In studying the effects of various fertilizers and soils on growing potatoes and barley, Fisher was conducting field studies which led to the design of statistical experiments. A complete experiment on the effectiveness of \( v \) different fertilizers on \( b \) types of plants would require \( b \) plots, each subdivided into \( v \) areas. This could be prohibitively expensive. An ‘incomplete’ one would test every type of plant with \( k < v \) different fertilizers such that any two fertilizers would be tested on \( \lambda \) different types of plants. ‘Balancing’ occurrence of pairs of treatments on exactly \( \lambda \) of \( b \) blocks of size \( k \) means the regular appearance of pairs of fertilizers on the same plant, allowing a complete covariance analysis of the results. This was Fisher’s great insight along with his focus not on one character at a time but a multivariate analysis. He introduced the idea of variance and maximum likelihood, established inequalities named for him (that a proper BIB requires \( b \geq v, r \geq k \)), and rapidly in the 1920s and 1930s established the field with mathematical rigour, writing his 1935 book, “The Design of Experiments” (Fisher 1935). The terminology introduced by Yates (Yates 1936) of \( v \)(arieties), \( t \)(reatments) and \( r \)(eplications) provides the symbols still in use today.

The next step in the development of Design Theory as the full-fledged branch of mathematics that it is today can be traced to Fisher’s trip to India in 1938 when he visited his friend P C Mahalanobis who had similarly pioneered the use of agricultural statistics in India, establishing a journal, Sankhya, and the Indian Statistical Institute in December 1931. A couple of young assistants in that group, most notably R C Bose, with physics and mathematics background, had been following Fisher’s idea of representing an \( n \)-sample by a point in \( n \)-dimensional Euclidean space, and were solving many design problems and constructing BIBs. They took up questions Fisher posed on statistical designs for controlled experiments, using their expertise in finite geometries, leading to the study based on Galois fields that forms the modern basis of the subject. We will return to this in Section 5.