Contents

Preface

Incomplete-block designs for experiments were first developed by Yates at Rothamsted Experimental Station. He produced a remarkable collection of designs for individual experiments. Two of them are shown, with the data from the experiment, in Example 4.3 on page 91 and Exercise 5.10 on page 110. This type of design poses two questions for statisticians: (i) what is the best way of choosing subsets of the treatments to allocate to the blocks, given the resource constraints? (ii) how should the data from the experiment be analysed?

Designs with partial balance help statisticians to answer both of these questions. The designs were formally introduced by Bose and Nair in 1939. The fundamental underlying concept is the association scheme, which was defined in its own right by Bose and Shimamoto in 1952.

Many experiments have more than one system of blocks. These can have complicated inter-relationships, like the examples in Section 7.1, which are all taken from real experiments. The general structure is called an orthogonal block structure. Although these were introduced quite independently of partially balanced incomplete-block designs, they too are association schemes. Thus association schemes play an important role in the design of experiments.

Association schemes also arise naturally in the theory of permutation groups, quite independently of any statistical applications. Much modern literature on association schemes is couched in the language of abstract algebra, with the unfortunate effect that it is virtually inaccessible to statisticians. The result is that many practising statisticians do not know the basic theory of the Bose–Mesner algebra for association schemes, even though it is this algebra that makes everything work. On the other hand, many pure mathematicians working in the subject have no knowledge of the subject's origin in, and continued utility in, designed experiments.

This book is an attempt to bridge the gap, and is intended to be accessible to both audiences. The first half is at a level suitable for final-year undergraduates (on the four-year MSci programme) or MSc students in Mathematics or Statistics. It assumes some linear algebra, modular arithmetic and probability, but no statistics and very little abstract algebra. The linear algebra is revised where it is needed (Sections 1.3, 2.1, 3.1), partly to establish the notation used in this book. The same is done for random variables in Section 4.2. Techniques which use finite fields are deliberately avoided, although the reader with some knowledge of them may be able to see where examples can be generalized.

After the basic theory in the first two chapters, the book has three main strands. The first one is the use of association schemes in designed experiments, which is developed in Chapters 4, 5, 7 and parts of 8 and 11. The second is the fruitful interplay between association schemes and partitions: see Chapters 6, 7, 9, 10 and 11. The third gives methods of creating new association schemes from old ones. This starts in Chapter 3 to give us an easy mechanism for developing examples, and continues in Chapters 9 and 10.

Chapters 1–6 form the heart of the book. Chapter 1 introduces association schemes and gives the three different ways of thinking about them: as partitions, as adjacency matrices, as coloured graphs. It also gives many families of association schemes. Chapter 2 moves straight to the Bose-Mesner algebra spanned by the adjacency matrices. The fact that this algebra is commutative implies thay these matrices have common eigenspaces, called strata. The relationship between the adjacency matrices and the stratum projectors is called the character table: this is the clue to all calculations in the association scheme. This chapter includes a section on techniques for actually calculating character tables: this should be useful for anyone who needs to make calculations in specific association schemes, for example to calculate the efficiency factors of a block design.

Chapter 3 introduces crossing and nesting, two methods of combining two association schemes to make a new one. Although these are not strictly necessary for an understanding of Chapters 4–6, they provide a wealth of new examples, and a glimpse, in Section 3.5, of the complicated structures that can occur in real experiments.

Chapters 4 and 5 cover incomplete-block designs. Chapter 4 gives the general theory, including enough about data analysis to show what makes a design good. In Chapter 5 this is specialized to partially balanced incomplete-block designs, where the Bose-Mesner algebra gives very pleasing results. Many of Yates's designs are re-examined in the light of the general results.

Chapter 6 introduces the machinery for calculating with partitions of a set. This leads immediately to the definition of orthogonal block structures and derivation of their properties. They yield association schemes which have explicit formulas for the character tables.

The next three chapters build on this core but are almost independent of each other. Chapter 7 covers designs where the experimental units have an orthogonal block structure more complicated than just one system of blocks. There are designs for row-column arrays, for nested blocks, for row-column arrays with split plots, and so on. The idea of partial balance from Chapter 5 is extended to these more complicated designs. Topics covered include efficiency of estimation, combining information from different strata, randomization. All of this needs the Bose–Mesner algebra.

Group theory is deliberately avoided in Chapters 1–7, because many statisticians do not know any. Cyclic designs are dealt with by using modular arithmetic without any appeal to more general group theory. However, the reader who is familiar with group theory will realise that all such results can be generalized, sometimes to all finite groups, sometimes just to Abelian ones. Chapter 8 revisits Chapters 1–7 and makes this generalization where it can. Later chapters have short sections referring back to this chapter, but the reader without group theory can omit them without losing the main story.

Chapter 9 is devoted to poset block structures, a special class of orthogonal block structures that is very familiar to statisticians. They were developed and understood long before the more general orthogonal block structures but, ironically, have a more complicated theory, which is why they are deferred to this part of the book. Part of the difficulty is that there are two partial orders, which can easily be confused with each other. The idea of poset block structures also gives a new method of combining several association schemes, which generalizes both crossing and nesting.

Chapters 10 and 11 are the most abstract, drawing on all the previous material, and heavily using the methods of calculating with partitions that were devoloped in Chapter 6. Chapter 10 looks at association schemes with an algebraist's eye, giving some relations between association schemes on different sets. It gives further constructions of new from old, and shows the important role of orthogonal block structures. Some of these results feed into Chapter 11, which looks at association schemes on the same set, trying to answer two statistically motivated questions from earlier chapters. Is there a 'simplest' association scheme with respect to which a given incomplete-block design is partially balanced? This has a positive answer. If there is more than one system of blocks, does it matter if they give partial balance for different association schemes? The answer is messy in general but detailed answers can be given for the three main classes of association schemes discussed in the book: orthogonal block structures, poset block structures, and Abelian group schemes.

The book concludes with two short chapters, looking to the future and summarizing the history.

The reader who simply wants an introduction to association schemes as they occur in designed experiments should read Chapters 1–6. For more emphasis on pure combinatorics, replace parts of Chapters 4 and 5 with parts of Chapters 8 and 10. The reader who designs experiments should also read Chapter 7 and one or both of Chapters 8 and 9. The reader who is more interested in the interplay between partitions and association schemes could read Chapters 1–3, 6 and 8–11.

This book differs from statistics books on incomplete-block designs, such as [115, 118, 121], in that it uses the Bose–Mesner algebra explicitly and centrally. Even books like [76, 178], which cover many partially balanced designs, do not make full use of the Bose–Mesner algebra. Furthermore, orthogonal block structures are developed here in their own right, with a coherent theory, and Chapter 7 gives a more detailed coverage of designs for orthogonal block structures than can be found in any other book on designed experiments.

On the other hand, this book differs from pure mathematical books on association schemes not just because of its attention to statistical matters. The inclusion of explicit sections on calculation (Sections 2.4, 5.2 and 6.4) is unusual. So is the emphasis on imprimitivite association schemes, which are others? deliberately downplayed in books such as [33]. On the other hand, I make no attempt to cover such topics as Krein parameters, polynomial schemes or the links with coding theory.

Most of the material in Chapters 6–7 and 9–11 has not previously appeared in book form, as far as I know.

Some topics in this book already have an established literature, with standard notation, although that may differ between research communities, for example t or v for the number of treatments. I have decideded to aim for consistency within the book, as far as is practicable, rather than consistency with the established literatures. Thus I am thinking more of the reader who is new to the subject and who starts at Chapter 1 than I am of the specialist who dips in. If your favourite notation includes an overworked symbol such as n, k, λ or P, be prepared to find something else. I hope that the glossary of notation on page 345 will be helpful.

Acknowledgements

My interest in this subject was lit by two courses I attended as a DPhil student. G. Higman lectured on graphs, groups and geometries, which included strongly regular graphs and generalized polygons; D. G. Higman gave a course on coherent configurations which became [100] . When I turned to statistics I studied simple orthogonal block structures, and I am grateful to T. P. Speed for pointing out that these could be viewed as association schemes. N. L. Johnson and S. Kotz invited me to write an article on partially balanced designs for the *Encyclopedia of Statistical Science*: I am very glad that they refused to accept my plea of ignorance so that I was forced to see how these designs fitted into the story. Fortunately, I was visiting the University of North Carolina at Chapel Hill at the time, and was able to spend some happy hours in the library reading work by R. C. Bose and his students, one of whom, I. M. Chakravarti, introduced me to Delsarte's thesis.

More recently, I am grateful to my colleagues P. J. Cameron and B. A. F. Wehrfritz for encouraging me to develop the material as a lecture course; to those colleagues and students in the Queen Mary Combinatorics Study Group who have commented on draft presentations of the material in Chapters 8–11; and to all those students who have directly or indirectly helped me to develop examples

and better ways of explaining things. P. J. Cameron read early drafts of the first six chapters and helped me with the second half of Chapter 13. Of course, the responsibilty for the choice of contents, and for the choice of names and notation where there are conflicts between different traditions, is mine, as are all mistakes in the text.

Web page

Updated information about this book, such as known errors, is available on the following web page.
`http://www.maths.qmw.ac.uk/~rab/Asbook`

R. A. Bailey
School of Mathematical Sciences
Queen Mary, University of London
Mile End Road
London E1 4NS
r.a.bailey@qmul.ac.uk

October 3, 2001