Simplifying Bayesian experimental design for multivariate partially exchangeable systems

Shaw, Simon Charles

Simplifying Bayesian Experimental Design for Multivariate Partially Exchangeable Systems

A thesis presented for the degree

of Doctor of Philosophy at the

University of Durham

Simon Charles Shaw

Department of Mathematical Sciences,

University of Durham,

Durham, DH1 3LE.

June 2000

Abstract

We adopt a Bayes linear approach to tackle design problems with many variables cross-classified in many ways. We investigate designs where we wish to sample individuals belonging to different groups, exploiting the powerful properties of the adjustment of infinitely second-order exchangeable vectors. The types of information we gain by sampling are identified with the orthogonal canonical directions. We show how we may express these directions in terms of the different factors of the model. This allows us to solve a series of lower dimensional problems, through which we may identify the different aspects of our adjusted beliefs with the different aspects of the choice of design, leading both to qualitative insights and quantitative guidance for the optimal choice of design. These subproblems have an interpretable form in terms of adjustment upon subspaces of the full problem and remain valid when we consider adjusting the underlying population structure and also for predicting future observables from past observation. We then examine the adjustment of finitely second-order exchangeable vectors, and show that the adjustment shares the same powerful properties as the adjustment in the infinite case. We show how if the finite sequence of vectors is extendible, then the differences in the adjustment of the sequence is quantitatively the same for all sequence lengths and it is easy to compare the qualitative differences. Extending to an infinite sequence allows us to draw comparisons between the finite and infinite modelling. Such comparisons may also be made when we consider sampling individuals belonging to different groups, where each group contains only a finite number of individuals.

Contents

1 Introduction: Bayes linear analysis for infinite second-order exchangeable sequences **10**

1.1 Bayesian Experimental Design 11

1.2 The problem 15

1.3 The quantification of uncertainty 15

1.3.1 The subjectivist perspective 15

1.3.2 A Bayesian approach 17

1.3.3 The Bayes linear approach 18

1.3.4 Prevision 18

1.3.5 The geometric interpretation of prevision 21

1.4 Formalising the problem 22

1.4.1 The examiner's problem 23

1.5 Second-order exchangeability 23

1.5.1 Full exchangeability 24

1.5.2 Problems with full exchangeability 26

1.5.3 Second-order exchangeability 27

1.6 The representation theorem for infinite second-order exchangeable beliefs 28

1.6.1 Applying the representation theorem for the example 29

1.7 The adjustment of beliefs 29

1.7.1 Revising previsions 30

1.7.2 Revising belief structures 33

1.7.3 Interpreting the adjustment 34

1.8 Resolution transforms 37

1.8.1 Bayes linear sufficiency 38

1.8.2 Resolution transforms for infinite exchangeable collections . . 39

1.9 Predictive adjustment 40

1.10 Implementations of the theory 41

1.11 Progressions 42

2 Simplifying Complex Designs: Bayes Linear Experimental Design for Grouped Multivariate Exchangeable Systems 44

2.1 Widening the scheme to include co-exchangeability 45

2.1.1 Partial Exchangeability 46

2.1.2 Co-exchangeability 48

2.1.3 Adjusting the mean components 49

2.1.4 Predicting future individuals 52

2.2 The examiner revisited 54

2.2.1 The examiner's considerations 54

2.2.2 Learning about the mean components 56

2.3 Grouped multivariate exchangeable systems 60

2.3.1 Specification of the system 60

2.3.2 Scaled exchangeability - a potential model approach to the specifications 64

2.4 Adjusting the mean components 67

2.4.1 Underlying canonical variable problem 68

2.4.2 The underlying canonical variable problem for the examiner . 70

2.4.3 Underlying canonical group problem 71

2.4.4 The underlying canonical group problem for the examiner . . 75

2.4.5 The adjustment of the mean components by the observed sample . . . 76

2.4.6 The adjustment of $[\mathcal{M}(\mathcal{C})]$ for the examiner example . . . 81

2.5 A link with analysis of variance models . . . 81

2.6 Predictive adjustment: predicting in already observed groups . . . 83

2.6.1 Analysis of groups and variables . . . 85

2.6.2 Adjustment of the full collection . . . 90

2.6.3 Example: the examiner uses prediction . . . 95

2.7 Predictive adjustment: predicting in unobserved groups . . . 96

2.7.1 Analysis of groups and variables . . . 97

2.7.2 Adjustment of the full collection . . . 100

2.7.3 Unobserved prediction for the examiner . . . 101

3 Multifactor multivariate exchangeable systems 102

3.1 Specifications of the model . . . 103

3.1.1 The examiner revisited . . . 106

3.1.2 A possible situation where our beliefs may be held . . . 107

3.2 Adjusting the mean components . . . 112

3.3 Underlying canonical variable and group problems . . . 113

3.3.1 The underlying canonical variable and group problems for the examiner . . . 119

3.4 The adjustment of the mean components by the observed sample . . . 120

3.5 The full solution to the examiner's problem . . . 123

3.6 Marginal tables . . . 125

3.6.1 Marginalising over any $k - r$ factors . . . 130

3.6.2 The examiner considers marginalisation . . . 132

3.7 Sliced tables . . . 133

3.7.1 Taking a general $f_{[\bar{\Delta}]}$-slice . . . 138

3.7.2 The examiner considers slicing 141

4 Modelling using finite second-order exchangeability 142

4.1 Introduction 143

4.2 Finite second-order exchangeability 145

4.2.1 The examiner's considerations 147

4.3 The representation theorem for finite second-order exchangeable beliefs 148

4.3.1 The examiner applies the representation theorem 149

4.4 The examiner's learning 149

4.5 Learning about the mean components 153

4.5.1 Resolution transforms for (finitely) exchangeable systems . . 154

4.6 Extendible and non-extendible exchangeable sequences 157

4.7 Linking the adjustments of finite and infinite second-order exchangeable sequences 161

4.8 Prediction of future individuals 164

4.9 Implementing the theory; linking with the underlying canonical variable problem 164

4.10 Co-exchangeable finitely exchangeable systems 165

4.10.1 Adjusting the mean components 167

4.11 Grouped multivariate exchangeable systems: the finite analogue . . . 169

4.12 Underlying canonical variable problem 170

4.12.1 The examiner problem in this finite setting 171

4.13 Underlying canonical finite group problems 172

4.13.1 The underlying canonical finite group problem for the examiner 176

4.14 The adjustment of the mean components by the observed sample . . 178

4.14.1 The full adjustment for the examiner 181

5 Extensions: one-level hierarchical co-exchangeability 182

5.1 The one-level hierarchical co-exchangeable model . . . 183

5.2 Sampling and learning about the components of the model . . . 184

5.3 Learning about the overall population collection . . . 186

5.4 Learning about the gth group residual . . . 187

5.5 Learning about the group means . . . 191

5.6 Comments and conclusions . . . 193

6 Conclusions, suggestions for further work 194

A Some useful notation 199

A.1 Matrices and sets . . . 199

A.2 Collections of random quantities . . . 199

A.3 Resolution transforms and canonical structures . . . 202

List of Tables

1.1 Resolutions for quantities in $[\mathcal{M}(\mathcal{C})]$, having seen the scores on a single exam paper . . . 35

1.2 Resolutions for the collection $\mathcal{M}(\mathcal{C})$, adjusting by $\mathcal{C}(n)$ . . . 35

1.3 Sample sizes required to achieve the given variance reductions for $\mathcal{M}(X_+)$ and $\mathcal{M}(X_-)$ . . . 36

2.1 Resolutions for the collection $\mathcal{M}(\mathcal{C}_+)$, adjusting by $\mathcal{C}(nI_2)$ . . . 57

2.2 Resolutions for the collection $\mathcal{M}(\mathcal{C}_-)$, adjusting by $\mathcal{C}(nI_2)$ . . . 58

3.1 The canonical directions for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_8)$. The multipliers are $\tau_+ = (1+\gamma)^{-\frac{1}{2}}$ and $\tau_- = (1-\gamma)^{-\frac{1}{2}}$. . . . 124

3.2 The corresponding canonical resolutions to the canonical directions of Table 3.1 for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_8)$. . . . 125

4.1 Resolutions for the quantities in $[\tilde{\mathcal{M}}(\mathcal{C})]$, having seen the scores on a single exam paper . . . 150

4.2 Resolutions for the quantities in $[\tilde{\mathcal{M}}(\mathcal{C})]$ having seen the scores on the first n exam papers . . . 151

4.3 Table showing the resolutions required to achieve the specified variance reductions in the given quantities . . . 152

Declaration

I declare that the material presented in this thesis has not been submitted previously for any degree in either this or any other university. It is based on research carried out between October 1996 and June 2000. Chapter 1 is an introduction, and as such no claim for originality is made. Some of the ideas behind Chapter 2 were the result of joint work with Michael Goldstein and the work on the adjustment of the mean components by the observed sample, as summarised by Theorem 13, has been published as Shaw & Goldstein (1999). The rest of the work for this thesis is entirely my own; the calculations in Chapter 3 were carried out using the Bayes linear compting package $[B/D]$ (see Wooff (1992)), which is available at the internet site *http://maths.dur.ac.uk/stats/bd/home.html*

Acknowledgements

I'd like to thank my parents for their unfailing support and faith in me, alongside my sister, Sally, and her partner, Stuart. I'd also like to acknowledge the guidance and confidence I've gained from everyone who ever had a cup of tea at Lyndhurst. Kieron should be thanked for sharing the rules of chess with me twenty years ago, and also infusing others with his passion for jazz.

I owe an immeasurable debt to my supervisor, Michael Goldstein, for his patient help and wonderful insights over the course of my PhD and also for his understanding, both on a personal and academic level. I'd like to thank the whole Maths department, and the Statistics department in particular, for its' friendliness and inspiration, during both my degree and PhD. Indeed, I'd like to express gratitude to anyone who has ever taught me. At William Farr, I have fond memories of the late Terence Leach for his depth of knowledge and enthusiasm for matters Lincolnian. The Maths department at De Aston, in particular my further Maths teachers Mr Neal and Mr Higgins, deserve great credit for the solid base they gave me and for first steering me down the road I now travel. The financial support provided by the UK's EPSRC was greatly appreciated.

I would like to dedicate this thesis to the memory of Dyfrig Williams. I will never forget the help and guidance he gave me.

Chapter 1

Introduction: Bayes linear analysis for infinite second-order exchangeable sequences

SUMMARY

This chapter is an introduction. In Section 1.1, we consider Bayesian experimental design and highlight a number of problems that we would like to overcome. In Subsection 1.3.1, we argue for the adoption of a subjectivist approach to the revision of belief and in Subsection 1.3.2 consider the Bayesian approach and problems caused by our abilities to only make limited belief specifications. As such, we advocate a system of partial belief adjustment; this system is commonly known as the Bayes linear approach and is motivated in Subsection 1.3.3. One of the lynchpins of statistical modelling is the use of exchangeability. In Section 1.5, we highlight full exchangeability and point out drawbacks to its adoption. Instead, we use second-order exchangeability, and this is reviewed in Subsection 1.5.3 and the representation theorem for second-order exchangeable sequences discussed in Section 1.6. The adjustment of beliefs in the Bayes linear framework is examined in Section 1.7, in particular highlighting the use of the resolution as a means of assessing the impact of an adjustment upon a quantity. Section 1.8 summarises the resolution transform and the related canonical resolutions and directions. These are tools that we shall find of great use in this thesis. In Subsection

1.8.1 we review the concept of Bayes linear sufficiency and illustrate this with the Bayes linear sufficiency of the sample means for a sample of individuals for adjusting exchangeable collections. In Subsection 1.8.2 we review the adjustment of second-order exchangeable sequences and the strong coherency conditions between adjustments with different sample sizes. These features are also exhibited in predictive adjustment which is examined in Section 1.9. In Section 1.10, we look at matrix implementations of the theory; within this thesis, we shall make heavy use of such implementations. Throughout this chapter, we use a simple motivating example concerning an examiner assessing marks on an exam paper as a means to illustrate the methodology.

1.1 Bayesian Experimental Design

The early work on the study of experimental design is typically assigned to the work of Fisher (1925, 1935) after he was set to work, at Rothamsted in 1919, to see what could be statistically gleaned from the years of records of experimental and observational data that had been collected there. Chapter 6 of Box (1978) provides an insight into Fisher's work on, and motivation for, the design of experiments.

To give a precise definition to what we mean by experimental design is hard, for as Deely (1992; p475) writes

> The expression "Experimental Design" has come to mean many things to many people ... In its simplest form I think of experimental design along the following lines. There is a core of experimental units about which we want to know something because we have some purpose in mind. There is some information available about these experimental units and generally it has some influence upon what we want to know. In this context then the design of experiments can be stated as simply determining which experimental units to use and how many.

This is a definition we find appealing and is the one we wish to bear in mind. Determining "which" and "how many" however is not simple. To borrow the title of Lad & Deely (1994) , we view "experimental design from a subjective utilitarian viewpoint". Personally speaking, the very nature of experimental design advocates a Bayesian approach. We shall consider the subjectivist perspective in Subsection 1.3.1 and the Bayesian standpoint in Subsection 1.3.2. Consider, for example, the opening lines of Cochran & Cox (1957; p1) when they write

> It is true that on many important aspects of experimentation the statistician has no expert knowledge. Nevertheless, in recent years, research workers have turned increasingly to statisticians for help both in planning their experiments and in drawing conclusions from the results. That this has happened is convincing evidence that statistics has something to contribute.

The statistician himself may have no expert knowledge on the nuances of the experiment and what is envisaged as potential outcomes. This does not mean however, that such information is not available, for the research worker/experimenter has this information. Indeed, it is his beliefs about the experiment that we, as the statistician, are attempting to model. Thus, we envisage a synergy between experts in different fields: the research worker with his beliefs and knowledge about the problem he wishes to investigate and the statistician who provides the tools to enable this analysis of beliefs.

Early work on Bayesian experimental design may be found in the work of Raiffa & Schlaifer (1961) and Lindley (1972). These provide a decision-theoretic approach to experimental design. Lindley (1972; p20) considers an experiment as a triplet $e = (X, \Theta, p(x|\theta))$, where X is a sample space of elements x; Θ is a parameter space of elements θ, and $p(x|\theta)$ a density function. We have a collection, E, of experiments e having a common Θ, together with a decision space D. The decision process is two part. Prior to choosing e and observing x, we determine which is the best e to choose from E. Based on our observation x, we choose a decision $d \in D$. Selecting e, yielding the data x, choosing d and yielding θ produces the utility $U(d, \theta, e, x)$. As Chaloner & Verdinelli (1995; p275) write

> Lindley's argument suggests that a good way to design experiments is to specify a utility function reflecting the purpose of the experiment, to regard the design choice as a decision problem and to select a design that maximises the expected utility.

Thus, the Bayesian solution to the experimental design problem is provided by equation (4.7) of Lindley (1972), namely

$$\max_{e} \int_{X} dx \max_{d} \int_{\Theta} d\theta \, U(d, \theta, e, x) p(\theta|x, e) p(x|e). \tag{1.1}$$

Notice that maximising this function is unlikely to be easy, even when closed forms exist for the integral. If closed forms do not exist, then numerical methods will be required to solve the integral. Observe that these will have to be applied for each decision, parameter, experiment and data point; a potentially large, if not infinite, set.

Recent reviews of work in Bayesian experimental design may be found in Verdinelli (1992), Chaloner & Verdinelli (1995) which also contains an extensive reference list to the field, and Toman (1999).

We perceive the experiment as being used to analyse beliefs, or more precisely the beliefs of the expert or experimenter. As Goldstein (1994a; p118) writes

> It is the analysis of beliefs which is fundamental - the data analysis is of interest purely as an ingredient from which informed beliefs may be constructed.

Hence, we believe that the type of utility function used is completely specific to the experiment at hand. For example, the costs associated with the experiment, be they financial or ethical, are specific to the experiment and experimenter. Typically, we want to make inferences about various combinations of the elements of the model, but it is unlikely that all combinations will be equally important; indeed different experimenters may have different views as to which combinations of the elements they would like to focus attention upon. Thus, a group decision may be required and achieving group consensus is unlikely to be straightforward. Different objectives will lead to different choices of optimum (for example in terms of variance reduction for the quantities of interest). Often it is difficult for formal design criteria to capture all of our aims, so that we need to have heuristic insights into the types of information that each design conveys. Further, introducing fully specified prior beliefs over a complex model may be a difficult elicitation problem, and may also make the optimal choice of design highly complex computationally. For example, as Farrow & Goldstein (1992; p613) write

> ... there may be an excessively large number of possible designs from which to choose and no clear rules to guide our search.

The specification of a utility function for the problem is also likely to be intricate. Typically, it may involve costs of more than one type. For example, in medical experiments there are likely to be not only financial costs associated with performing the experiment but also ethical costs, such as potential side-effects that patients involved in the experiment might incur. Farrow & Goldstein (1992) point out that there may be a reluctance for the experimenter to specify trade-offs between the different costs and also between the costs and benefits. They propose using graphical methods to display the results and allow the experimenter to make the final choice. As they point out on p613, 'although we might not locate the "best" design, we can display the characteristics of a large number of designs including those which are "good" under various different trade-off regimes.' Another problem is pointed out by Lad & Deely (1994). Specifications of the utilities might depend on the relationship of the experimenter to the problem. Different doctors or concerned family members might proffer different values. A doctor may be more concerned with survival; a patient with quality of life. To quote Lad & Deely (1994; p276)

> The important thing here is to realize that some such value-judgement must be made if a design decision is to be concluded, and different people may make this evaluation in different respectable ways, depending on their personal concerns in the matter. It is in this way that the utilities become relevant to the particular problem being addressed.

Thus, experimental design with many variables cross-classified in many ways is a challenging problem for Bayes analysis. In this thesis, we shall concentrate upon looking at various designs, as opposed to solving a specific problem. Whilst we advocate that the individual nature of each experiment means that we should be solving specific problems as opposed to developing a toolbox of design, we emphasise once more that we view an experiment as a means of helping to analyse our beliefs and an understanding of how limited belief specifications are altered by data is an area that merits attention. Verdinelli (1992; p473) points out that "the need for deriving tractable analytical results has overshadowed the need for more realistic assumptions". Instead, we focus upon making modelling as accurately as possible, within our limited specification capabilities, whilst still hoping to maintain a tractable analysis. We aim to i) simplify the elicitation requirements ii) tame the computational problems for the design choice iii) provide qualitative insights into the effectiveness of the different choices of design.

We shall now review the methodology upon partial belief specifications that we wish to use. We motivate the methodology using the simple example we explain below.

1.2 The problem

An examination has been sat by a number of candidates. The exam consists of a number of compulsory questions, each question being marked out of ten. Any question not attempted receives a mark of zero. Each question consists of a number of distinct parts designed to test the candidates' abilities in a number of ways. For example, there is an initial straightforward part on which each candidate is expected to score highly, and then progressively harder parts, designed to separate the weaker students from the more able.

The marking of the exam is overseen by an examiner who has a number of immediate questions about the exam. Principally, he has to ensure that the exam is of the desired level of difficulty and also that there is not a discernable difference in the severity of the questions. If the examiner discovers any differences, he may rescale the marking schemes of the individual questions to try to eliminate the discrepancies. Of course, any changes to the mark scheme means that any scripts already marked would need to be reappraised. The examiner would like to minimise the number of scripts that need reappraising, but balance this against the desire to check that the exam is fair.

1.3 The quantification of uncertainty

The examiner is faced with uncertainty over the performance of the candidates in the exam. However, he has some knowledge, garnered through his years of teaching and examining, as to how he feels that students will fare on the exam; indeed much of this knowledge has already been used in the initial setting of the exam, and the design of the first marking scheme. Thus, the examiner is willing and able to quantify beliefs about certain quantities of interest and he would like to analyse these beliefs in the light of new information, the exam data. We now explore how he may go about doing this.

1.3.1 The subjectivist perspective

Throughout this thesis, we adopt a subjectivist approach to the revision of belief. Personally, it seems inherently clear that this is the right way to proceed. An event can occur once, and once only. For example, every toss of a coin is different: the air conditions will have, however slight, changed; the coin may have altered, however minimally, in shape as a result of it having been tossed previously; the tosser, be it machine or human, will be more fatigued; even for seemingly a straightforward process as tossing a coin more than once we can draw up sundry differences between the circumstances of each coin toss. Thus, attempting to found a theory based upon a frequency measure of repeatable events stumbles at the first hurdle. My viewpoint is that I am able to make an assessment of my own personal uncertain knowledge for any event I feel inclined to do so for. I also believe that any other individual is also able to make an assessment of their uncertainty for the same event and it may not be the same assessment I make. The quintessential account of the subjectivist perspective may be found in the work of Bruno de Finetti, in particular the two volume *Theory of Probability*, de Finetti (1974, 1975). The following quotation from the preface perhaps best summarises the position

> Probabilistic reasoning - always to be understood as subjective - merely stems from our being uncertain about something. It makes no difference whether the uncertainty relates to an unforseeable future, or to an unnoticed past, or to a past doubtfully reported or forgotten; it may even relate to something more or less knowable (by means of a computation, a logical deduction, etc.) but for which we are not willing or able to make the effort; and so on ... the only relevant thing is uncertainty - the extent of our own knowledge and ignorance. The actual fact of whether or not the events considered are in some sense *determined*, or known by other people, and so on, is of no consequence.

Cifarelli & Regazzini (1996) provide a summary of de Finetti's work as well as an extensive bibliography of his scientific output and of references dealing with later developments of his ideas. Chapter 8. of von Plato (1994) treats at length the ideas of de Finetti in the context of the development of probability theory.

The book by Kyburg & Smokler (1964/1980) provides a selection of articles on the early developments of the subjective standpoint, whilst the work of Savage (1954, 1981) should also be considered. Shafer (1986) provides a re-examination of Savage's (1954) argument for subjective expected utility. Fishburn (1986) surveys the development of subjective probability. The book by Lad (1996) attempts to make the viewpoint of de Finetti more accessible, and the first chapter provides an exemplary introduction to the philosophical questions and historical development of the subjectivist standpoint.

1.3.2 A Bayesian approach

We seek a methodology for organising and analysing our subjective beliefs in a systematic and logical fashion. The most familiar methodology is the so-called Bayesian approach. In a Bayesian approach, beliefs are represented by a joint probability measure for all of the random quantities of interest; a random quantity being any well-defined quantity about whose value we are uncertain, see de Finetti (1974; Section 2.3.3). Conditioning arguments allow us to update, via Bayes theorem (see Bayes (1764) and Laplace (1774/1986)), our beliefs about some of these random quantities given the values of the remaining random quantities. Text book developments of the Bayesian methodology may be found in, for example, Lindley (1965), Bernardo & Smith (1994), and O'Hagan (1994). On the surface, Bayesianism seems attractive, but there are immediate practical and logical dilemmas.

De Finetti (1973) draws a distinction between what he calls the Bayesian standpoint and Bayesian techniques. The latter consists of using standardized models and prior distributions and Bayes theorem to update beliefs, instead of, as de Finetti says, "carefully keeping realistic adherence to the specific features of each particular case and to the true opinion of the person concerned" and thus "Bayesian techniques, if considered as merely formal devices, are no more trustworthy than any other tool of the plentiful arsenal of the objectivist Statistics". We are concerned with adopting the Bayesian standpoint, and this is what we mean by Bayesian, where we attempt to model our true feelings for the problem at hand. However, even allowing for this, there are still problems with the adoption of the methods developed from this standpoint.

As the updating of beliefs in the Bayesian paradigm is done by conditioning, it is necessary that full probability distributions are specified. We require that, at least in principle, all possible data outcomes can be specified. As Goldstein (1981) points out, often qualitative data can not be expressed in such an exhaustive form; we will observe previously unconsidered features in our data outcomes. Full specification also means that an extremely large number of statements of prior knowledge are required to express beliefs to such a level of detail, a level that we often have neither the inclination nor ability to reach. The response to the discussion of Goldstein (1990) quickly shows how even a seemingly simple problem results in a "monster probability specification". Goldstein (1994a) shows how this problem is exacerbated when we have to tackle harder problems:

> "We require theory and methods to help us to think more clearly about our uncertainties. The larger and more complex

> the problem, the more such help is needed, but the less help formal methods seem able to provide. To make progress, it seems unavoidable that we must root the theory in our actual limited abilities to specify and analyse beliefs."

1.3.3 The Bayes linear approach

In an attempt to resolve these difficulties and to establish a fully subjective approach to belief revision, a series of papers by Goldstein (1981, 1986a, 1986b, 1988a, 1988b, 1991, 1994b) and Goldstein & Wooff (1998) have developed a system of partial belief adjustment based upon the revision of prevision, as defined in the following section. Summaries of the methodology may be found in Goldstein (1999, 2000). Goldstein (1994a) lists an irreducible minimum of belief considerations that allow the construction of a fully subjective approach to belief revision:

1. quantitative judgements of belief about the magnitudes of the quantities of interest
2. expression of uncertainty in our judgements of the magnitudes
3. judgement on the relationship between the magnitudes of the quantities

A simple quantification is sought, in order to keep the specification process as simple as possible and this is achieved through expectation statements; the above quantifications being achieved by the expectation of the quantity, the variance of the quantity and the covariance between the two quantities, respectively. Specification of the expectations is made directly, so that expectation is treated as primitive. This approach has been termed the Bayes linear approach and is the approach we shall follow in this thesis. Linear Bayesian methods have been considered by other authors, see for example Stone (1963) and Hartigan (1969) .

1.3.4 Prevision

In treating expectation as primitive, we follow the development of de Finetti (1974, 1975). An alternative treatment of expectation as primitive may be found in Whittle (1992). His approach is from a different foundational perspective, arguing that the long-term average of a variable is empirically meaningful and that expectation is the idealization of this. Thus, his approach is similar to that of the frequentist approach to probability.

De Finetti (1974; Chapter 3) suggests that when confronted with uncertainty we do not remain agnostic between possible alternatives but instead feel a strong inclination that certain alternatives, as opposed to others, will turn out to be true. As de Finetti (1974; p72) writes

> "Uncertain things remain uncertain, but we attribute to the various uncertain events a greater or lesser degree of that new factor which is extralogical, subjective and personal and which expresses these attitudes ... Prevision, in the sense we give to the term and approve of, consists in considering, after careful reflection, all the possible alternatives, in order to distribute among them, in the way which will appear most appropriate, one's own expectations, one's own sensations of probability."

To each random quantity X, there corresponds the individual's evaluation $E(X)$, the prevision of X. We seek an operationally defined measurement of $E(X)$. Lad (1996; Section 2.1) defines this to be "a specified procedure of action which, when followed, yields a number". In terms of prevision, this reduces to finding the fair price of X in the following sense. Your uncertain knowledge is quantified by you evaluating your preference between having a claim to a specified gain (eg. a specified amount of money) and a claim to an unknown gain (eg. an unknown amount of money), X. The specified gain that you considered equivalent to X is the fair price of X and is called your prevision, being denoted by $E(X)$. By equivalent we mean that you would willingly exchange the unknown gain, X, for the specified certain gain, $E(X)$ and vice versa.

De Finetti (1974; Section 3.3.6) gives a criterion for obtaining the measurement. $E(X)$ is the value $\overline{x}$ which you would choose if, having made this choice, you were to suffer a penalty L given by

$$L = \left(\frac{X - \overline{x}}{k}\right)^2 \tag{1.2}$$

for some unit of loss, k. If the quantity $X = E$, an event, then your prevision is also called your probability for the event, denoted $P(E)$. As Wilkinson (1995) points out, this definition is not without its flaws, for example it makes no recourse to utility. The definition assumes that for money bets we make an assumption of, what de Finetti calls, "rigidity in the face of risk". However, this is not really valid for individuals tend to be risk averse for large sums of money, and the value of a certain sum of money is not the same for each individual. In his addendum to

Section 2.4, Lad (1996) poses the problem of "how much would you pay to receive \$1 if your local nuclear power plant has an accident and releases high intensity radioactive contaminants into the river?" As Lad explains, the resolution to this is not easy, requiring not only consideration of the utility for the monetary yield, but also the utility of what's happening. De Finetti (1974; p77) asks the question:

> "Are the conclusions which we draw after observing the actual behaviour of an individual, directly making decisions in which he has a real interest, more reliable than those based on the preferences which he expresses when confronted with a hypothetical situation or decision? Both the direct interest and the lack of it might on the one hand favour, and on the other obstruct, the calmness and accuracy, and hence the reliability, of the evaluations."

In the power plant example, the consequences of winning the bet, namely that of the contamination of the river and the consequences of that might mean that we'd prefer to lose the bet. Ramsey (1926) (see Kyburg & Smokler (1964; p77)) is also aware of the difficulties posed by such propositions and he considers that propositions where the options offered are not of desire are ethically neutral. Thus, the power plant proposition is not an ethically neutral proposition. A similar consideration is taken by Lad (1996; p59, p63-67) in one of his three qualifications to the definition of prevision. De Finetti is aware of such problems and .considers them in Section 3.2 of his magnum opus and also Footnote a. in Kyburg & Smokler (1964; p102) of the translation of de Finetti (1937), but these are more digressions than solutions. Solutions have been suggested, for example the loss could be considered in units of probability currency, see Walley (1991). To delve too deeply into these dilemmas would be to move away from the thrust of this thesis and so we shall proceed assuming that there is a well-defined prevision, although always bearing in mind de Finetti's (1974; p76) observation that "every measurement procedure and device should be used with caution, and its results carefully scrutinized".

There is one requirement of prevision and that is that your assertions are coherent. By this we mean that you do not assert previsions for which you will make a certain loss. De Finetti shows that the following two considerations are necessary and sufficient for coherence:

$$E(X+Y) = E(X)+E(Y); \tag{1.3}$$

$$\inf X \le E(X) \le \sup X. \tag{1.4}$$

1.3.5 The geometric interpretation of prevision

De Finetti (1974; Sections 2.8, 4.17) formulated a geometric interpretation of an individual's current previsions. For a general collection of random quantities $\mathcal{B} = \{\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2, \dots\}$, construct the space $\langle\mathcal{B}\rangle$ of finite linear combinations of the elements of $\mathcal{B}$. The quantity $\mathcal{B}_0$ is a random quantity taking the value $\mathcal{B}_0 = 1$ with certainty (see de Finetti (1974; p51)). A typical element, $\mathcal{A}$, of $\langle\mathcal{B}\rangle$ is then

$$\mathcal{A} \;=\; v_0 + \sum_u \nu_u \mathcal{B}_{v_u}, \tag{1.5}$$

where $\{v_1, v_2, \dots\}$ is a general finite subset of integers and $v_0 = v_0\mathcal{B}_0$. We view $\langle\mathcal{B}\rangle$ as a vector space by considering each $\mathcal{B}_v$ as a vector with linear combinations, $\mathcal{A}$, of random quantities represented as the corresponding linear combination of vectors. The linear representation is such that specifying previsions for each $\mathcal{B}_v$ uniquely fixes your prevision for all linear combinations, $\mathcal{A}$, but not in general for any other function. Thus, random quantities $\hat{\mathcal{B}}$ and $\hat{\mathcal{B}}^2$ would correspond to different vectors $\mathcal{B}_v$ and $\mathcal{B}_{v'}$ in our representation as they are not, in general, linearly related.

Having fixed the linear structure, we add the geometric framework by forming the inner product space $[\mathcal{B}]$ from the minimal closure of $\langle\mathcal{B}\rangle$ by imposing the following inner product and norm for $\mathcal{A}$, $\mathcal{A}^* \in \langle\mathcal{B}\rangle$,

$$(\mathcal{A}, \mathcal{A}^*) \;=\; Cov(\mathcal{A}, \mathcal{A}^*); \tag{1.6}$$
$$\| \mathcal{A} \|^2 \;=\; Var(\mathcal{A}). \tag{1.7}$$

We restrict $\mathcal{B}$ to elements with finite prior variance. This inner product generated by covariance is, as de Finetti (1974; Section 4.17) argues, a natural way to form a geometric interpretation of prevision. Since the norm of a quantity corresponds to its standard deviation then large norms are attached to quantities with large uncertainty, whilst orthogonality between two quantities corresponds to these quantities being uncorrelated. We have to be careful though for strictly this inner product should be formed over the closure of the equivalence classes of random quantities which differ by a constant. If we neglect to do so, then we will have non-zero vectors (non-zero constants) corresponding to a zero norm. Goldstein (1981; p108) follows the convention of standardising every quantity $\mathcal{A}$ by subtracting its prior mean rather than use equivalence classes. Goldstein (1986b; p200) then explains how by considering the inner product, $[\mathcal{B}^*]$, given by $\{\mathcal{A}, \mathcal{A}^*\} = E(\mathcal{A}\mathcal{A}^*)$ over the unstandardised quantities we may understand the inner product given by covariance over the standardised quantities. Observe that $E(\mathcal{A}) = \{\mathcal{A}, \mathcal{B}_0\}\mathcal{B}_0$ so that the prevision of each $\mathcal{A}$ is the projection of $\mathcal{A}$ into $[\mathcal{B}_0]$. The inner product given by covariance over the standardised quantities, $[\mathcal{B}]$, is then equivalent to the subspace $[\mathcal{B}_0]^{\perp}$ of $[\mathcal{B}^*]$. In the work that we shall develop in this thesis, our primary interest lies in measuring uncertainty and examining the changes in uncertainty in the light of data as opposed to the previsions explicitly and so we may proceed by using the standardised quantities. However, for completeness, in this introductory chapter we shall give the details of how we may revise our previsions.

Goldstein (1986a) terms $[\mathcal{B}]$ a belief structure with base $b([\mathcal{B}]) = \mathcal{B}$. Two subspaces $[\mathcal{B}^*]$ and $[\mathcal{B}^\dagger]$ are said to be orthogonal, written $[\mathcal{B}^*] \perp [\mathcal{B}^\dagger]$, if every element of the collection $\mathcal{B}^*$ is uncorrelated with every element of $\mathcal{B}^\dagger$.

The base $\mathcal{B}$ will not contain all the quantities for which you are prepared to give a prevision, rather it will be a subset of your choice. For example, it may contain only those random quantities that you are interested in revising beliefs about in the light of new information.

Consider the examiner. He is about to make a series of measurements, such as the total score on question v, on an exam candidate. He introduces into $\mathcal{B}$ the quantities he will explicitly express beliefs about. He must decide whether to include quantities such as (score on question $v)^2$, (score on question v/score on question w) and so on as well as the simple score quantities. He has the option of not including some of these quantities (most likely, he will choose to exclude almost all of them). Contrast this with the scenario under the usual Bayes specification. Here, for any specified quantity, it is an implicit requirement to state the prevision of all possible functions of the quantity. Often, this is an incredibly difficult and largely unnecessary task.

Constrained only by the requirement that we are able to make the specifications, we may include as many functional forms as we feel relevant to the problem. In this way, we may specify whatever product order of moments we desire. A full probability specification is then the limit of the above approach in specifying all joint prior moments. Hence, the above framework affords us the opportunity to work with whatever level of detail we feel appropriate to the problem in question.

1.4 Formalising the problem

At this point, we draw the reader's attention to Appendix A which contains a summary of potentially useful notation.

We have a sequence of individuals upon whom we wish to make a series of measurements. We are interested in traits common to the individuals and so we elect to make the same series of measurements on each individual. We gather these together as the collection $\mathcal{C} = \{X_1, X_2, \dots\}$, finite or infinite, where each X_v is a real valued function of the quantities that will be measured on the individuals. From $\mathcal{C}$ we generate the collection for each individual. Let $\mathcal{C}_i = \{X_{1i}, X_{2i}, \dots\}$ be the values of the measurements for the ith individual. Thus, X_{vi} denotes the value of the vth quantity for the ith individual and $\mathcal{C}_i$ is a collection of observable random quantities. Form the full collection $\mathcal{C}^*$, the union of all of the elements in all of the collections $\mathcal{C}_i$, where i ranges over some indexing set of individuals. We specify our prior mean, variance and covariance for each pair of quantities.

1.4.1 The examiner's problem

The examiner decides that he is only willing and able to specify beliefs about the marks awarded on each question. For simplicity of exposition, we shall consider that there are only two questions; the execution of the theory is the same for any number of questions. Thus, he intends to make a series of measurements $\mathcal{C} = \{X_1, X_2\}$ on a sample from a collection of exam candidates. From $\mathcal{C}$ he generates the collection for each candidate. Let $\mathcal{C}_i = \{X_{1i}, X_{2i}\}$ be the values of the measurements for the ith candidate. Thus, X_{vi} denotes the mark on the vth question for the ith candidate sitting the exam. The examiner forms the full collection $\mathcal{C}^*$, the collection of all the marks for all of the candidates sitting the exam. The examiner specifies his prior mean, variance and covariance for each pair of quantities.

1.5 Second-order exchangeability

The examiner, who does not know the identity of the candidates, considers the sequence of exam results for the individuals. He judges that there are similarities between his beliefs in that his beliefs seem not to depend on the individuals, in short the individuals are exchangeable. Having adopted an approach of limited belief specification, our interest lies in second-order exchangeability, but first we must make clear what we mean by exchangeability and also make explicit any problems we may find with this approach.

1.5.1 Full exchangeability

De Finetti (1931) introduced the concept that was to become known as exchangeability in terms of personal judgements about sequences of quantities. Kyburg & Smokler (1980; p15) suggest that the concept of exchangeable events is the most crucial component for the subjective theory for it provides a bridge between the classical procedures of statistical inference and subjective probability. The paper of de Finetti (1937) contains the details of this link. To quote Kyburg & Smokler (1980; p15-6)

> "de Finetti shows that the classical limit theorems on which many forms of statistical inference depend hold just as well for sequences of exchangeable events as they do for the sequences of independent, equiprobable events to which they have traditionally been applied. He shows, for example, that in the case of a sequence of exchangeable events, a person, whatever be the opinion with which he starts out, must, if he is to be coherent in his beliefs, after a sufficient amount of observation come to assign a probability to the type of event in question that is close to its observed relative frequency."

He does this through what became known as 'de Finetti's representation theorem'. Although some of the results had been derived before (see Dale (1985) for a detailed study) it was de Finetti's understanding of exchangeability in terms of subjectivity that was the breakthrough; as he says himself "the above-mentioned representation theorem, together with every other more or less original result in my conception of probability theory, should not be considered as a discovery. Everything is essentially the fruit of a thorough examination of the subject matter, carried out in an unprejudiced manner, with the aim of rooting out nonsense" (de Finetti (1974; pxii)).

So what do we mean by exchangeability? We shall borrow an example from de Finetti (1959). Suppose that we have a sequence of trials of the same phenomenon, for example an archer firing arrows at a target from the same location. Suppose that after six attempts, he has recorded the sequence of successes S and failures F in the order FFFSFS. How might we view this sequence, especially if we wish to consider future throws? We might see it as the start of an improvement that is bound to continue, or that odd-numbered trials are almost all failures whilst the even trials have a greater proportion of successes; there could be any number of systematic dependencies upon order that we perceive. However, in the case of exchangeability, we wish to exclude notions of dependence on ordering, exchangeability is where we perceive symmetry with respect to order. O'Hagan (1994; Section 4.40) provides a definition for exchangeable events that we use here.

Definition 1 *You regard events $A_1, A_2, \ldots, A_n$ as being exchangeable if*

$$P\left(\cap_{j=1}^{k} A_{i_j}\right) = P\left(\cap_{j=1}^{k} A_j\right) \tag{1.8}$$

for all $k = 1, 2, \ldots, n$ and for all $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

Notice that since we are dealing with the probability of events, this definition is consistent with our prevision notation. As O'Hagan points out, it is clear from this definition how the belief of exchangeability reduces the specification burden from $2^n - 1$ probabilities to the n probabilities

$$d_k = P\left(\cap_{i=1}^{k} A_i\right) \text{ for } k = 1, \ldots, n. \tag{1.9}$$

This definition may be extended to cover an infinite sequence of events. The infinite sequence of events $A_1, A_2, \ldots$ are (infinitely) exchangeable if every finite subsequence is exchangeable. We are now in a position to state the representation theorem for exchangeable events. We adapt Proposition 4.1 of Bernardo & Smith (1994).

Theorem 1 *If $A_1, A_2, \ldots$ is an infinitely exchangeable sequence of events (so $A_1 = 1$ if A_1 occurs, 0 otherwise), there exists a distribution function Q such that*

$$P(A_1, \ldots, A_n) = \int_0^1 \prod_{i=1}^{n} \theta^{A_i}(1-\theta)^{1-A_i} dQ(\theta), \tag{1.10}$$

where,

$$Q(\theta) = \lim_{n \to \infty} P\left[\frac{Y_n}{n} \leq \theta\right], \tag{1.11}$$

with $Y_n = \sum_{i=1}^{n} A_n$, and $\theta = \lim_{n\to\infty}(Y_n/n)$.

Where the limits mean convergent in distribution; see Bernardo & Smith (1994, section 3.2.3). Thus, it is as if, conditional on θ, the $A_1, \ldots, A_n$ are a random sample from a Bernoulli distribution with parameter θ.

We may extend our exchangeability definition to include random quantities. Borrowing from Lad (1996; Definition 5.14) we make the following definition:

Definition 2 *You regard random quantities* $X_1, X_2, \ldots, X_n$ *as being exchangeable if you assert a prevision for the product events of the form* $(X_{i_1} = x_{i_1})(X_{i_2} = x_{i_2}) \cdots (X_{i_k} = x_{i_k})$ *that is constant for every permutation of the numbers* $x_{i_1}, x_{i_2}, \ldots,$ x_{i_k} *for all* $k = 1, 2, \ldots, n$ *and for all* $1 \leq i_1 < i_2 < \ldots < i_k \leq n$.

The representation theorem for exchangeable events may then be generalised to encompass exchangeable random quantities. Further details on exchangeability may be found in Galambos (1982) and the references therein, whilst detailed references to representation theorems for exchangeable sequences may be found in Section 4.8.1. of Bernardo & Smith (1994).

1.5.2 Problems with full exchangeability

As Goldstein (1994b) points out, there are two major drawbacks to the representation theorem. Firstly, although we have reduced the prior specification process through the judgement of exchangeability, we are still required to make a specification which is far more detailed than it would ever be possible for us to make. We need to specify the d_k's of equation (1.9) for all possible values of k, an impossible task, unless we start with the representation theorem.

Difficulties also arise with our beliefs about future quantities in the sequence and our attempts to use induction. Point 5. of de Finetti (1959) (see de Finetti (1972; p194)) states that:

> "Inductive reasoning is nothing other than reckoning $P(H|E)$, the probability of H after the observation of E, in accordance with Bayes' theorem-or, equivalently, according to the theorem of compound probability, of which Bayes' theorem is a corollary."

As de Finetti points out, $P(H|E)$ and $P(\overline{H}|E)$ are proportional to $P(EH)$ and $P(E\overline{H})$ so that experience adds no new elements for evaluation, it merely reduces the number of possible cases. Thus, to return to the archer problem of the previous section, if we wish to evaluate the probability of hitting the target on the seventh attempt we need only compare the probabilities that would have been attributed to the cases FFFSFSS and FFFSFSF. In particular, if we viewed the attempts as being exchangeable initially, we would also need to view the future shots as being exchangeable for the inductive argument of point 5 to hold. There are various problems with this. Firstly, why is there any link between our current and future beliefs? Goldstein (1997) concerns itself purely with this question and suggests that the inductive reasoning suggested above should act as a prior inference for our posterior judgements. Furthermore, why should we be certain that our future events will be exchangeable? Is our judgement of exchangeability not something that we might want to explore through sampling? Goldstein (1994b) considers the issues at hand in revising exchangeable beliefs.

1.5.3 Second-order exchangeability

As explained in the previous section, full exchangeability imposes constraints upon our beliefs that are so harsh, that, in practice, it is unlikely that they are ever likely to be realised. We restrict attention instead upon second-order exchangeability which places sufficient structure to allow for tractable belief revision without placing too harsh a constraint upon our prior beliefs.

We regard a sequence of quantities as second-order exchangeable if our first and second-order beliefs about the sequence is unaffected by permuting the order of the sequence. We have the following definition:

Definition 3 *The collection of measurements C is second-order exchangeable over the full collection C^* if*

$$\begin{aligned} E(X_{vi}) &= m_v \ \forall v, i; && (1.12)\\ Cov(X_{vi}, X_{wi}) &= d_{vw} \ \forall v, w, i; && (1.13)\\ Cov(X_{vi}, X_{wj}) &= c_{vw} \ \forall v, w, i \neq j. && (1.14) \end{aligned}$$

If C^* is, at least in principle, the union of an infinite number of individuals' collections, then we say that the sequence is infinitely exchangeable.

Initially, the examiner judged that the candidates were exchangeable. Additionally, the examiner also judged that the questions were exchangeable. He made the following specifications:

$$\begin{aligned} E(X_{vi}) &= 0 \ \forall v, i; && (1.15)\\ Cov(X_{vi}, X_{wi}) &= \begin{cases} 5 & \text{if } v = w, \ \forall i; \\ 3.75 & \text{if } v \neq w, \ \forall i; \end{cases} && (1.16)\\ Cov(X_{vi}, X_{wj}) &= \begin{cases} 1 & \text{if } v = w, \ \forall i \neq j; \\ 0.25 & \text{if } v \neq w, \ \forall i \neq j. \end{cases} && (1.17) \end{aligned}$$

Thus, the examiner has chosen to standardise his quantities to have mean zero.

1.6 The representation theorem for infinite second-order exchangeable beliefs

Goldstein (1986a) derives a representation theorem for infinitely exchangeable measurements as follows.

Theorem 2 *If $\mathcal{C}$ is second-order exchangeable over the individuals, then we may introduce the further collections of random quantities $\mathcal{M}(\mathcal{C}) = \{\mathcal{M}(X_1), \mathcal{M}(X_2), \ldots,\}$, and $\mathcal{R}_i(\mathcal{C}) = \{\mathcal{R}_i(X_1), \mathcal{R}_2(X), \ldots\}$, and write*

$$X_{vi} = \mathcal{M}(X_v) + \mathcal{R}_i(X_v), \tag{1.18}$$

where

$$\mathcal{M}(X_v) = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} X_{vi}; \tag{1.19}$$

the limit here being in mean square (see Bernardo & Smith (1994, section 3.2.3). The collections $\mathcal{M}(\mathcal{C})$ and $\mathcal{R}_i(\mathcal{C})$ satisfy the following relationships

$$E(\mathcal{M}(X_v)) = m_v \ \forall v; \tag{1.20}$$

$$E(\mathcal{R}_i(X_v)) = 0 \ \forall v, i; \tag{1.21}$$

$$Cov(\mathcal{M}(X_v), \mathcal{M}(X_w)) = c_{vw} \ \forall v, w; \tag{1.22}$$

$$Cov(\mathcal{M}(X_v), \mathcal{R}_j(X_w)) = 0 \ \forall v, w, j; \tag{1.23}$$

$$Cov(\mathcal{R}_i(X_v), \mathcal{R}_j(X_w)) = \begin{cases} d_{vw} - c_{vw} & \text{if } i = j \ \forall v, w; \\ 0 & \text{otherwise.} \end{cases} \tag{1.24}$$

The quantities in $\mathcal{M}(\mathcal{C})$ and in the $\mathcal{R}_i(\mathcal{C})$ are not observable. Instead, the quantities in $\mathcal{M}(\mathcal{C})$ may be viewed as being analogous to underlying population means for the various quantities, with the collections $\mathcal{R}_i(\mathcal{C})$ viewed as the individual discrepancies from the overall means. As noted in Goldstein (1994b), this representation theorem is operational in practice as well as in principle. We are required solely to elicit the actual belief statements m_v, d_{vw}, c_{vw} over the observable sequence, as described in equations (1.12) - (1.14). Thus, application of the representation theorem here remains feasible.

We may extend the representation theorem to encompass linear combinations. Let $\langle\mathcal{C}\rangle$ be the collection of finite linear combinations of the elements of $\mathcal{C}$, so that

$$\mathcal{X} = \sum_u \eta_u X_{v_u}, \tag{1.25}$$

where $\{v_1, v_2, \dots\}$ is a general finite subset of integers, is a typical element of $\langle \mathcal{C} \rangle$. For each $\mathcal{X} \in \langle \mathcal{C} \rangle$, we construct the corresponding quantity for each individual i, namely the value

$$\mathcal{X}_i = \sum_u \eta_u X_{v_u i}, \tag{1.26}$$

and denote by $\langle \mathcal{C}_i \rangle$ the collection of finite linear combinations of $\{Y_i\}$. We may then apply the representation theorem to $\mathcal{X} \in \langle \mathcal{C} \rangle$, for each individual i, to yield

$$\mathcal{X}_i = \mathcal{M}(\mathcal{X}) + \mathcal{R}_i(\mathcal{X}), \tag{1.27}$$

where $\mathcal{M}(\mathcal{X}) = \sum_u \eta_u \mathcal{M}(X_{v_u}) \in \langle \mathcal{M}(\mathcal{C}) \rangle$ and $\mathcal{R}_i(\mathcal{X}) = \mathcal{X}_i - \mathcal{M}(\mathcal{X}) = \sum_u \eta_u \mathcal{R}_i(X_{v_u})$ $\in \langle \mathcal{R}_i(\mathcal{C}) \rangle$. $\langle \mathcal{M}(\mathcal{C}) \rangle$ and $\langle \mathcal{R}_i(\mathcal{C}) \rangle$ are, respectively, the collection of finite linear combinations of the elements of $\mathcal{M}(\mathcal{C})$ and $\mathcal{R}_i(\mathcal{C})$.

1.6.1 Applying the representation theorem for the example

The examiner reasons that the students taking the exam are from a (potentially) infinite population. Thus, from the specifications the examiner has made, we may introduce the collection of random quantities $\mathcal{M}(\mathcal{C}) = \{\mathcal{M}(X_1), \mathcal{M}(X_2)\}$, the collection of underlying mean scores for the questions, and $\mathcal{R}_i(\mathcal{C}) = \{\mathcal{R}_i(X_1), \mathcal{R}_i(X_2)\}$, the individual residuals from the mean score and write:

$$X_{vi} = \mathcal{M}(X_v) + \mathcal{R}_i(X_v). \tag{1.28}$$

The induced beliefs over the newly introduced random quantities are as follows:

$$E(\mathcal{M}(X_v)) = 0 \ \forall v; \tag{1.29}$$

$$E(\mathcal{R}_i(X_v)) = 0 \ \forall v, i; \tag{1.30}$$

$$Cov(\mathcal{M}(X_v), \mathcal{M}(X_w)) = \begin{cases} 1 & \text{if } v = w; \\ 0.25 & \text{if } v \neq w; \end{cases} \tag{1.31}$$

$$Cov(\mathcal{M}(X_v), \mathcal{R}_j(X_w)) = 0 \ \forall v, w, j; \tag{1.32}$$

$$Cov(\mathcal{R}_i(X_v), \mathcal{R}_j(X_w)) = \begin{cases} 4 & \text{if } i = j \text{ and } v = w; \\ 3.5 & \text{if } i = j \text{ and } v \neq w; \\ 0 & \text{otherwise.} \end{cases} \tag{1.33}$$

1.7 The adjustment of beliefs

We have looked at how the examiner may go about quantifying his beliefs about the examination. However, the examiner is concerned with the effect of new information on the collection of previsions he has specified. In terms of our development, the examiner is interested in learning about various linear combinations of the $\mathcal{M}(\mathcal{C})$. For example, he is interested in the overall total,

$$\mathcal{M}(X_+) = \mathcal{M}(X_1) + \mathcal{M}(X_2), \tag{1.34}$$

as a means of assessing the difficulty of the paper, and also

$$\mathcal{M}(X_-) = \mathcal{M}(X_1) - \mathcal{M}(X_2), \tag{1.35}$$

illustrating the difference in marks between the first and the second question. We need to consider the effect of new information upon the beliefs.

1.7.1 Revising previsions

The previous sections have summarised the static features of beliefs. In a Bayes linear analysis, we are interested in what happens to the elements of the geometric structure if we receive some new information. Suppose that we are to receive the values of a data collection $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2, \dots)$, where $\mathcal{D}_0$ is the unit constant. This will cause us to revise the values of the previsions we had assigned for the elements in $[\mathcal{B}]$.

Definition 4 *For a random quantity $\mathcal{A} \in \langle \mathcal{B} \rangle$, the linear combination*

$$E_{\mathcal{D}}(\mathcal{A}) = \xi_0 + \sum_u \xi_u \mathcal{D}_{v_u}, \tag{1.36}$$

where the coefficients are chosen to minimise

$$E\{(\mathcal{A} - \xi_0 - \sum_u \xi_u \mathcal{D}_{v_u})^2\} \tag{1.37}$$

over all collections $(\xi_0, \xi_1, \dots)$, is termed the adjusted expectation of $\mathcal{A}$ given $\mathcal{D}$. The minimal value is termed the adjusted variance of $\mathcal{A}$ given $\mathcal{D}$, written $Var_{\mathcal{D}}(\mathcal{A})$.

As Goldstein (1983) shows, to be coherent adjusted expectation requires that

1. For any quantities $\mathcal{A}$, $\mathcal{A}^*$ and constants θ and ϑ, we have

$$E_{\mathcal{D}}(\theta\mathcal{A} + \vartheta\mathcal{A}^*) = \theta E_{\mathcal{D}}(\mathcal{A}) + \vartheta E_{\mathcal{D}}(\mathcal{A}^*). \tag{1.38}$$

2. For any $\mathcal{A}$, we have

$$E(E_{\mathcal{D}}(\mathcal{A})) = E(\mathcal{A}). \tag{1.39}$$

An equivalent definition of $E_{\mathcal{D}}(\mathcal{A})$ is as the element $\mathcal{A}^{\ddagger} \in [\mathcal{D}]$ which minimises $\| \mathcal{A} - \mathcal{A}^{\ddagger} \|$. Thus, $E_{\mathcal{D}}(\mathcal{A})$ is the orthogonal projection of $\mathcal{A}$ into $[\mathcal{D}]$. $Var_{\mathcal{D}}(\mathcal{A})$ is then the squared orthogonal distance from $\mathcal{A}$ to $[\mathcal{D}]$. We may thus decompose $\mathcal{A}$ into the sum of two uncorrelated components,

$$\mathcal{A} = E_{\mathcal{D}}(\mathcal{A}) + \{\mathcal{A} - E_{\mathcal{D}}(\mathcal{A})\}. \tag{1.40}$$

Definition 5 *The residual form*

$$[\mathcal{A}/\mathcal{D}] = \mathcal{A} - E_{\mathcal{D}}(\mathcal{A}) \tag{1.41}$$

is called the adjusted version of $\mathcal{A}$.

Adjusted quantities thus obey the following properties:

$$E([\mathcal{A}/\mathcal{D}]) = 0; \tag{1.42}$$

$$Cov(E_{\mathcal{D}}(\mathcal{A}), [\mathcal{A}/\mathcal{D}]) = 0. \tag{1.43}$$

Hence, by taking the variance of equation (1.40) and using equation (1.43), we may decompose the variance of $\mathcal{A}$ as

$$Var(\mathcal{A}) = Var(E_{\mathcal{D}}(\mathcal{A})) + Var([\mathcal{A}/\mathcal{D}]) \tag{1.44}$$

$$= Var(E_{\mathcal{D}}(\mathcal{A})) + Var_{\mathcal{D}}(\mathcal{A}). \tag{1.45}$$

As Hartigan (1969; p447) points out, if $\mathcal{A}$, $\mathcal{D}_1$, $\mathcal{D}_2$, $\ldots$ are jointly normal, then these definitions coincide with the usual definitions of conditional expectation and variance; that is $E_{\mathcal{D}}(\mathcal{A}) = E(\mathcal{A}|\mathcal{D})$ and $Var_{\mathcal{D}}(\mathcal{A}) = Var(\mathcal{A}|\mathcal{D})$. In linear regression analysis, Goel & DeGroot (1980) show that only normal distributions have linear posterior expectations. Goldstein (1976; p51) is aware of such restrictions and proposes a Bayes linear formulation of the regression model; this is an early version of the Bayes linear methodology he goes on to develop. In Section 1., Goel & DeGroot (1980) also provide a brief summary and references to examples when the posterior expectation of the population mean is a linear function of the sample observation. The first interpretation of belief adjustment of Goldstein (1999; p30) provides an explanation to the link to normality, but as he explains in the fourth interpretation, there are more foundational perspectives. This provides a detailed argument for the justification of Hartigan (1969; p452) when he writes

> "A referee has expressed reservations about the usefulness of linear expectation and variance - "Why should a Bayesian be

interested in such things? Are they meant to represent approximations to the expectation and variance of the posterior distribution? If so, then the restriction to linear expectation is no more innocuous than is the assumption of normality itself." A Bayesian (or non-Bayesian) should be interested in linear expectation and variance because it is possible to use them, without requiring the full Bayesian probability apparatus, but still retaining the basic Bayesian form of combining prior and present data."

Having observed $\mathcal{D}$, we would like a means of assessing the impact of the adjustment upon $\mathcal{A}$. Observe equation (1.45). $Var(E_{\mathcal{D}}(\mathcal{A}))$ is the variability of $\mathcal{A}$ which is accounted for by the variability in $\mathcal{D}$ as $E_{\mathcal{D}}(\mathcal{A})$ is determined when we observe $\mathcal{D}$. $Var([\mathcal{A}/\mathcal{D}])$ is the residual variance of $\mathcal{A}$ given $\mathcal{D}$. We thus make the following definition.

Definition 6 *The variance of $\mathcal{A}$ resolved by $\mathcal{D}$ is defined to be the quantity*

$$RVar_{\mathcal{D}}(\mathcal{A}) \quad = \quad Var(E_{\mathcal{D}}(\mathcal{A})). \tag{1.46}$$

Thus, from equation (1.45), we write the variance partition for $\mathcal{A}$ as

$$Var(\mathcal{A}) \quad = \quad Var_{\mathcal{D}}(\mathcal{A}) + RVar_{\mathcal{D}}(\mathcal{A}). \tag{1.47}$$

Thus, intuitively, the observation of $\mathcal{D}$ is expected to be informative for $\mathcal{A}$ if $Var_{\mathcal{D}}(\mathcal{A})$ is large relative to $Var(\mathcal{A})$. This motivates the following definition.

Definition 7 *The resolution, $R_{\mathcal{D}}(\mathcal{A})$, for $\mathcal{A}$ induced by $\mathcal{D}$ is defined to be the quantity*

$$R_{\mathcal{D}}(\mathcal{A}) \quad = \quad \frac{Var(\mathcal{A}) - Var_{\mathcal{D}}(\mathcal{A})}{Var(\mathcal{A})}. \tag{1.48}$$

The resolution ratio, $RR_{\mathcal{D}}(\mathcal{A})$, for $\mathcal{A}$ induced by $\mathcal{D}$ is defined to be the quantity

$$RR_{\mathcal{D}}(\mathcal{A}) \quad = \quad \frac{Var(\mathcal{A}) - Var_{\mathcal{D}}(\mathcal{A})}{Var_{\mathcal{D}}(\mathcal{A})}. \tag{1.49}$$

The resolution and resolution ratio give equivalent simple, scale free, measures of the impact of the adjustment upon $\mathcal{A}$. If $R_{\mathcal{D}}(\mathcal{A})$ is near zero, then, relative to our prior knowledge about $\mathcal{A}$, we do not expect the Bayes linear analysis of the sample to be informative for $\mathcal{A}$. A value of $R_{\mathcal{D}}(\mathcal{A})$ close to one suggests that the Bayes linear analysis of the sample is expected to be highly informative for $\mathcal{A}$. Likewise, if $RR_{\mathcal{D}}(\mathcal{A})$ is very large, then we expect the Bayes linear analysis of the sample to be highly informative. Typically, see Goldstein & Wooff (1997, 1998) for examples, we may use such summaries to aid the choice of optimal sample sizes for design problems where variance reduction of certain quantities of interest is viewed as a benefit, balanced against the cost of the sample.

1.7.2 Revising belief structures

In a belief analysis, we are interested not just how our beliefs change with the observation of data for a single quantity $\mathcal{A}$ but also for the entire belief structure, $[\mathcal{B}]$. Goldstein (1988a) deals extensively with this problem and derives various properties of such adjustments. We shall briefly summarise them in this section. For any collection $\mathcal{K} = \{\mathcal{K}_1, \mathcal{K}_2, \ldots\}$ of random quantities, $[\mathcal{K}/\mathcal{D}]$ denotes the set $\{[\mathcal{K}_1/\mathcal{D}], [\mathcal{K}_2/\mathcal{D}], \ldots\}$.

Definition 8 *If $[\mathcal{B}]$ is a belief structure with base $b([\mathcal{B}])$, then the belief structure $[\mathcal{B}/\mathcal{D}]$ is the belief structure with base $b([\mathcal{B}/\mathcal{D}]) = [b([\mathcal{B}])/\mathcal{D}]$. $[\mathcal{B}/\mathcal{D}]$ is the belief structure $[\mathcal{B}]$ adjusted by $\mathcal{D}$.*

The following properties of adjusted belief structures are described in Section 2.6 of Goldstein (1988a).

Theorem 3 *Adjusted belief structures satisfy the following properties:*

1. *(a) $[\mathcal{B}/\mathcal{D}] = 0$ if and only if $[\mathcal{B}]$ is contained in $[\mathcal{D}]$.*

   (b) $[\mathcal{B}/\mathcal{D}] = [\mathcal{B}]$ if and only if $[\mathcal{B}]$ and $[\mathcal{D}]$ are orthogonal spaces.

2. *For any $[\mathcal{B}]$, $[\mathcal{B}^\dagger]$ and $[\mathcal{D}]$, we have*

$$[(\mathcal{B} + \mathcal{B}^\dagger)/\mathcal{D}] \;=\; [\mathcal{B}/\mathcal{D}] + [\mathcal{B}^\dagger \mathcal{D}]. \tag{1.50}$$

3. *For any collection of belief structures $[\mathcal{B}_1^\dagger]$, $[\mathcal{B}_2^\dagger]$, $\ldots$, $[\mathcal{B}_j^\dagger]$, we have*

$$[\mathcal{B}_1^\dagger] + [\mathcal{B}_2^\dagger] + \cdots + [\mathcal{B}_j^\dagger] \;=\; [\mathcal{D}_1^\dagger] + [\mathcal{D}_2^\dagger] + \cdots + [\mathcal{D}_j^\dagger], \tag{1.51}$$

where $[\mathcal{D}_1^\dagger] = [\mathcal{B}_1^\dagger]$, and for $i \geq 2$, $[\mathcal{D}_i^\dagger] = [\mathcal{B}_i^\dagger/(\mathcal{B}_1^\dagger + \cdots + \mathcal{B}_{i-1}^\dagger)]$. The spaces $[\mathcal{D}_1^\dagger]$, $\ldots$, $[\mathcal{D}_j^\dagger]$ are mutually orthogonal.

4. *For any $[\mathcal{B}]$ and any $[\mathcal{D}_1^\dagger]$ orthogonal to $[\mathcal{D}_2^\dagger]$, we have*

$$[\mathcal{B}/(\mathcal{D}_1^\dagger + \mathcal{D}_2^\dagger)] \;=\; [[\mathcal{B}/\mathcal{D}_1^\dagger]/\mathcal{D}_2^\dagger] \;=\; [[\mathcal{B}/\mathcal{D}_2^\dagger]/\mathcal{D}_1^\dagger]. \tag{1.52}$$

5. *For any* $[\mathcal{B}]$, $[\mathcal{D}_1^\dagger]$, *and* $[\mathcal{D}_2^\dagger]$, *we have*

$$[\mathcal{B}/(\mathcal{D}_1^\dagger + \mathcal{D}_2^\dagger)] = [[\mathcal{B}/\mathcal{D}_1^\dagger]/[\mathcal{D}_2^\dagger/\mathcal{D}_1^\dagger]]. \tag{1.53}$$

6. *For any* $[\mathcal{D}_1^\dagger]$, $[\mathcal{D}_2^\dagger]$, *we have*

$$E_{(\mathcal{D}_1^\dagger + \mathcal{D}_2^\dagger)} = E_{\mathcal{D}_1^\dagger} + E_{[\mathcal{D}_2^\dagger/\mathcal{D}_1^\dagger]}. \tag{1.54}$$

7. *For any* $[\mathcal{D}_1^\dagger]$, $[\mathcal{D}_2^\dagger]$ *and any* $\mathcal{A}$, $\mathcal{A}^*$ *in* $[\mathcal{B}]$ *we have*

$$\begin{aligned}(E_{(\mathcal{D}_1^\dagger + \mathcal{D}_2^\dagger)}(\mathcal{A}), E_{(\mathcal{D}_1^\dagger + \mathcal{D}_2^\dagger)}(\mathcal{A}^*)) &= \\ (E_{\mathcal{D}_1^\dagger}(\mathcal{A}), E_{\mathcal{D}_1^\dagger}(\mathcal{A}^*)) + (E_{[\mathcal{D}_2^\dagger/\mathcal{D}_1^\dagger]}(\mathcal{A}), E_{[\mathcal{D}_2^\dagger/\mathcal{D}_1^\dagger]}(\mathcal{A}^*)).\end{aligned} \tag{1.55}$$

In particular, for any $[\mathcal{D}]$, *we can decompose the inner product over* $[\mathcal{B}]$ *as*

$$(\mathcal{A}, \mathcal{A}^*) = ((E_{\mathcal{D}}(\mathcal{A}), E_{\mathcal{D}}(\mathcal{A}^*)) + (E_{[\mathcal{B}/\mathcal{D}]}(\mathcal{A}), E_{[\mathcal{B}/\mathcal{D}]}(\mathcal{A}^*)). \tag{1.56}$$

If both $\mathcal{B}$ and $\mathcal{D}$ are finite, then we may implement the theory through the use of matrices. Consider adjusting the vector of quantities $\mathcal{B} = [\mathcal{B}_1 \dots \mathcal{B}_p]^T$, by another vector of quantities $\mathcal{D} = [\mathcal{D}_1 \dots \mathcal{D}_q]^T$. We reproduce the first two statements of Lemma 1 of Goldstein & Wooff (1998).

Lemma 1 *The vector of adjusted expectations* $E_{\mathcal{D}}(\mathcal{B})$, *and the adjusted variance matrix* $Var_{\mathcal{D}}(\mathcal{B})$ *are computed as*

$$E_{\mathcal{D}}(\mathcal{B}) = E(\mathcal{B}) + Cov(\mathcal{B}, \mathcal{D})\{Var(\mathcal{D})\}^\dagger(\mathcal{D} - E(\mathcal{D})); \tag{1.57}$$

$$Var_{\mathcal{D}}(\mathcal{B}) = Var(\mathcal{B}) - Cov(\mathcal{B}, \mathcal{D})\{Var(\mathcal{D})\}^\dagger Cov(\mathcal{D}, \mathcal{B}). \tag{1.58}$$

$A^\dagger$ represents the Moore-Penrose generalised inverse of A, see Penrose (1955). Equations (1.57) and (1.58) may be considered as the generalisations of equations (1.3) and (1.11) of Mouchart & Simar (1980).

1.7.3 Interpreting the adjustment

Suppose that the examiner receives the scores on a single exam paper, and wishes to examine the effect of those scores on his prior beliefs. Table 1.1 shows the effect of this adjustment. There are a number of points to make about this adjustment. The combinations $\mathcal{M}(X_+)$ and $\mathcal{M}(X_-)$ (see equations (1.34) and (1.35)) are mutually uncorrelated. The quantities thus form an orthogonal basis for $[\mathcal{M}(\mathcal{C})]$ and any

Component	Resolution (sample size 1)
$\mathcal{M}(X_+)$	$\lambda_+ = \frac{1}{7}$
$\mathcal{M}(X_-)$	$\lambda_- = \frac{3}{5}$
$l_1\mathcal{M}(X_1) + l_2\mathcal{M}(X_2)$	$\frac{(l_1+l_2)^2}{2(l_1^2+l_2^2)}\lambda_+ + \frac{(l_1-l_2)^2}{2(l_1^2+l_2^2)}\lambda_-$

Table 1.1: Resolutions for quantities in $[\mathcal{M}(\mathcal{C})]$, having seen the scores on a single exam paper

Component	Resolution (sample size n)
$\mathcal{M}(X_+)$	$\lambda_{(n)+} = \frac{n\lambda_+}{(n-1)\lambda_+ +1} = \frac{n}{n+6}$
$\mathcal{M}(X_-)$	$\lambda_{(n)-} = \frac{n\lambda_-}{(n-1)\lambda_- +1} = \frac{3n}{3n+2}$
$l_1\mathcal{M}(X_1) + l_2\mathcal{M}(X_2)$	$\frac{(l_1+l_2)^2}{2(l_1^2+l_2^2)}\lambda_{(n)+} + \frac{(l_1-l_2)^2}{2(l_1^2+l_2^2)}\lambda_{(n)-}$

Table 1.2: Resolutions for the collection $\mathcal{M}(\mathcal{C})$, adjusting by $\mathcal{C}(n)$

element $l_1\mathcal{M}(X_1)+l_2\mathcal{M}(X_2) \in [\mathcal{M}(\mathcal{C})]$ may be expressed as a linear combination of the two stated quantities. Moreover, we can show that for this particular orthogonal basis that the resolution of $l_1\mathcal{M}(X_1)+l_2\mathcal{M}(X_2)$ may be found as a weighted average of the resolutions of these two quantities. Thus, we expect to learn most about quantities that are proportional to $\mathcal{M}(X_-)$ and least about quantities proportional to $\mathcal{M}(X_+)$. This is very useful for the examiner. $\mathcal{M}(X_-)$ summarises the differences between difficulty in the questions, something that he would like to pay attention to, and $\mathcal{M}(X_+)$ is the average score received on the paper.

Now consider the effect of the adjustment when the examiner is going to receive a sample of n scripts. We'll consider the same quantities. Table 1.2 summarises the adjustment. There are a number of general points to make again. Notice once more that the quantities $\mathcal{M}(X_+)$ and $\mathcal{M}(X_-)$ form an orthogonal basis for $[\mathcal{M}(\mathcal{C})]$ that, coupled with the corresponding resolutions, completely summarises the adjustment. As the resolution for any quantity in $[\mathcal{M}(\mathcal{C})]$ can be expressed as a weighted average for these quantities, we can easily see that we expect to learn most about quantities proportional to $\mathcal{M}(X_-)$, the differences between the questions and least in the direction of the mean score on the exam, $\mathcal{M}(X_+)$. Quantities in $[\mathcal{M}(\mathcal{C})]$ that are highly correlated with the $\mathcal{M}(X_-)$ are expected to be learnt most about. Notice also the straightforward modification of the resolutions belonging to the $\mathcal{M}(X_+)$ and $\mathcal{M}(X_-)$. It is thus straightforward for the examiner to assess the impact of taking any sample size he pleases. Moreover, it is easy to observe the effect for any quantity, for any sample size, by the third quantity in Table 1.2 in which the only dependence on n is through $\lambda_{(n)+}$ and $\lambda_{(n)-}$. Thus, the examiner may use

Quantity	Sample size required for variance reduction of		
	50 percent	90 percent	95 percent
$\mathcal{M}(X_+)$	6	54	114
$\mathcal{M}(X_-)$	1	6	13

Table 1.3: Sample sizes required to achieve the given variance reductions for $\mathcal{M}(X_+)$ and $\mathcal{M}(X_-)$

the information, in combination with the costs associated with the sampling process balanced against the benefits for the variance reduction, to establish optimal sample sizes under various criteria.

Suppose, for example, that the examiner would like to know the sample size required to achieve a proportionate variance reduction of 50 percent, 90 percent and 95 percent for the two quantities of interest. These figures are easily calculable from the resolutions and are shown in Table 1.3.

Thus, the examiner can see that he quickly expects to reduce the variation in his uncertainty about whether the questions are of roughly similar standard, indeed by a sample of size 13, he will have achieved a proportionate variance reduction of over 95 percent. However, he must take a sample 9 times bigger to achieve a similar variance reduction for the total on the paper. Note that a sample size of 114 will guarantee that all the quantities in $[\mathcal{M}(\mathcal{C})]$ have a proportionate variance reduction of at least 95 percent. Notice also how the examiner may use this information to his advantage. He may feel that it is necessary to assess whether the questions are of roughly the same level of difficulty, and a small sample size allows him to do this. The task of aligning the questions through mark scheme adaption is likely to be complicated, and so the examiner may be relieved that he is able to investigate this quickly. He may further decide that having done this, knowledge about the overall total is not that important. He may decide to simply adjust the overall pass rate once all the papers are marked, especially when faced with the comparably large sample he would have to take to reduce much of his uncertainty about the total mark.

We now show that the observations made above are not specific to this example, but in fact hold for general adjustments of second-order exchangeable beliefs.

1.8 Resolution transforms

In a typical analysis, we observe a data collection $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots\}$ and wish to evaluate the effect of the collection $[\mathcal{D}]$ upon the expectations and variances of a collection of beliefs $[\mathcal{B}]$ where $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \dots\}$. The resolution transform for each $\mathcal{A} \in [\mathcal{B}]$ is defined to be

$$T_{\mathcal{D}}(\mathcal{A}) \;=\; E_{\mathcal{B}}\{E_{\mathcal{D}}(\mathcal{A})\}. \tag{1.59}$$

Thus, for each $\mathcal{A} \in [\mathcal{B}]$, $T_{\mathcal{D}}(\mathcal{A})$ is the point in $[\mathcal{B}]$ which is closest to $E_{\mathcal{D}}(\mathcal{A})$. Goldstein (1981) shows that $T_{\mathcal{D}}$ is a bounded, self-adjoint operator on $[\mathcal{B}]$, satisfying for each $\mathcal{A} \in [\mathcal{B}]$,

$$Var_{\mathcal{D}}(\mathcal{A}) \;=\; Var(\mathcal{A}) - Cov(\mathcal{A}, T_{\mathcal{D}}(\mathcal{A})). \tag{1.60}$$

Hence, substituting equation (1.60) into equation (1.48), we have that

$$R_{\mathcal{D}}(\mathcal{A}) \;=\; \frac{Cov(\mathcal{A}, T_{\mathcal{D}}(\mathcal{A}))}{Var(\mathcal{A})}. \tag{1.61}$$

If either of the collections $\mathcal{B}$ or $\mathcal{D}$ are finite dimensional, then Goldstein (1981) showed that we may extract for $T_{\mathcal{D}}$ a set of eigenvectors, $Z = \{Z_1, Z_2, \dots\}$, which form a basis for $[\mathcal{B}]$ and corresponding ordered eigenvalues, $1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$. If both the collections $\mathcal{B}$ and $\mathcal{D}$ are infinite then the eigenstructure of $T_{\mathcal{D}}$ may be more complicated unless a certain compactness condition holds. Details may be found in Goldstein (1981).

Definition 9 *The eigenvalue λ_j is termed the jth canonical resolution, and the corresponding eigenvector Z_j is termed the jth canonical direction.*

The canonical resolutions and directions satisfy the following properties. The Z_j have expectation 0, are mutually uncorrelated, and are scaled to have prior variance 1. As Z forms a basis for $[\mathcal{B}]$, each $\mathcal{A} \in [\mathcal{B}]$ may be expressed as $\mathcal{A} = \sum_{j=1}^{r} Cov(\mathcal{A}, Z_j)Z_j$. In addition, however, we may use equations (1.60) and (1.61) to express the adjusted variance and resolution for each $\mathcal{A} \in [\mathcal{B}]$ as a linear combination of the adjusted variances and resolutions, respectively, of the Z_i as follows:

$$Var_{\mathcal{D}}(\mathcal{A}) \;=\; \sum_{j=1}^{r}(1 - \lambda_j)Cov(\mathcal{A}, Z_j)^2; \tag{1.62}$$

$$R_{\mathcal{D}}(\mathcal{A}) \;=\; \frac{\sum_{j=1}^{r} \lambda_j Cov(\mathcal{A}, Z_j)^2}{Var(\mathcal{A})}. \tag{1.63}$$

Hence, constrained by being uncorrelated with $(Z_1, \ldots, Z_j)$, Z_{j+1} is the element of $[\mathcal{B}]$ maximising the resolution. Thus, we expect to learn most about elements of $[\mathcal{B}]$ having strong correlations with the directions with large resolutions. The canonical directions thus identify the types of information that we expect to gain by sampling, with the quantification of how much we learn in each direction being provided by the canonical resolutions.

1.8.1 Bayes linear sufficiency

In a full probabilistic setting, data d is uninformative about the parameters of interest b if d is independent of b. In which case we have $p(b|d) = p(b)$ where $p(\cdot)$ is the probability density function. Sufficiency occurs when part of the data is uninformative for our learning. Suppose we divide our data into $d = (d_1, d_2)$ and suppose that $p(b|d_1, d_2) = p(b|d_1)$. Then, if d_1 is known, observing d_2 provides no inference about b: it is sufficient to observe d_1. Further details of this general probabilistic notion of sufficiency may be found in O'Hagan (1994; p68-70) and Kendall & Stuart (1979; p22-24).

Bayes linear sufficiency was first developed in Goldstein (1986b), before being formalised in Goldstein & O'Hagan (1996). Bayes linear sufficiency may be seen as the appropriate unbracing of the general probabilistic notion of sufficiency for comparing estimators over linear systems where only a partial prior specification has been made. Suppose we observe a data collection $\mathcal{D} = \{\mathcal{D}_1, \ldots, \mathcal{D}_r, \mathcal{D}_{r+1}, \ldots \mathcal{D}_s\}$ and we wish to use $[\mathcal{D}]$ to learn linearly about a collection of beliefs $[\mathcal{B}]$ where $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots\}$. Then we would say that the collection $[\mathcal{D}_{suff}]$, where $\mathcal{D}_{suff} = \{\mathcal{D}_1, \ldots, \mathcal{D}_r\}$ was Bayes linear sufficient for $[\mathcal{D}]$ for adjusting the collection $[\mathcal{B}]$ if there was no linear way we could use $[\mathcal{D}]$ to learn more about $[\mathcal{B}]$ if we have already adjusted by $[\mathcal{D}_{suff}]$. Precise definitions may be found in Goldstein & O'Hagan (1996).

Bayes linear sufficiency for exchangeable collections

Suppose that we are about to observe a sample of n exchangeable collections, which we label for convenience, $\mathcal{C}_1, \ldots, \mathcal{C}_n$ and we let $\mathcal{C}(n) = \cup_{i=1}^{n} \mathcal{C}_i$. We then want to use this data to revise our beliefs over the mean collection $\mathcal{M}(\mathcal{C})$ and also over future collections of observables $\mathcal{C}_r$ where $n < r$. We use the following notational shorthands for our adjusted quantities:

$$E_n(X) = E_{\mathcal{C}(n)}(X); \tag{1.64}$$

$$Var_n(X) = Var_{\mathcal{C}(n)}(X); \tag{1.65}$$

$$E_{\mathcal{M}}(X) = E_{\mathcal{M}(\mathcal{C})}(X). \tag{1.66}$$

For each $\mathcal{X} \in [\mathcal{C}]$, denote the average of the first n values by

$$\mathcal{S}_n(\mathcal{X}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i, \tag{1.67}$$

and the collection of average values by $\mathcal{S}_n(\mathcal{C}) = \{\mathcal{S}_n(X_1), \mathcal{S}_n(X_2), \ldots\}$. Goldstein & Wooff (1998) show that the collection of sample averages is Bayes linear sufficient for the full sample for adjusting expectations about both the population mean quantities and future observations from the population. They derive the following theorem:

Theorem 4 *If $\mathcal{C}$ is exchangeable over individuals then $\mathcal{S}_n(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(n)$ both for adjusting the collection $\mathcal{M}(\mathcal{C})$ and also for adjusting any collection $\mathcal{C}_r$ $(r > n)$.*

Thus, in terms of the examination problem, the examiner does not need the individual marks from each paper if he only wants to learn about the $[\mathcal{M}(\mathcal{C})]$. It is sufficient for him to know the average score on each question.

1.8.2 Resolution transforms for infinite exchangeable collections

Goldstein & Wooff (1998) discuss the representation theorem for infinite second-order exchangeable beliefs. They show that the resolution transform for the underlying population structure induced by a second-order exchangeable sample has, no matter what the sample size, essentially the same form. Letting $T_n(\cdot) = E_{\mathcal{M}}\{E_n(\cdot)\}$ denote the resolution transform for the mean collection $\mathcal{M}(\mathcal{C})$, based on n observations, $\mathcal{C}(n)$, they derive the following theorem:

Theorem 5 *For infinite exchangeable collections, the eigenvectors of T_n are the same for each n. Further, if eigenvector Y has eigenvalue λ for T_1, then the corresponding eigenvalue $\lambda_{(n)}$ for Y as an eigenvector of T_n is*

$$\lambda_{(n)} = \frac{n\lambda}{(n-1)\lambda + 1}. \tag{1.68}$$

The canonical directions Y_i for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(n)$ thus remain the same for all sample sizes and are termed the canonical directions induced by exchangeability. As Goldstein & Wooff (1998) remark, this is qualitatively important as the underlying features of the adjustment remain the same no matter what the sample size, whilst exploitation of (1.68) allows the simplification of design problems where the sample size required to achieve specified variance reduction over elements of the $[\mathcal{M}(\mathcal{C})]$ is to be chosen. Goldstein & Wooff (1998) derive the following corollary to the theorem:

Corollary 1 *Suppose that Y is an eigenvector of T_1 with eigenvalue $\lambda > 0$. Then the sample size n required to achieve a proportionate variance reduction of κ for Y $(0 < \kappa < 1)$, that is so that $Var_n(W) \leq (1-\kappa)Var(W)$, is*

$$n \;\geq\; \frac{\kappa}{(1-\kappa)} \times \frac{(1-\lambda)}{\lambda}. \tag{1.69}$$

Further, if the minimal eigenvalue of T_1 is $\lambda_{min} > 0$, then a sample size of $\{\kappa/(1-\kappa)\}\{(1-\lambda_{min})/\lambda_{min}\}$, rounded up, is the minimum sample which is sufficient to achieve a proportionate variance reduction of κ for every element of $[\mathcal{M}(\mathcal{C})]$.

Think back to the examination example. The appropriately scaled $\mathcal{M}(X_+)$ and $\mathcal{M}(X_-)$ are the canonical directions induced by exchangeability, with $\lambda_{(n)+}$ and $\lambda_{(n)-}$ following from Theorem 5. By considering Table 1.3, an application of Corollary 1 shows us that sample sizes of 6, 54 and 114 are the minimum sample sizes that will achieve a proportionate variance reduction of 50 percent, 90 percent and 95 percent respectively, for every element of $[\mathcal{M}(\mathcal{C})]$ as $\mathcal{M}(X_+)$ is the smallest canonical direction.

1.9 Predictive adjustment

Goldstein & Wooff (1998) also discuss the effect of the observation of a sample of n individuals for predicting the values for a further collection of r individuals who are second-order exchangeable with those in the sample. Thus, if we let $\mathcal{C}(n;r) = \cup_{j=n+1}^{n+r}\mathcal{C}_j$ denote the collection of future individuals, then we are interested in the adjustment of $\mathcal{C}(n;r)$ by $\mathcal{C}(n)$. For $Y \in [\mathcal{C}]$, we let

$$\mathcal{S}_{(n;r)}(Y) \;=\; \frac{1}{r}\sum_{j=n+1}^{n+r} Y_j, \tag{1.70}$$

and let $\mathcal{S}_{(n;r)}(\mathcal{C}) = \{\mathcal{S}_{(n;r)}(X_1), \mathcal{S}_{(n;r)}(X_2), \dots\}$. The predictive resolution transform, given a sample of size n, assessed for the r further individuals $\mathcal{C}(n;r)$ is denoted as $T_{(n;r)}(\cdot) = E_{(n;r)}\{E_n(\cdot)\}$ where $E_{(n;r)}(X) = E_{\mathcal{C}(n;r)}(X)$. Goldstein & Wooff (1998) show that the canonical resolutions corresponding to non-zero canonical resolutions of the adjustment of $[\mathcal{C}(n;r)]$ by $\mathcal{C}(n)$ share, up to a scale factor to ensure a prior variance of one, the same co-ordinate representation as those of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(n)$. We repeat Theorem 3 of their work.

Theorem 6 *Suppose that Y_i is an eigenvector of T_s, for each s, with eigenvalues $\lambda_{(r)i}$, $\lambda_{(n)i}$ for T_r, T_n respectively. Thus, $\mathcal{M}(Y_i^*) = \lambda_{(1)i}Y_i$. Then $\mathcal{S}_{(n;r)}(Y_i^*)$ is an eigenvector of $T_{(n;r)}$ with eigenvalue*

$$\lambda_{(n;r)i} = \lambda_{(n)i}\lambda_{(r)i}. \tag{1.71}$$

Further, all eigenvectors of $T_{(n;r)}$ are of this form.

Hence, the crucial point is, as Goldstein & Wooff (1998) explain, that predictive adjustment and adjustment over the population structure share the same qualitative features. Thus, there are similar implications for design and interpretation in the predictive case as in the population structure adjustment.

1.10 Implementations of the theory

In this thesis, we shall largely concentrate upon adjustments where the number of variables we wish to measure on the individuals is finite. Thus, we shall consider that for each individual, we shall make a total of v_0 measurements. For example, we may, analogously, write $\mathcal{C}_i$ as the $v_0 \times 1$ vector, $\mathcal{C}_i = [X_{1i} \dots X_{v_0 i}]^T$ and $\mathcal{C}(n)$ as the $nv_0 \times 1$ vector, $\mathcal{C}(n) = [\mathcal{C}_1^T \dots \mathcal{C}_n^T]^T$.

In this case, we may implement the theory through the use of matrices. Consider adjusting the vector of quantities $\mathcal{B} = [\mathcal{B}_1 \dots \mathcal{B}_p]^T$, by another vector of quantities $\mathcal{D} = [\mathcal{D}_1 \dots \mathcal{D}_q]^T$. In this thesis, we wish to consider cases where the variance matrices under consideration are strictly positive definite and hence invertible. If they are not invertible, we obtain the corresponding results to those that we develop that by considering the adjustment over the linear span of the columns of the matrices that we need to invert. For our needs, we rewrite Lemma 1 of Goldstein & Wooff (1998), see also Lemma 1 in this thesis, as the following lemma.

Lemma 2 *The vector of adjusted expectations $E_{\mathcal{D}}(\mathcal{B})$, the adjusted variance matrix*

$Var_{\mathcal{D}}(\mathcal{B})$, and the resolution transform $T_{\mathcal{D}}$ are computed as

$$\begin{aligned} E_{\mathcal{D}}(\mathcal{B}) &= E(\mathcal{B}) + Cov(\mathcal{B},\mathcal{D})\{Var(\mathcal{D})\}^{-1}(\mathcal{D} - E(\mathcal{D})); & (1.72) \\ Var_{\mathcal{D}}(\mathcal{B}) &= Var(\mathcal{B}) - Cov(\mathcal{B},\mathcal{D})\{Var(\mathcal{D})\}^{-1}Cov(\mathcal{D},\mathcal{B}); & (1.73) \\ T_{\mathcal{D}} &= \{Var(\mathcal{B})\}^{-1}Cov(\mathcal{B},\mathcal{D})\{Var(\mathcal{D})\}^{-1}Cov(\mathcal{D},\mathcal{B}). & (1.74) \end{aligned}$$

In this full invertibility case, equations (1.57) and (1.58) reduce to equations (1.72) and (1.73), which are, respectively, equations (1.3) and (1.11) of Mouchart & Simar (1980).

In the case of exchangeable sequences, we collect the $\{c_{vw}\}$ (see equation (1.22)) into the $v_0 \times v_0$ matrix C with (v,w)th entry $(C)_{vw} = c_{vw}$ and the $\{e_{vw} = d_{vw} - c_{vw}\}$ (see equation (1.24)) into the $v_0 \times v_0$ matrix E with (v,w)th entry $(E)_{vw} = c_{vw}$. We may then write

$$\begin{aligned} Var(\mathcal{M}(\mathcal{C})) &= C; & (1.75) \\ Cov(\mathcal{R}_i(\mathcal{C}),\mathcal{R}_j(\mathcal{C})) &= \begin{cases} E & \text{if } i = j; \\ 0 & \text{otherwise.} \end{cases} & (1.76) \end{aligned}$$

As Goldstein & Wooff (1998) show, T_n, the resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(n)$ may be expressed as

$$T_n = \{C + (1/n)E\}^{-1}C. \qquad (1.77)$$

The eigenvectors of T_n are the same for each n, with simple modifications for the eigenvalues. They may be found by solving the generalised eigenvalue problem

$$CY = (C + E)Y\Lambda, \qquad (1.78)$$

which is the form of T_1. This generalised eigenvalue problem will play a significant role throughout this thesis.

1.11 Progressions

In this chapter, we have motivated and explored the use of Bayes linear methods through the example. However, even within this seemingly straightforward example, there are questions that need answering. What happens if all the scripts are not marked by the same marker? Is our analysis completely wasted, or is there scope for continuation through partial exchangeability? Indeed, what happens if there is more than one marker of the scripts? The examiner assumes that the candidates are selected from an infinite population. This is clearly an approximation. Is there any way we can remove this, by tackling the problem through finite second-order exchangeability? If so, can we establish the effect of the examiner's assumption of infinite second-order exchangeability? In the remainder of this thesis, we shall develop methods which enable us to answer these questions, and to develop a methodology that is applicable to a wide range of problems in experimental design.

Chapter 2

Simplifying Complex Designs: Bayes Linear Experimental Design for Grouped Multivariate Exchangeable Systems

SUMMARY

Having summarised the work on the adjustment of second-order exchangeable sequences in Chapter 1, we consider elaborating our modelling through the use of partial exchangeability. In Subsection 2.1.1 we review adoptions of partial exchangeability within the fully specified Bayesian framework before reviewing, in Subsection 2.1.2, the type of partial exchangeability we shall use in the Bayes linear analysis we consider in this thesis, namely co-exchangeability. We are concerned with observing collections of individuals in order to learn about the underlying population structure and the prediction of future individuals and in Subsections 2.1.3 and 2.1.4 we show that the sample means are Bayes linear sufficient to perform these adjustments. For our work, we find it more illuminating to adopt a model, described in Section 2.3, that, whilst being co-exchangeable,is not the most general co-exchangeable model. In Section 2.4 we consider the adjustment of the mean components for this model and show that the canonical structure for the adjustment may be derived by solving a series of subproblems; an underlying canonical variable problem, as de-

scribed in Subsection 2.4.1, and an underlying canonical group problem, as described in Subsection 2.4.3. Both of these subproblems have an interpretable form in terms of adjustment upon subspaces of the full problem. In Subsection 2.4.5, we show that the analysis of these subproblems completely determine the analysis for the full adjustment. In Sections 2.6 and 2.7, we consider the prediction of future individuals, firstly when the individuals belong to groups where we have already observed individuals, and secondly where the future individuals belong to previously unobserved groups. Once more, we may find a series of subproblems for which the solution of which enable the solution of the full adjustment, as explained in Subsections 2.6.2 and 2.7.2. The subproblems may again be viewed as underlying canonical variable and group problems, and are discussed in Subsections 2.6.1 and 2.7.1. In both cases, the underlying canonical variable problem is that given in Subsection 2.4.1, and the underlying canonical group problems may be interpreted in an analogous manner to that for the adjustment of the mean components. Throughout the chapter, we illustrate the theory through the simple examiner example.

2.1 Widening the scheme to include co-exchangeability

In this thesis we have argued that a subjectivistic approach to the revision of belief should be adopted and that we are concerned with attempting to accurately model the opinion of the person concerned and investigate how these beliefs alter with data. Indeed, this is why we observe data, for to quote Goldstein (1992) "Why is a data analysis of interest? Because it can influence someone's beliefs! Our beliefs and how they change are what is fundamental - data analysis is simply a means, not an end in itself." Thus, our goal is realistic and tractable belief modelling and how these beliefs are influenced by the data and our natural progression is to investigate whether we are modelling as realistically as we can. Previously, we have reviewed the analysis of infinite second-order exchangeable sequences. This work was motivated and illustrated by a scenario where the examiner received the scores on the exam papers sat by a number of individuals. In reality however, the examiner may have made simplifications to his modelling. He may have followed the lines that Goldstein & Wooff (1994) suggest are often made , for

> " ... in a complicated analysis, we might have to make thousands of numerical specifications. Typically, we will make such specifications by imposing pragmatic simplifications, e.g. treating 'almost exchangeable' units as exchangeable, 'almost uncorrelated' quantities as uncorrelated."

For example, the examiner may know that the papers are marked by different markers and the case of second-order exchangeability across all the scripts is an idealised simplification. Given more time, he may be willing and able to think about explicitly incorporating the different markers into his beliefs. Alternatively, he may have overlooked the different markers when he made his belief specification, but has, on reflection, decided to incorporate them into his beliefs. Exchangeability should always be viewed as a simplification of our beliefs. What we need is guidance about how and where we should elaborate our modelling. De Finetti (1959), translated as Chapter 9 of de Finetti (1972), perhaps best sums up the situation:

> " ... there is no harm in beginning with an oversimplified formulation. What is important is not to stop there, not to consider simplicity as a mark of perfection, but as a useful characteristic to be exploited for the first step ... we must widen the scheme and consider partial exchangeability. But we shall do it gradually ... passing directly to the most general case would be to renounce all possibility of illuminating the varied aspects of the question that merit interest."

2.1.1 Partial Exchangeability

Return to the examiner example. He now wishes to explicitly build into his model the fact that the papers are marked by different markers. He might reason that each marker will mark consistently and not suffer from factors such as fatigue or disheartenment. He may then consider that the sequence of individuals marked by the same marker is exchangeable, but be less certain that this judgement is appropriate for the relationships between different markers. He might judge, for example, that one marker is more thorough than another, or that the different levels of marker experience may influence their marking. Another example is provided by

Bernardo & Smith (1994). Suppose that a drug is administered at a number of different dose levels to both male and female patients across a broad age spectrum. One is unlikely to judge that the whole sequence is exchangeable, but within each combination of dose level, sex and age bracket, we might deem it appropriate to judge the patients as exchangeable. It may be that these distinctions are slight, for as Bruno (1964), translated as Chapter 10 of de Finetti (1972) states, "in our example of the effectiveness of a new drug, we may be reluctant to admit a difference in the drug according to sex. Only after a long series of trials the difference may appear significant and could then be translated into different probabilities for the two groups".

We could think of many diversions away from overall judgements of exchangeability to those of partial exchangeability; further examples may be found, for example, in Lad (1996).

A precise definition of partial exchangeability is hard, for as Bernardo & Smith (1994) write:

> "Clearly, there are many possible forms of departure from overall judgements of exchangeability to those of partial exchangeability and so a formal definition of the term does not seem appropriate. In general, it simply signifies that there may be additional "labels" on the random quantities with exchangeable judgements made separately for each group of random quantities having the same additional labels."

Much work has been carried out on various forms of partial exchangeability, involving full exchangeability, and the development of representation theorems from these judgements. The earliest development may be tracked to de Finetti (1938), whilst further extensions may be found in Freedman (1962, 1963) and Diaconis & Freedman (1980a). Diaconis (1988) provides a good review, whilst Aldous (1985) provides a brief introduction. For a textbook grouping of models via partial exchangeability see Bernardo & Smith (1994), Section 4.6.

The judgement of partial exchangeability is a useful extension in fulfilling our aims of models reflecting our actual beliefs; the extensions mentioned above are considered from a full exchangeability perspective and thus, our concerns which attach themselves to full exchangeability are not removed by the adoption of partial exchangeability in these cases. We wish to consider partial exchangeability in the light of second-order exchangeability. One extension, that we find useful for this thesis is that of co-exchangeability which we now describe.

2.1.2 Co-exchangeability

Goldstein (1986a) introduces the concept of co-exchangeable belief structures and we give an overview of that development in this subsection. Assume $\mathcal{C}_1 = \{X_{11}, X_{12}, \dots\}$ is a collection of measurements made on a series of individuals, so that $\mathcal{C}_{1i} = \{X_{11i}, X_{12i}, \dots\}$ are the measurements for the ith individual in the collection. Let $\mathcal{C}_1^*$ be the union of all of the elements in all of the collections $\mathcal{C}_{1i}$, where i ranges over some indexing set of individuals, and suppose that the collection of measurements $\mathcal{C}_1$ is second-order exchangeable over the full collection $\mathcal{C}_1^*$. Now imagine that $\mathcal{C}_2 = \{X_{21}, X_{22}, \dots\}$ is another series of measurements made on a different collection of individuals; the measurements need not be the same as those made in the previous collection and let $\mathcal{C}_{2j} = \{X_{21j}, X_{22j}, \dots\}$ be the measurements for the jth individual in this collection.

Definition 10 *We say that $\mathcal{C}_{2j}$ is exchangeable with the belief structure $\mathcal{C}_1^*$ if*

$$Cov(X_{1vi}, X_{2wj}) = \hat{c}_{vw} \; \forall v, w, i. \tag{2.1}$$

Now, form the union of all of the elements in all of the collections $\mathcal{C}_{2j}$ by letting j range over some indexing set of individuals, and call this $\mathcal{C}_2^*$. We shall consider that the collection of measurements $\mathcal{C}_2$ is second-order exchangeable over the full collection $\mathcal{C}_2^*$. Co-exchangeable belief structures are defined as follows:

Definition 11 *Two exchangeable systems, $\mathcal{C}_1^*$ and $\mathcal{C}_2^*$, are termed co-exchangeable if each $\mathcal{C}_{1i}$ is exchangeable with $\mathcal{C}_2^*$ and each $\mathcal{C}_{2j}$ is exchangeable with $\mathcal{C}_1^*$. This is equivalent to the requirement that*

$$Cov(X_{1vi}, X_{2wj}) = \hat{c}_{vw} \; \forall v, w, i, j. \tag{2.2}$$

Thus, two second-order exchangeable collections are termed co-exchangeable if the joint second-order specification is invariant under permutation. The definition can be extended to encompass as many systems as we desire. We term $\mathcal{C}_1^*$, $\mathcal{C}_2^*$, $\mathcal{C}_3^*$, ... a chain of co-exchangeable exchangeable sequences if we specify

$$E(X_{gvi}) = m_{gv} \; \forall g, v, i; \tag{2.3}$$

$$Cov(X_{gvi}, X_{gwi}) = d_{gvw} \; \forall g, v, w, i; \tag{2.4}$$

$$Cov(X_{gvi}, X_{gwj}) = c_{ggvw} \; \forall g, v, w, i \neq j; \tag{2.5}$$

$$Cov(X_{gvi}, X_{hwj}) = c_{ghvw} \; \forall g \neq h, v, w, i, j. \tag{2.6}$$

Hence, $\mathcal{C}_1^*$, $\mathcal{C}_2^*$, $\mathcal{C}_3^*$, ... are a chain of co-exchangeable exchangeable sequences if they are pairwise co-exchangeable. Notice that there need not be any relationship between the pairwise consideration of, say the first and second sequences and the first and third sequences. Indeed, in full generality, we may make a different series of measurements on each sequence. The representation theorem for exchangeable sequences may then be adapted to incorporate co-exchangeability as the following theorem, see Goldstein (1986a), shows:

Theorem 7 *Suppose that $\mathcal{C}_1^*$, $\mathcal{C}_2^*$, $\mathcal{C}_3^*$, ... are a sequence of co-exchangeable infinite exchangeable systems. Then for each system g we introduce the further collections of random quantities $\mathcal{M}(\mathcal{C}_g) = \{\mathcal{M}(X_{g1}), \mathcal{M}(X_{g2}), \dots\}$, and $\mathcal{R}_i(\mathcal{C}_g) = \{\mathcal{R}_i(X_{g1}), \mathcal{R}_i(X_{g2}), \dots\}$, and $\forall g, v, i$ write*

$$X_{gvi} = \mathcal{M}(X_{gv}) + \mathcal{R}_i(X_{gv}), \tag{2.7}$$

where

$$\mathcal{M}(X_{gv}) = \lim_{n_g \to \infty} \frac{1}{n_g} \sum_{i=1}^{n_g} X_{gvi}, \tag{2.8}$$

with the limit in mean square. The collections $\mathcal{M}(\mathcal{C}_g)$ and $\mathcal{R}_i(\mathcal{C}_g)$ satisfy the following relationships

$$\begin{aligned}
E(\mathcal{M}(X_{gv})) &= m_{gv} \ \forall g, v; && (2.9)\\
E(\mathcal{R}_i(X_{gv})) &= 0 \ \forall g, v, i; && (2.10)\\
Cov(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hw})) &= c_{ghvw} \ \forall g, h, v, w; && (2.11)\\
Cov(\mathcal{M}(X_{gv}), \mathcal{R}_j(X_{hw})) &= 0 \ \forall g, h, v, w, j; && (2.12)\\
Cov(\mathcal{R}_i(X_{gv}), \mathcal{R}_j(X_{hw})) &= \begin{cases} d_{gvw} - c_{ggvw} & \text{if } g = h,\ i = j\ \forall v, w; \\ 0 & \text{otherwise.} \end{cases} && (2.13)
\end{aligned}$$

From equations (2.9) - (2.13) we may deduce the relationships between the collections $\mathcal{M}(\mathcal{C}_g)$, $\mathcal{R}_i(\mathcal{C}_g)$ and $\mathcal{C}_g$. In particular, notice that by adding equations (2.11) and (2.12) we get

$$Cov(\mathcal{M}(X_{gv}), X_{hwj}) = c_{ghvw} \ \forall g, h, v, w. \tag{2.14}$$

2.1.3 Adjusting the mean components

Suppose that we make observations in g_0 systems, observing $n_g > 0$ individuals in the gth system for $g = 1, \dots, g_0$. We wish to use these observations to adjust our beliefs over the mean collection, $\mathcal{M}(\mathcal{C}) = \{\mathcal{M}(\mathcal{C}_1), \dots, \mathcal{M}(\mathcal{C}_{g_0})\}$. Thus in the

gth system, we will observe the n_g exchangeable collections, $\mathcal{C}_{g1}, \ldots, \mathcal{C}_{gn_g}$ where $\mathcal{C}_{gi} = \{X_{g1i}, \ldots, X_{gv_0i}\}$. The complete sample in the gth system is gathered together as $\mathcal{C}_g(n_g) = \cup_{i=1}^{n_g} \mathcal{C}_{gi}$. The total sample is $\mathcal{C}(N) = \cup_{g=1}^{g_0} \mathcal{C}_g(n_g)$. For $\mathcal{X}_g \in \langle \mathcal{C}_g \rangle$, let

$$\mathcal{S}_{n_g}(\mathcal{X}_g) \quad = \quad \frac{1}{n_g} \sum_{i=1}^{n_g} \mathcal{X}_{gi}, \tag{2.15}$$

and let $\mathcal{S}_{n_g}(\mathcal{C}_g) = \{\mathcal{S}_{n_g}(X_{g1}), \ldots, \mathcal{S}_{n_g}(X_{gv_0})\}$ be the collection of sample averages for the gth system. The complete collection of sample averages are denoted by $\mathcal{S}_N(\mathcal{C}) = \{\mathcal{S}_{n_1}(\mathcal{C}_1), \ldots, \mathcal{S}_{n_{g_0}}(\mathcal{C}_{g_0})\}$. For each $X_{gvi} \in \mathcal{C}(N)$ let

$$\mathcal{T}_i(X_{gv}) \quad = \quad X_{gvi} - \mathcal{S}_{n_g}(X_{gv}), \tag{2.16}$$

so that $\mathcal{T}_i(\mathcal{C}_g(n_g)) = \{\mathcal{T}_i(X_{g1}), \ldots, \mathcal{T}_i(X_{gv_0})\}$ is the collection of residuals of the ith individual in the gth system from the collection of sample means corresponding to that system. The complete collection of sample mean residuals for the gth system is then $\mathcal{T}(\mathcal{C}_g(n_g)) = \cup_{i=1}^{n_g} \mathcal{T}_i(\mathcal{C}_g(n_g))$ and $\mathcal{T}(\mathcal{C}(N)) = \cup_{g=1}^{g_0} \mathcal{T}(\mathcal{C}_g(n_g))$ is the complete collection of sample mean residuals. We have the following lemma.

Lemma 3 *The second-order relationships between the $\mathcal{M}(\mathcal{C}_g)$s, the $\mathcal{S}_{n_g}(\mathcal{C}_g)$s and the $\mathcal{T}_i(\mathcal{C}_g(n_g))$s may be expressed as*

$$\begin{aligned}
Cov(\mathcal{M}(X_{gv}), \mathcal{S}_{n_h}(X_{hw})) &= c_{ghvw} \ \forall g, h, v, w; & (2.17)\\
Cov(\mathcal{S}_{n_g}(X_{gv}), \mathcal{S}_{n_h}(X_{hw})) &= \begin{cases} c_{ggvw} + \frac{1}{n_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h \ \forall v, w; \\ c_{ghvw} & \text{otherwise}; \end{cases} & (2.18)\\
Cov(\mathcal{S}_{n_g}(X_{gv}), \mathcal{T}_j(X_{hw})) &= 0 \ \forall g, h, v, w, j; & (2.19)\\
Cov(\mathcal{M}(X_{gv}), \mathcal{T}_j(X_{hw})) &= 0 \ \forall g, h, v, w, j; & (2.20)\\
Cov(\mathcal{T}_i(X_{gv}), \mathcal{T}_j(X_{hw})) &= \begin{cases} \frac{n_g - 1}{n_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h,\ i = j;\ \forall v, w; \\ -\frac{1}{n_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h,\ i \neq j;\ \forall v, w; \\ 0 & \text{otherwise}. \end{cases} & (2.21)
\end{aligned}$$

Proof - The results follow immediately from the definitions of the $\mathcal{S}_{n_g}(\mathcal{C}_g)$s (see equation (2.15)) and the $\mathcal{T}_i(\mathcal{C}_g(n_g))$s (see equation (2.16)) and the specifications given by equations (2.11) - (2.13). □

The resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{S}_N(\mathcal{C})$ is denoted by $T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$ and the resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$ is denoted by $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]}$. We now show that the collection of sample means is Bayes linear sufficient for the full sample for adjusting the collection of mean components.

Theorem 8 *If $\mathcal{C}_1^*, \ldots, \mathcal{C}_{g_0}^*$ are a chain of co-exchangeable infinite exchangeable systems then $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting the collection $\mathcal{M}(\mathcal{C})$.*

Proof - Note from equation (2.19) we have that $[\mathcal{S}_N(\mathcal{C})] \perp [\mathcal{T}(\mathcal{C}(N))]$ and from equation (2.20) we have that $[\mathcal{M}(\mathcal{C})] \perp [\mathcal{T}(\mathcal{C}(N))]$. This is all we need as we now verify. Note that $[\mathcal{C}(N)] \subseteq [\mathcal{S}_N(\mathcal{C})] + [\mathcal{T}(\mathcal{C}(N))]$. We have

$$\begin{aligned}
[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)] &= [\mathcal{M}(\mathcal{C})/(\mathcal{S}_N(\mathcal{C}) + \mathcal{T}(\mathcal{C}(N)))] && (2.22)\\
&= [[\mathcal{M}(\mathcal{C})/\mathcal{T}(\mathcal{C}(N))]/[\mathcal{S}_N(\mathcal{C})/\mathcal{T}(\mathcal{C}(N))]] && (2.23)\\
&= [\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})], && (2.24)
\end{aligned}$$

where equation (2.22) follows from Property 5 of Theorem 3 and equation (2.23) follows by orthogonality and Property 1 of Theorem 3. □

Corollary 2 *The following equations, for each $\mathcal{M}(\mathcal{X}) \in [\mathcal{M}(\mathcal{C})]$, are necessary and sufficient conditions for the Bayes linear sufficiency of $\mathcal{S}_N(\mathcal{C})$ for $\mathcal{C}(N)$ for adjusting the collection $\mathcal{M}(\mathcal{C})$.*

$$\begin{aligned}
E_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(\mathcal{X})) &= E_N(\mathcal{M}(\mathcal{X})); && (2.25)\\
Var_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(\mathcal{X})) &= Var_N(\mathcal{M}(\mathcal{X})); && (2.26)\\
T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}(\mathcal{M}(\mathcal{X})) &= T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]}(\mathcal{M}(\mathcal{X})). && (2.27)
\end{aligned}$$

Proof - From Property 6 of Theorem 3 we have

$$\begin{aligned}
E_N(\mathcal{M}(\mathcal{X})) &= E_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(\mathcal{X})) + E_{[\mathcal{T}(\mathcal{C}(N))/\mathcal{S}_N(\mathcal{C})]}(\mathcal{M}(\mathcal{X})) && (2.28)\\
&= E_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(\mathcal{X})) + E_{\mathcal{T}(\mathcal{C}(N))}(\mathcal{M}(\mathcal{X})) && (2.29)\\
&= E_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(\mathcal{X})), && (2.30)
\end{aligned}$$

where equation (2.29) follows since $[\mathcal{S}_N(\mathcal{C})] \perp [\mathcal{T}(\mathcal{C}(N))]$ and equation (2.30) follows since $[\mathcal{M}(\mathcal{C})] \perp [\mathcal{T}(\mathcal{C}(N))]$. Then, from Goldstein (1990; p153), we have that

$$T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}(\mathcal{M}(\mathcal{X})) = T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]}(\mathcal{M}(\mathcal{X})). \qquad (2.31)$$

Equation (2.26) thus follows immediately via equation (1.62). □

Thus, if we want to adjust the mean components corresponding to a sequence of co-exchangeable infinite exchangeable systems by observing individuals in each of the systems then we need only the resulting sample means to perform the adjustment.

2.1.4 Predicting future individuals

In addition to using the sample to learn about the mean components, $\mathcal{M}(\mathcal{C})$, we may also want to use the sample for predicting the values of future individuals we may want to observe. We shall consider two scenarios. Firstly that we'd like to predict the values for future individuals in systems where we have already observed some individuals. The second case is where we observe individuals belonging to exchangeable systems that we have not currently observed any individuals from.

In the first instance, consider that we wish to examine the impact of the observation of the sample of n_g individuals in the gth system for $g = 1, \dots, g_0$ for predicting future individuals belonging to some subset of these g_0 systems. Thus, we are interested in predicting in a subset $\{\nu_1, \dots, \nu_{h_0}\}$ of the systems $\{1, \dots, g_0\}$ and in the ν_hth system, we wish to predict the values for the next r_{ν_h} individuals as we have already observed the first n_{ν_h} individuals in this system. Without loss of generality, we assume that the systems are labelled so that we are interested in predicting in the first $h_0 \leq g_0$ systems. Let $\mathcal{C}(n_h; r_h) = \cup_{i=n_h+1}^{n_h+r_h} \mathcal{C}_{hi}$ denote the collection of individuals in the hth system that we would like to predict the values for. We denote the complete collection of individuals we would like to predict the values for as $\mathcal{C}(N; R) = \cup_{h=1}^{h_0} \mathcal{C}(n_h; r_h)$. $\langle \mathcal{C}(N; R) \rangle$ denotes the collection of linear combinations of the elements of $\mathcal{C}(N; R)$ and $[\mathcal{C}(N; R)]$ the inner product space formed from $\langle \mathcal{C}(N; R) \rangle$ in the usual way. We denote the resolution transform for the adjustment of $[\mathcal{C}(N; R)]$ by $\mathcal{C}(N)$ as $T_{[\mathcal{C}(N;R)/\mathcal{C}(N)]}$ and $T_{[\mathcal{C}(N;R)/\mathcal{S}_N(\mathcal{C})]}$ denotes the resolution transform for the adjustment of $[\mathcal{C}(N; R)]$ by $\mathcal{S}_N(\mathcal{C})$.

Theorem 9 *If $\mathcal{C}_1^*, \dots, \mathcal{C}_{g_0}^*$ are a chain of co-exchangeable infinite exchangeable systems then $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting the collection $\mathcal{C}(N; R)$. Equivalently, for any $Y_h \in [\mathcal{C}_h]$ where $h \in \{1, \dots, h_0\}$, $j \in \{n_h + 1, \dots, n_h + r_h\}$ we have*

$$\begin{aligned} E_{\mathcal{C}(N)}(Y_{hj}) &= E_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(Y_h)); & (2.32)\\ Var_{\mathcal{C}(N)}(Y_{hj}) = Var_{\mathcal{S}_N(\mathcal{C})}(Y_{hj}) &= Var_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(Y_h)) + Var(\mathcal{R}_j(Y_h)), & (2.33)\end{aligned}$$

and for all $Y \in [\mathcal{C}(N; R)]$ we have

$$T_{[\mathcal{C}(N;R)/\mathcal{C}(N)]}(Y) = T_{[\mathcal{C}(N;R)/\mathcal{S}_N(\mathcal{C})]}(Y). \quad (2.34)$$

Proof - Notice that for each $X_{hwj} \in \mathcal{C}_{hj}$ where $h \in \{1, \dots, h_0\}$, $j \in \{n_h+1, \dots, n_h+r_h\}$, we have

$$Cov(\mathcal{T}_i(X_{gv}), X_{hwj}) = 0, \quad (2.35)$$

so that $[\mathcal{C}(N;R)] \perp [\mathcal{T}(\mathcal{C}(N))]$ and Bayes linear sufficiency follows in the same way as the proof to Theorem 8. Notice that by adding equations (2.12) and (2.13) we have that if $h \in \{1,\ldots,h_0\}$, $j \in \{n_h+1,\ldots,n_h+r_h\}$, $g \in \{1,\ldots,g_0\}$, $i \in \{1,\ldots,n_g\}$ then for all v,w,

$$Cov(X_{gvi}, \mathcal{R}_j(X_{hw})) \;=\; 0, \tag{2.36}$$

so that $[\mathcal{R}_j(\mathcal{C}_g)] \perp [\mathcal{C}(N)]$ and by Property 1b of Theorem 3,

$$Var_{\mathcal{C}(N)}(\mathcal{R}_j(X_{vw})) \;=\; Var(\mathcal{R}_j(X_{vw})), \tag{2.37}$$

and equation (2.33) follows. □

In the second instance, consider that we wish to examine the impact of the observation of the sample of n_g individuals in the first g_0 exchangeable systems for predicting the first r_h individuals in the hth system, where $h \in \{g_0+1,\ldots,g_0+h_0\}$. Thus, we are interested in predicting in h_0 previously unobserved systems. Let $\mathcal{C}_h(r_h) = \cup_{i=1}^{r_h}\mathcal{C}_{hi}$ denote the collection of individuals in the hth system about whom we are interested in predicting the values for. The complete collection of future individuals of individuals is $\mathcal{C}(R/N) = \cup_{h=g_0+1}^{g_0+h_0}\mathcal{C}_h(r_h)$. We denote the resolution transform for the adjustment of $[\mathcal{C}(R/N)]$ by $\mathcal{C}(N)$ as $T_{[\mathcal{C}(R/N)/\mathcal{C}(N)]}$ and $T_{[\mathcal{C}(R/N)/\mathcal{S}_N(\mathcal{C})]}$ denotes the resolution transform for the adjustment of $[\mathcal{C}(R/N)]$ by $\mathcal{S}_N(\mathcal{C})$.

Theorem 10 *If $\mathcal{C}_1^*, \ldots, \mathcal{C}_{g_0}^*, \mathcal{C}_{g_0+1}^*, \ldots, \mathcal{C}_{g_0+h_0}^*$ are a chain of co-exchangeable infinite exchangeable systems, then $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting the collection $\mathcal{C}(R/N)$. Equivalently, for any $Y_h \in [\mathcal{C}_h]$ where $h \in \{g_0+1,\ldots,g_0+h_0\}$, $j \in \{1,\ldots,r_h\}$ we have*

$$E_{\mathcal{C}(N)}(Y_{gj}) \;=\; E_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(Y_g)); \tag{2.38}$$
$$Var_{\mathcal{C}(N)}(Y_{gj}) \;=\; Var_{\mathcal{S}_N(\mathcal{C})}(Y_{gj}) \;=\; Var_{\mathcal{S}_N(\mathcal{C})}(\mathcal{M}(Y_g)) + Var(\mathcal{R}_j(Y_g)); \tag{2.39}$$

and for all $Y \in [\mathcal{C}(R/N)]$ we have

$$T_{[\mathcal{C}(R/N)/\mathcal{C}(N)]}(Y) \;=\; T_{[\mathcal{C}(R/N)/\mathcal{S}_N(\mathcal{C})]}(Y). \tag{2.40}$$

Proof - Follows in a similar way to Theorem 9 by noting that for each $X_{hwj} \in \mathcal{C}_{hj}$ where $h \in \{g_0+1,\ldots,g_0+h_0\}$, $j \in \{1,\ldots,r_h\}$ we have

$$Cov(\mathcal{T}_i(X_{gv}), X_{hwj}) \;=\; 0. \tag{2.41}$$

□

We may proceed from here; but in order to motivate our future work we return to our running examiner example.

2.2 The examiner revisited

2.2.1 The examiner's considerations

Suppose that the exam papers are to be marked by two markers, marker 1 and marker 2 and that the examiner is aware of who will mark each script. The examiner denotes the measurements, for the ith individual being marked by the gth marker, by $\mathcal{C}_{gi} = \{X_{g1i}, X_{g2i}\}$. Thus, X_{gvi} denotes the mark on the vth question for the ith individual marked by the gth marker. He denotes by $\mathcal{C}_g^*$ the complete collection of marks for all the individuals marked by the gth marker. In our earlier discussion and development, the examiner was interested in learning about the relationships between the questions. Whilst he remains interested in these, his interests have also expanded, for, additionally he has questions about the markers. Chiefly, he is concerned with ensuring that the markers are marking to the same benchmark.

The examiner considers his beliefs about the X_{gvi}. The examiner asserts that if he did not know which marker was responsible for which script, he would view the total sequence $\mathcal{C}^* = \mathcal{C}_1^* \cup \mathcal{C}_2^*$ (of course, without knowledge of the markers, the label g in X_{gvi} would carry no meaning to the examiner) as being second-order exchangeable over the individuals. The second-order specification he is willing and able to make for the relationships between the individuals would be that that he made in Chapter 1, see equations (1.15), (1.16) and (1.17). The examiner now contemplates the effect of knowing the markers upon his beliefs.

Initially, the examiner is unsure exactly how the markers differ; this is one aspect he would like to investigate. For example, all markers have received the same training and instruction on how to mark the current exam paper. The examiner first concentrates upon his beliefs about the relationships between individuals marked by the same marker. He judges that the individuals are second-order exchangeable, and for initial simplicity that the individuals are drawn from a potentially infinite sequence of individuals who could have been marked by the marker under consideration. Having designed the questions with the intention of them being of comparable difficulty and judged them exchangeable, the examiner sees no reason why, when separating his beliefs to include markers, this aspect would be lost and moreover how each marker would effect this. As such, he judges that individuals in the same group share the same specification as that he specified when he elected not to take account of the different markers in his modelling. That is, his specification is the same as that given in equations (1.15), (1.16) and (1.17), namely:

$$E(X_{gvi}) = 0 \ \forall g, v, i; \tag{2.42}$$

$$Cov(X_{gvi}, X_{gwi}) = \begin{cases} 5 & \text{if } v = w, \ \forall g, i; \\ 3.75 & \text{if } v \neq w, \ \forall g, i; \end{cases} \tag{2.43}$$

$$Cov(X_{gvi}, X_{gwj}) = \begin{cases} 1 & \text{if } v = w, \ \forall i \neq j; \\ 0.25 & \text{if } v \neq w, \ \forall i \neq j. \end{cases} \tag{2.44}$$

Once more, the examiner has standardised the random quantities by subtracting the prior mean. For each group g, $\mathcal{C}_g$ is second-order exchangeable over the individuals and the examiner may introduce the further collections of random quantities $\mathcal{M}(\mathcal{C}_g) = \{\mathcal{M}(X_{g1}), \mathcal{M}(X_{g2})\}$ and $\mathcal{R}_i(\mathcal{C}_g) = \{\mathcal{R}_i(X_{g1}), \mathcal{R}_i(X_{g2})\}$ and write $X_{gvi} = \mathcal{M}(X_{gv}) + \mathcal{R}_i(X_{gv})$. The examiner now needs solely to consider the relationships across groups. The examiner judges that the exchangeable systems $\mathcal{C}_1$, $\mathcal{C}_2$ are co-exchangeable. That is, he assesses that $Cov(X_{1vi}, X_{2wj}) = c_{12vw}$ for all v, w, i, j. He also judges that they are co-exchangeable across questions, so that $Cov(X_{1vi}, X_{2vj}) = d_1$ and $Cov(X_{1vi}, X_{2wj}) = d_2$ for all $v \neq w$, i, j, where d_1 and d_2 are constants. Equivalently, the examiner judges that $Cov(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hv})) = d_1$ and $Cov(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hw})) = d_2$. The examiner muses over this specification. He judges that $Corr(X_{gvi}, X_{hvj}) \leq Corr(X_{gvi}, X_{gvj})$ and $Corr(X_{gvi}, X_{hwj}) \leq Corr(X_{gvi}, X_{gwj})$. By considering our previous specifications, we see that, for constants c_1 and c_2, this reduces to $Cov(X_{gvi}, X_{hvj}) = c_1 Cov(X_{gvi}, X_{gvj})$ and $Cov(X_{gvi}, X_{hwj}) = c_2 Cov(X_{gvi}, X_{gwj})$. The examiner judges that $c_1 = \gamma = c_2$. By considering $Corr(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hv}))$, it is clear that the examiner must specify $-1 \leq \gamma \leq 1$.

The examiner's full beliefs are then:

$$E(\mathcal{M}(X_{gv})) = 0 \ \forall g, v; \tag{2.45}$$

$$E(\mathcal{R}_i(X_{gv})) = 0 \ \forall g, v, i; \tag{2.46}$$

$$Cov(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hw})) = \begin{cases} \gamma_{gh} & \text{if } v = w, \ \forall g, h; \\ 0.25\gamma_{gh} & \text{if } v \neq w, \ \forall g, h; \end{cases} \tag{2.47}$$

$$Cov(\mathcal{R}_i(X_{gv}), \mathcal{R}_j(X_{hw})) = \begin{cases} 4 & \text{if } g = h, \ i = j, \ v = w; \\ 3.5 & \text{if } g = h, \ i = j, \ v \neq w, \ \forall g, i; \\ 0 & \text{otherwise,} \end{cases} \tag{2.48}$$

where $\gamma_{gh} = 1$ if $g = h$, and $\gamma_{gh} = \gamma$ otherwise, for $-1 \leq \gamma \leq 1$. Note that equation (2.45) follows since we have standardised our quantities to have prior mean zero. Notice that since for each marker group we have replicated the problem over the variables that we considered previously, then if we restrict attention to learning in a single marker group having observed data drawn exclusively from that group, then we would replicate the example of Chapter 1. Our interest now is in what happens when we consider learning from more than one group, and also across the groups. Can we link the single group learning with the multi-group learning? Also, what role does the parameter γ play in the design process? Can we gain insights into the effect of changing this parameter?

2.2.2 Learning about the mean components

Suppose that the examiner observes a sample of n exchangeable collections, $\mathcal{C}_{g1}$, $\ldots$, $\mathcal{C}_{gn}$ from each of the two groups. He uses these observations to adjust his beliefs over the mean components collections $\mathcal{M}(\mathcal{C}) = \mathcal{M}(\mathcal{C}_1) \cup \mathcal{M}(\mathcal{C}_2)$. Write $\mathcal{C}(nI_2) = \cup_{g=1}^{2} \cup_{i=1}^{n} \mathcal{C}_{gi}$ to denote the full sample.

Recall that the single group specification is the same as that made when there was no differentiation between markers. One of our interests lies in seeing whether this single group analysis provides us with any guidance in the multi-group case. Thus, for each group, the examiner will construct the analogous quantities to those of equations (1.34) and (1.35), namely

$$\mathcal{M}(X_{g+}) \quad = \quad \mathcal{M}(X_{g1}) + \mathcal{M}(X_{g2}), \tag{2.49}$$

the overall total of the question scores in the gth marker group and

$$\mathcal{M}(X_{g-}) \quad = \quad \mathcal{M}(X_{g1}) - \mathcal{M}(X_{g2}), \tag{2.50}$$

the difference in question scores in the gth marker group. The corresponding resolutions, as drawn from Table 1.2, are

$$\lambda_{(n)g+} \quad = \quad \frac{n}{n+6}; \tag{2.51}$$

$$\lambda_{(n)g-} \quad = \quad \frac{3n}{3n+2}. \tag{2.52}$$

The examiner collects the two group question totals together to form the collection $\mathcal{M}(\mathcal{C}_+) = \{\mathcal{M}(X_{1+}),\ \mathcal{M}(X_{2+})\}$ and similarly he collects the two group question differences together to form the collection $\mathcal{M}(\mathcal{C}_-) = \{\mathcal{M}(X_{1-}), \mathcal{M}(X_{2-})\}$. Let $\langle\mathcal{M}(\mathcal{C}_+)\rangle$, $\langle\mathcal{M}(\mathcal{C}_-)\rangle$ be the collections of linear combinations of the elements of $\mathcal{M}(\mathcal{C}_+)$ and $\mathcal{M}(\mathcal{C}_-)$ respectively. In the usual way we view $\langle\mathcal{M}(\mathcal{C}_+)\rangle$, $\langle\mathcal{M}(\mathcal{C}_-)\rangle$ as vector spaces and construct the corresponding inner product spaces $[\mathcal{M}(\mathcal{C}_+)]$, $[\mathcal{M}(\mathcal{C}_-)]$ over the $\langle\mathcal{M}(\mathcal{C}_+)\rangle$, $\langle\mathcal{M}(\mathcal{C}_-)\rangle$. Notice that every element $l_1\mathcal{M}(X_{1+}) + l_2\mathcal{M}(X_{2+}) \in [\mathcal{M}(\mathcal{C}_+)]$ is uncorrelated with every element $l_1'\mathcal{M}(X_{1-}) + l_2'\mathcal{M}(X_{2-}) \in [\mathcal{M}(\mathcal{C}_-)]$ so that $[\mathcal{M}(\mathcal{C}_+)]$ is orthogonal to $[\mathcal{M}(\mathcal{C}_-)]$.

The examiner considers first the effect of the adjustment of elements in the collection $\mathcal{M}(\mathcal{C}_+)$. Immediate quantities of interest are

$$\begin{aligned}\mathcal{M}(X_{++}) \quad &= \quad \mathcal{M}(X_{1+}) + \mathcal{M}(X_{2+}) &(2.53)\\ &= \quad (\mathcal{M}(X_{11}) + \mathcal{M}(X_{12})) + (\mathcal{M}(X_{21}) + \mathcal{M}(X_{22})), &(2.54)\end{aligned}$$

Component	Resolution (sample size n in each group)
$\mathcal{M}(X_{++}) = \mathcal{M}(X_{1+}) + \mathcal{M}(X_{2+})$	$\lambda_{(n)++} = \frac{n(1+\gamma)}{n(1+\gamma)+6}$
$\mathcal{M}(X_{-+}) = \mathcal{M}(X_{1+}) - \mathcal{M}(X_{2+})$	$\lambda_{(n)-+} = \frac{n(1-\gamma)}{n(1-\gamma)+6}$
$l_1\mathcal{M}(X_{1+}) + l_2\mathcal{M}(X_{2+})$	$\frac{(1+\gamma)(l_1+l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\lambda_{(n)++} + \frac{(1-\gamma)(l_1-l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\lambda_{(n)-+}$

Table 2.1: Resolutions for the collection $\mathcal{M}(\mathcal{C}_+)$, adjusting by $\mathcal{C}(nI_2)$

the total score of the questions marked by either marker; notice how this is proportional to the average score on each question, and

$$\begin{aligned}\mathcal{M}(X_{-+}) &= \mathcal{M}(X_{1+}) - \mathcal{M}(X_{2+}) &(2.55)\\ &= (\mathcal{M}(X_{11}) + \mathcal{M}(X_{12})) - (\mathcal{M}(X_{21}) + \mathcal{M}(X_{22})), &(2.56)\end{aligned}$$

the difference in the totals between the markers. This will give an idea of differences between the markers. The examiner would expect that this quantity is zero if there was no difference between the markers. Notice that the combinations $\mathcal{M}(X_{++})$ and $\mathcal{M}(X_{-+})$ are mutually uncorrelated and thus form an orthogonal basis for $[\mathcal{M}(\mathcal{C}_+)]$. The examiner may draw up Table 2.1 to summarise the effect of the adjustment of $[\mathcal{M}(\mathcal{C}_+)]$ by the sample $\mathcal{C}(nI_2)$. If $\gamma \in [0, 1]$ then for all n the largest resolution is $\lambda_{(n)++}$ corresponding to quantities proportional to $\mathcal{M}(X_{++})$, and the smallest resolution is $\lambda_{(n)-+}$ corresponding to quantities proportional to $\mathcal{M}(X_{-+})$. Notice that the resolution for each element $l_1\mathcal{M}(X_{1+}) + l_2\mathcal{M}(X_{2+})$ of $[\mathcal{M}(\mathcal{C}_+)]$ may be found as a weighted average of the resolutions for these two quantities. The quantities $\mathcal{M}(X_{++})$, $\mathcal{M}(X_{-+})$, when suitably scaled to have prior variance one, are the canonical directions of the adjustment , with $\lambda_{(n)++}$, $\lambda_{(n)-+}$ the canonical resolutions. We expect to learn most about quantities in $[\mathcal{M}(\mathcal{C}_+)]$ that are highly correlated with $\mathcal{M}(X_{++})$, and least for those quantities highly correlated with $\mathcal{M}(X_{-+})$. If $\gamma \in [-1, 0]$, the same comments apply, only $\mathcal{M}(X_{-+})$ is proportional to the canonical direction with the largest resolution and $\mathcal{M}(X_{++})$ proportional to the canonical direction with smallest resolution. Notice that since these features do not depend upon the sample size, the qualitative features of the adjustment within $[\mathcal{M}(\mathcal{C}_+)]$ remain the same for all sample sizes. Thus, for adjustment within this space, the elegant features of Goldstein & Wooff (1998) are preserved.

Observe the role played by the parameter γ. As $\gamma \to 1$, the examiner learns progressively less about the differences between the means of the questions across markers. Notice that $\gamma = 1$ corresponds to a correlation of one between the markers. In this case, notice that $\lambda_{(n)++} = \lambda_{(2n)+}$ (see Table 1.2) and learning is only along $\mathcal{M}(X_{++})$. This is not surprising, for it is as if the examiner had a single exchangeable

Component	Resolution (sample size n in each group)
$\mathcal{M}(X_{+-}) = \mathcal{M}(X_{1-}) + \mathcal{M}(X_{2-})$	$\lambda_{(n)+-} = \frac{3n(1+\gamma)}{3n(1+\gamma)+2}$
$\mathcal{M}(X_{--}) = \mathcal{M}(X_{1-}) - \mathcal{M}(X_{2-})$	$\lambda_{(n)--} = \frac{3n(1-\gamma)}{3n(1-\gamma)+2}$
$l_1\mathcal{M}(X_{1-}) + l_2\mathcal{M}(X_{2-})$	$\frac{(1+\gamma)(l_1+l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\lambda_{(n)+-} + \frac{(1-\gamma)(l_1-l_2)^2}{2(l_1^2+2\gamma l_1 l_2+l_2^2)}\lambda_{(n)--}$

Table 2.2: Resolutions for the collection $\mathcal{M}(\mathcal{C}_-)$, adjusting by $\mathcal{C}(nI_2)$

sequence and took a sample of size $2n$: the labelling of markers is superfluous. $\mathcal{M}(X_{-+})$ has zero mean and variance $5(1-\gamma)$, so in this case corresponds to a quantity with zero variance. Similarly, as $\gamma \to 0$, then $\lambda_{(n)-+} \to \lambda_{(n)++}$, and the learning becomes progressively more even in the two directions. A value of $\gamma = 0$ corresponds to the two markers being uncorrelated. In this case, notice that $\lambda_{(n)++} = \lambda_{(n)-+} = \lambda_{(n)+}$ and also that the resolutions corresponding to $\mathcal{M}(X_{1+})$ and $\mathcal{M}(X_{2+})$ are also $\lambda_{(n)+}$. Thus, it is as if the examiner considered two completely separate exchangeable sequences and observed a sample of size n in each. As $\gamma \to -1$, the examiner learns progressively less about the overall total of the quantities and his learning concentrates upon the differences between the totals of the questions across the markers. This is not surprising. A choice of $\gamma = -1$ corresponds to a correlation of minus one between the markers and, in effect, $\mathcal{M}(X_{2-})$ is the analogous quantity to $\mathcal{M}(X_{1+})$. Notice that $\lambda_{(n)-+} = \lambda_{(2n)+}$ and learning is only along $\mathcal{M}(X_{-+})$. $\mathcal{M}(X_{++})$ has variance $5(1+\gamma)$, so in this case has zero variance.

The examiner now considers the effect of the adjustment of elements in $[\mathcal{M}(\mathcal{C}_-)]$. Immediate quantities of interest are

$$\begin{aligned} \mathcal{M}(X_{+-}) &= \mathcal{M}(X_{1-}) + \mathcal{M}(X_{2-}) && (2.57) \\ &= (\mathcal{M}(X_{11}) + \mathcal{M}(X_{21})) - (\mathcal{M}(X_{12}) + \mathcal{M}(X_{22})), && (2.58) \end{aligned}$$

which is proportional to the difference in question averages, averaged across markers and

$$\begin{aligned} \mathcal{M}(X_{--}) &= \mathcal{M}(X_{1-}) - \mathcal{M}(X_{2-}) && (2.59) \\ &= \mathcal{M}(X_{11}) - \mathcal{M}(X_{12}) - \mathcal{M}(X_{21}) + \mathcal{M}(X_{22}), && (2.60) \end{aligned}$$

a quantity reflecting a measure of differences between the questions. The combinations $\mathcal{M}(X_{+-})$ and $\mathcal{M}(X_{--})$ are uncorrelated and form an orthogonal basis for $[\mathcal{M}(\mathcal{C}_-)]$. The examiner may then construct Table 2.2 to summarise the effect of the adjustment of $[\mathcal{M}(\mathcal{C}_-)]$ by the sample $\mathcal{C}(nI_2)$. The key point to observe here is that the adjustment over each $[\mathcal{M}(\mathcal{C}_-)]$ is essentially the same as that over $[\mathcal{M}(\mathcal{C}_+)]$, with

$\lambda_{(n)++}$ replaced by $\lambda_{(n)+-}$ and $\lambda_{(n)-+}$ replaced by $\lambda_{(n)--}$ and the same linear combinations for the components with the same co-ordinate representation, so that each of the comments that we made when discussing $[\mathcal{M}(\mathcal{C}_+)]$ remain valid for elements in $[\mathcal{M}(\mathcal{C}_-)]$.

Observe what has, in effect, happened. Within each group we solved a variable problem to find the canonical structure for the adjustment of the underlying mean components by a sample of individuals drawn from that group. This yields canonical directions proportional to $\mathcal{M}(X_{g+})$ and $\mathcal{M}(X_{g-})$. The examiner's specifications meant that the variable problem was the same in each group, that is the canonical directions in each group shared the same co-ordinate representation. The analogous canonical directions were gathered together across groups to form the collections $[\mathcal{M}(\mathcal{C}_+)]$ and $[\mathcal{M}(\mathcal{C}_-)]$. These collections were then adjusted, the adjustment being the same in each collection. Thus, we solved a single variable problem and then a single group problem. We remark further that the ratio

$$\frac{RR_{\mathcal{C}(nI_2)}(\mathcal{M}(X_{+-}))}{RR_{\mathcal{C}(nI_2)}(\mathcal{M}(X_{--}))} = \frac{1+\gamma}{1-\gamma} \tag{2.61}$$

remains the same for all choices of numeric specifications in equations (2.43) and (2.44), provided that the γ proportionality is maintained. Additionally, note that

$$\frac{RR_{\mathcal{C}(nI_2)}(\mathcal{M}(X_{++}))}{RR_{\mathcal{C}(nI_2)}(\mathcal{M}(X_{+-}))} = \frac{RR_{\mathcal{C}(nI_2)}(\mathcal{M}(X_{+-}))}{RR_{\mathcal{C}(nI_2)}(\mathcal{M}(X_{--}))}, \tag{2.62}$$

so that this ratio remains constant across the two collections. Changing the numeric specifications in equations (2.43) and (2.44) only changes the $\mathcal{M}(X_{g+})$, $\mathcal{M}(X_{g-})$ to four other quantities, say $\mathcal{M}(X_{g'+'})$ and $\mathcal{M}(X_{g'-'})$. The adjustment over the corresponding collections $\mathcal{M}(\mathcal{C}_{'+'})$ and $\mathcal{M}(\mathcal{C}_{'-'})$ will share the same features as those described above, with the ratio of resolution ratios for these quantities still given by $\{1+\gamma\}/\{1-\gamma\}$.

Notice that the canonical directions to $[\mathcal{M}(\mathcal{C}_+)]$ and $[\mathcal{M}(\mathcal{C}_-)]$ have interpretable forms. $\mathcal{M}(X_{++})$ and $\mathcal{M}(X_{-+})$ are quantities concerning similarities and differences between markers and so we could associate $[\mathcal{M}(\mathcal{C}_+)]$ with marker investigations, whilst $\mathcal{M}(X_{-+})$, $\mathcal{M}(X_{--})$ are quantities concerning similarities and differences between questions across markers and so we could associate $[\mathcal{M}(\mathcal{C}_-)]$ with question investigations.

If we consider a general element $Y = l_1\mathcal{M}(X_{11}) + l_2\mathcal{M}(X_{12}) + l_3\mathcal{M}(X_{21}) + l_4\mathcal{M}(X_{22})$ and wish to examine the effect of the adjustment upon it, then we may show that the resolution of Y may be expressed as:

$$R_{\mathcal{C}(nI_2)}(Y) = \frac{c_{Y++}^2\lambda_{(n)++} + c_{Y-+}^2\lambda_{(n)-+} + c_{Y+-}^2\lambda_{(n)+-} + c_{Y--}^2\lambda_{(n)--}}{c_{Y++}^2 + c_{Y-+}^2 + c_{Y+-}^2 + c_{Y--}^2}, \tag{2.63}$$

where we define $c_{Y++} = Cov(Y, \mathcal{M}(X_{++}))$, $c_{Y-+} = Cov(Y, \mathcal{M}(X_{-+}))$, $c_{Y+-} = Cov(Y, \mathcal{M}(X_{+-}))$ and $c_{Y--} = Cov(Y, \mathcal{M}(X_{--}))$. Thus, the four quantities discussed previously are canonical directions for the full adjustment over all the mean components. Thus, the single group canonical directions retain a fundamental role when we consider the adjustment over groups. By solving them over the group structure, we may find the complete canonical structure for the design.

In summary, we decompose the structure of the design to separate the problem into uncorrelated spaces that share the same properties. Within each space we find uncorrelated directions that summarise the types of information we expect to learn, both for quantities within that space, and also for a general quantity when the directions from each space are collected together. We solve the problem for a single group and then solve across groups for each group canonical resolution in turn. The problem across the groups has the same format for each of the single group resolutions. We see how such decompositions ease the task of picking effective designs. For example, if our primary goal is to choose sample sizes to achieve specific variance reductions, then examination of these spaces shows how we may assess the effects of different sample sizes. Also, we see how the choice of γ only affects the problem across the groups, so that we can easily assess the impact of adjusting γ. These properties apply to a wide class of complex design problems as we now explain.

2.3 Grouped multivariate exchangeable systems

2.3.1 Specification of the system

We are interested in making a series of measurements $\mathcal{C} = \{X_1, \dots, X_{v_0}\}$ on a collection of individuals, where each individual can be classified as coming from one of g_0 groups. For each individual, we wish to measure the same set of v_0 variables. Let $\mathcal{C}_{gi} = \{X_{g1i}, \dots, X_{gv_0i}\}$ be the values of the measurements for the ith individual in the gth group. We collect the measurements for all the individuals in the gth group together as the collection $\mathcal{C}_g^*$ and the total collection of measurements, namely those for all the individuals in all the groups, are collected together in the collection $\mathcal{C}^*$. We judge that the individuals in each group may be thought of as being sampled from a potentially infinite population and we consider that each $\mathcal{C}_g^*$ is second-order exchangeable over the individuals and $\mathcal{C}_1^*$, $\mathcal{C}_2^*$, ..., $\mathcal{C}_{g_0}^*$ are pairwise co-exchangeable across the individuals. That is, our specifications take the form of equations (2.3)

- (2.6). Applying the representation theorem of Goldstein (1986a), reproduced as Theorem 7 in this thesis, we may introduce the collection of random quantities $\mathcal{M}(\mathcal{C}_g) = \{\mathcal{M}(X_{g1}), \ldots, \mathcal{M}(X_{gv_0})\}$, the collection of underlying mean components for the gth group and $\mathcal{R}_i(\mathcal{C}_g) = \{\mathcal{R}_i(X_{g1}), \ldots, \mathcal{R}_i(X_{gv_0})\}$, the collection of residual components for the ith individual in the gth group. We may write for each g, i:

$$X_{gi} \;=\; \mathcal{M}(\mathcal{C}_g) + \mathcal{R}_i(\mathcal{C}_g). \tag{2.64}$$

The complete collection of mean components is given by $\mathcal{M}(\mathcal{C}) = \cup_g \mathcal{M}(\mathcal{C}_g)$. As we are considering a finite number of groups and variables we regard each $\mathcal{C}_g$, $\mathcal{M}(\mathcal{C}_g)$, $\mathcal{R}_i(\mathcal{C}_g)$ as the analogous row vectors, so that, for example, $\mathcal{M}(\mathcal{C}_g) = [\mathcal{M}(X_{g1}) \ldots \mathcal{M}(X_{gv_0})]^T$. We organise our belief specifications as matrices. From equations (2.11) and (2.12), the judgement of second-order co-exchangeability means that our specifications for the covariance structure of the $\mathcal{M}(\mathcal{C}_g)$s and the $\mathcal{R}_i(\mathcal{C}_g)$s are of the form $Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{M}(\mathcal{C}_h)) = C_{gh}$ and $Var(\mathcal{R}_i(\mathcal{C}_g)) = E_g$ where C_{gh}, E_g are general $v_0 \times v_0$ nonnegative definite matrices with (v, w)th entries $(C_{gg})_{vw} = c_{ggvw}$, $(E_g)_{vw} = e_{gvw}$ respectively, C_{gh} is the general $v_0 \times v_0$ matrix with (v, w)th entry $(C_{gh})_{vw} = c_{ghvw}$. The variance and covariance specifications for the $\mathcal{C}_g$s may be determined from those for the $\mathcal{M}(\mathcal{C}_g)$s and $\mathcal{R}_i(\mathcal{C}_g)$s.

We are required to specify $g_0 v_0$ expectations, $E(\mathcal{M}(X_{gv}))$. We assume that this does not daunt us, and is specified. Having done so, we may adjust our quantities so that they all have expectation zero. The reason for doing this is so that if we consider our linear spaces as inner product spaces, then we do not have non-zero quantities having norm zero. We thus standardise each $\mathcal{M}(X_{gv})$ by subtracting its expectation. Goldstein (1986b; p200) explains how we can link the inner-product representation with the actual means specified for the (unstandardised) mean components. Notice from the representation theorem that the residuals already have expectation zero.

We are also required to specify the $v_0 \times v_0$ matrix $Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{M}(\mathcal{C}_h)) = C_{gh}$ for each g, h. With g_0 possible groups, this involves the specification of $(1/2)g_0(g_0 + 1)$ such matrices. We could go ahead and specify them, but to do so may be to ignore de Finetti's advice. We want to avoid the prospect of renouncing *all possibility of illuminating the varied aspects of the question that merit interest.* We might judge that the mean components are second-order exchangeable over groups, that is $C_{gg} = F$ for all g and $C_{gh} = P$ for all $g \neq h$. This leads to a hierarchical model and we shall return to this in Chapter 5. We might judge that the groups are in fact co-exchangeable so that not only would we have no dependence upon groups in the mean component specification, but also $E_g = E$ for all g.

We are considering designs where individuals in each group are being measured by the same set of v_0 variables. For example, the exam questions; or in medical terms, measurements made by the doctor when performing a medical, such as height and weight. It could be that our prior knowledge is more detailed about the variables under consideration than about the groups. The examiner may have more knowledge about the relationships between the questions, a doctor might know very well the usual relationships between say, height and weight. What may be less clear is exactly how these relationships will be affected by the groups. The examiner may not really see how the markers would differ or that they would treat each question differently and likewise the doctor may feel that factors like age or sex may affect his measurements in the same way. The desire for accurate modelling leads one to acknowledge that the second-order exchangeability of individuals exists in certain groups whereas the judgement may be less valid across these groups. To acknowledge the potential for difference in the groups is enough; there is no harm in feeling that the groups may make a difference but not knowing what this difference is. Indeed, this may be the goal of the experiment. The doctor may wish to examine the differences between 'old' and 'young' men and his data may reveal the difference to be greater, or in more varied areas than he expected. Similarly, the examiner may not know how seemingly equally trained markers will differ, but on experience he knows they might. Whilst he may not judge that they will treat the questions differently, on observing the data he may begin to see that new markers are struggling on a particular question, a situation he had previously opined would not be the case.

Such circumstances may lead us to consider starting with fully exchangeable groups. This would lead to the representation

$$X_{gvi} = \mathcal{M}(X_v) + \mathcal{R}_{gi}(X_v), \tag{2.65}$$

where $Cov(\mathcal{R}_{gi}(X_v), \mathcal{R}_{hj}(X_w)) = \delta_{gh}\delta_{ij}e_{vw}$. The notation δ_{ab} represents the quantity 1 when $a = b$ and 0 otherwise.

There may be circumstances that allow us to relax the judgement of fully exchangeable groups, perhaps by increasing the uncertainty about certain groups and also between groups. A natural way to do this is to consider scalings. Consider the example of the exam problem. The examiner may be confident in having an 'underlying' model for the relationships between the questions, which he is able to represent by $\mathcal{M}(\mathcal{C}) = [\mathcal{M}(X_1) \dots \mathcal{M}(X_{v_0})]^T$. However, he acknowledges that the markers may not conform to this; some are liable to mark too strictly and others are likely to be more lenient and so he will have to rescale his 'underlying' model to take account of this. The doctor may also think scalings may be applicable to his considerations. Variables such as height may be scaled by age. Such scalings may or may not be known, but the sort of considerations we wish to make involve judgements of the following type.

In this chapter, we are interested in specifications where the effect of the groups is acting multiplicatively upon the variables. Thus, we are interested in specifications where the effect of the group can be judged to be occurring constantly across the variables. Within the same group, we consider that $Corr(\mathcal{M}(X_{gv}), \mathcal{M}(X_{gw}))$ only depends upon the variables v, w. Across different groups, we consider that it is the choice of group that is crucial. We want to consider that

$$Cov(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hw})) = Cov(\mathcal{M}(X_{gw}), \mathcal{M}(X_{hv})), \tag{2.66}$$

and also that $Corr(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hv}))$ only depends on the groups, g and h.

We wish to consider circumstances where we are only willing and able to make a specification for the covariance structure as follows:

$$\begin{aligned} Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{M}(\mathcal{C}_h)) &= \alpha_{gh} C \; \forall g, h; & (2.67) \\ Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{R}_j(\mathcal{C}_h)) &= 0 \; \forall g, h, j; & (2.68) \\ Cov(\mathcal{R}_i(\mathcal{C}_g), \mathcal{R}_j(\mathcal{C}_h)) &= \begin{cases} \beta_g E & \text{if } g = h, \, i = j; \\ 0 & \text{otherwise,} \end{cases} & (2.69) \end{aligned}$$

where C, E are general $v_0 \times v_0$ nonnegative definite matrices with (v, w)th entries $(C)_{vw} = c_{vw}$, $(E)_{vw} = e_{vw}$ respectively. Let A be the $g_0 \times g_0$ matrix with (g, h)th entry $(A)_{gh} = \alpha_{gh}$, and let B be the $g_0 \times g_0$ diagonal matrix with (g, g)th entry $(B)_{gg} = \beta_g$. Thus, in matrix terms, equation (2.67) may be expressed as:

$$Var(\mathcal{M}(\mathcal{C})) = A \otimes C, \tag{2.70}$$

where $(A \otimes C)$ denotes the direct product of A and C. For further details and properties of the direct product see Searle et al. (1992). By letting $l = [l_1 \dots l_{g_0}]^T$ and considering $Var(\sum_{g=1}^{g_0} l_g \mathcal{M}(X_{gv})) = c_{vv} l^T A l$ for $c_{vv} \neq 0$ we see that A must be nonnegative definite. For simplicity of exposition, in this thesis we shall assume that the matrices of interest are of full rank. Specifically, C, E, A are positive definite and for each g, $\beta_g > 0$. Thus, the results that we shall develop are concerned with designs where the covariance matrices for the underlying mean components are proportional, and the residual variance matrices are also proportional. As in Goldstein & Wooff (1998), if we do not have invertibility, we obtain corresponding results over the linear span of the columns of the matrices that we construct below.

2.3.2 Scaled exchangeability - a potential model approach to the specifications

The following subsection contains an example situation of when the beliefs expressed in equations (2.67) - (2.69) may hold; we emphasise that this provides guidance towards an interpretation of the beliefs as scaled exchangeability, but that the model given in equation (2.71) need not hold in all cases when we have judged our beliefs as in equations (2.67) - (2.69).

Suppose that we judge that the groups act multiplicably on the mean components, that is we view that

$$\mathcal{M}(\mathcal{C}_g) \quad = \quad \mathcal{A}_g \mathcal{M}(X), \tag{2.71}$$

where $\mathcal{A}_g$ and $\mathcal{M}(X)$ are random quantities. Since we have adjusted our quantities to have prior mean zero, then $E(\mathcal{A}_g \mathcal{M}(X)) = 0$. $\mathcal{A}_g$ is a single random quantity and we may collect all the $\mathcal{A}_g$'s together into the $g_0 \times 1$ vector $\mathcal{A} = [\mathcal{A}_1 \dots \mathcal{A}_{g_0}]^T$. $\mathcal{M}(X) = [\mathcal{M}(X_1) \dots \mathcal{M}(X_{v_0})]^T$ is a $v_0 \times 1$ vector containing v_0 random quantities. We could, for example, view that the effect of the different groups is to scale each variable in the group in the same way. Consider the example of the exam problem. The examiner may be happy with his underlying model, represented by $\mathcal{M}(X)$ (see equation (2.65)) but also that to take into account the effect of each marker he needs to scale this, this scaling being given by $\mathcal{A}_g$. Currently, these scalings are unknown to the examiner, but he has prior knowledge about them and is willing and able to specify them. It may be that if this model, as given by equation (2.71), holds then we are interested in learning about the $\mathcal{A}_g$s and $\mathcal{M}(X)$. However, since this subsection is designed to illustrate a case when our beliefs may hold, we bypass this question, instead concentrating only on learning about the products, $\mathcal{A}_g \mathcal{M}(X)$. The examiner might judge that new markers tend to mark a little on the generous side. Alternatively, the performance of established markers from previous years might be known and applied to the scalings for them.

Having established the relationship given by equation (2.71), we still want a second-order specification for the $\mathcal{M}(\mathcal{C})$s, that is we want to create the inner product $[\mathcal{M}(\mathcal{C})]$. Suppose that we judge that there is no interaction between the quadratic products of the $\mathcal{A}_g$s and the $\mathcal{M}(X_v)$s, that is we judge that

$$Cov(\mathcal{A}_g \mathcal{A}_h, \mathcal{M}(X_v)\mathcal{M}(X_w)) \quad = \quad 0 \;\forall g, h, v, w. \tag{2.72}$$

Lemma 4 *If we judge that the model given by equation (2.71) is appropriate and* *we consider that equation (2.72) is a valid specifications then for all g, h we have:*

$$\begin{aligned} Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{M}(\mathcal{C}_h)) &= \\ &\{Cov(\mathcal{A}_g, \mathcal{A}_h) + E(\mathcal{A}_g)E(\mathcal{A}_h)\}\left\{Var(\mathcal{M}(X)) + E(\mathcal{M}(X))E(\mathcal{M}(X))^T\right\}. \quad (2.73) \end{aligned}$$

Thus, the specification for $[\mathcal{M}(\mathcal{C})]$ *is determined by the specification of the two spaces* $[\mathcal{A}]$ *and* $[\mathcal{M}(X)]$.

Proof - If the model given by equation (2.71) holds, then

$$\begin{aligned} Cov(\mathcal{M}(X_{gv}), \mathcal{M}(X_{hw})) &= E(\mathcal{M}(X_{gv})\mathcal{M}(X_{hw})) && (2.74) \\ &= E(\mathcal{A}_g\mathcal{M}(X_v)\mathcal{A}_h\mathcal{M}(X_w)) && (2.75) \\ &= E(\mathcal{A}_g\mathcal{A}_h\mathcal{M}(X_v)\mathcal{M}(X_w)) && (2.76) \\ &= E(\mathcal{A}_g\mathcal{A}_h)E(\mathcal{M}(X_v)\mathcal{M}(X_w)), && (2.77) \end{aligned}$$

where equation (2.77) follows from equation (2.72). Noting that

$$Cov(\mathcal{A}_g, \mathcal{A}_h) = E(\mathcal{A}_g\mathcal{A}_h) - E(\mathcal{A}_g)E(\mathcal{A}_h), \quad (2.78)$$

and

$$Cov(\mathcal{M}(X_v), \mathcal{M}(X_w)) = E(\mathcal{M}(X_v)\mathcal{M}(X_w)) - E(\mathcal{M}(X_v))E(\mathcal{M}(X_w)) \quad (2.79)$$

yields the result. □

It is clear that both the $g_0 \times g_0$ matrix $E(\mathcal{A})E(\mathcal{A})^T$ and the $v_0 \times v_0$ matrix $E(\mathcal{M}(X))E(\mathcal{M}(X))^T$ are nonnegative definite. Setting

$$\begin{aligned} a_{gh} &= \{Cov(\mathcal{A}_g, \mathcal{A}_h) + E(\mathcal{A}_g)E(\mathcal{A}_h)\}; && (2.80) \\ c_{vw} &= Cov(\mathcal{M}(X_v), \mathcal{M}(X_w)) + E(\mathcal{M}(X_v))E(\mathcal{M}(X_w)), && (2.81) \end{aligned}$$

then we have for this model that

$$Var(\mathcal{M}(\mathcal{C})) = A \otimes C, \quad (2.82)$$

which conforms with the specification given in equation (2.70).

Notice here that we only require equation (2.72). It may also be the case that we judge that there is no interaction between the $\mathcal{A}_g$s and the $\mathcal{M}(X_v)$s, so that

$$Cov(\mathcal{A}_g, \mathcal{M}(X_v)) = 0 \ \forall g, v. \quad (2.83)$$

In this instance, we have that $E(\mathcal{M}(X_{gv})) = E(\mathcal{A}_g)E(\mathcal{M}(X_v))$. Taking $E(\mathcal{M}(X_v)) = 0$ makes our notion of the effect of the groups scaling an underlying model clearer.

Suppose that we also judge that the residual components may be scaled in a similar vein. That is we consider that the following model is valid:

$$\mathcal{R}_i(\mathcal{C}_g) = \mathcal{B}_g\mathcal{R}_{gi}(X), \tag{2.84}$$

where

$$Cov(\mathcal{B}_g\mathcal{B}_h, (\mathcal{R}_{g'i}(X_v))^2) = 0 \; \forall g, h, g', v, i; \tag{2.85}$$

$$Cov(\mathcal{R}_{gi}(X), \mathcal{R}_{hj}(X)) = \begin{cases} E & \text{if } g = h,\ i = j; \\ 0 & \text{otherwise.} \end{cases} \tag{2.86}$$

Equation (2.86) is a natural one to take. We are considering the residual components to be scaled across groups and that is the only effect of the group. Thus, the unscaled residual component should be uncorrelated with all different individuals and not depend on the group. This is what the judgement of equation (2.86) achieves. Notice that this scaling allows us more scope than just having the same residual variance for all of the individuals, so that $\mathcal{B}_g = 1$ for all g. We have the following lemma; the proof is identical to that of Lemma 4.

Lemma 5 *If we judge that the model given by equation (2.84) is appropriate and we consider that equations (2.85) and (2.86) are valid specifications then for all g, h, i, j we have:*

$$Cov(\mathcal{R}_i(\mathcal{C}_g), \mathcal{R}_j(\mathcal{C}_h)) = \begin{cases} \{Var(\mathcal{B}_g) + E(\mathcal{B}_g)^2\}E & \textit{if } g = h,\ i = j; \\ 0 & \textit{otherwise.} \end{cases} \tag{2.87}$$

Thus, the specification for each $[\mathcal{R}_i(\mathcal{C}_g)]$ is determined by the specification of the spaces $[\mathcal{B}]$ and each $[\mathcal{R}_{gi}(X)]$.

Setting $\beta_g = \{Var(\mathcal{B}_g) + E(\mathcal{B}_g)^2\}$ and letting B be the $g_0 \times g_0$ diagonal matrix with (g, g)th entry $(B)_{gg} = \beta_g$ then we have for this model that

$$Cov(\mathcal{R}_i(\mathcal{C}_g), \mathcal{R}_j(\mathcal{C}_h)) = \begin{cases} \beta_g E & \text{if } g = h,\ i = j; \\ 0 & \text{otherwise,} \end{cases} \tag{2.88}$$

which conforms with the specification given in equation (2.69). Notice that the total model is

$$X_{gi} = \mathcal{A}_g\mathcal{M}(X) + \mathcal{B}_g\mathcal{R}_{gi}(X). \tag{2.89}$$

We may think of this model as one of scaled exchangeability. To emphasise this, there is nothing to prevent us from judging that $\mathcal{A}_g = \mathcal{B}_g$ so that

$$X_{gi} = \mathcal{A}_g\{\mathcal{M}(X) + \mathcal{R}_{gi}(X)\}, \tag{2.90}$$

perhaps corresponding to $X_{gi} = \mathcal{A}_g X_i$. Thus, the effect of the introduction of the groups is to scale our quantities, which we then regard as exchangeable. That is the X_is, where $X_i = \mathcal{A}_g^{-1}\mathcal{M}(X_g)$ form a second-order exchangeable sequence in the sense of Chapter 1. Thus, we see immediately the link between this scaled model and the fully exchangeable groups model, which leads to the representation

$$X_{gvi} = \mathcal{M}(X_v) + \mathcal{R}_{gi}(X_v). \tag{2.91}$$

2.4 Adjusting the mean components

We wish to take a sample of size $n_g > 0$ from the gth group for $g = 1, \ldots, g_0$ and use these observations to adjust our beliefs over the mean collection, $\mathcal{M}(\mathcal{C}) = \{\mathcal{M}(\mathcal{C}_1), \ldots, \mathcal{M}(\mathcal{C}_{g_0})\}$. Again, we analogously think of this as the corresponding $g_0 v_0 \times 1$ column vector, $\mathcal{M}(\mathcal{C}) = [\mathcal{M}(\mathcal{C}_1)^T \ldots \mathcal{M}(\mathcal{C}_{g_0})^T]^T$. We collect the sample sizes together into the matrix $N = diag(n_1, \ldots, n_{g_0})$. Thus in the gth group, we will observe the n_g exchangeable collections, $\mathcal{C}_{g1}, \ldots, \mathcal{C}_{gn_g}$ where $\mathcal{C}_{gi} = \{X_{g1i}, \ldots, X_{gv_0i}\}$. The complete sample in the gth group is gathered together as $\mathcal{C}_g(n_g) = \cup_{i=1}^{n_g}\mathcal{C}_{gi}$. The total sample is $\mathcal{C}(N) = \cup_{g=1}^{g_0}\mathcal{C}_g(n_g)$. In our analogous vector form, we have $\mathcal{C}_{gi} = [X_{g1i} \ldots X_{gv_0i}]^T$, $\mathcal{C}_g(n_g) = [\mathcal{C}_{g1}^T \ldots \mathcal{C}_{gn_g}^T]^T$ and $\mathcal{C}(N) = [\mathcal{C}_1(n_1)^T \ldots \mathcal{C}_{g_0}(n_{g_0})^T]^T$.

Let $\mathcal{S}_{n_g}(\mathcal{C}_g) = \{\mathcal{S}_{n_g}(X_{g1}), \ldots, \mathcal{S}_{n_g}(X_{gv_0})\}$ be the collection of sample averages for the gth group. The complete collection of sample averages are denoted by $\mathcal{S}_N(\mathcal{C}) = \{\mathcal{S}_{n_1}(\mathcal{C}_1), \ldots, \mathcal{S}_{n_{g_0}}(\mathcal{C}_{g_0})\}$. Using the analogous vector notation, we have $\mathcal{S}_{n_g}(\mathcal{C}_g) = [\mathcal{S}_{n_g}(X_{g1}) \ldots \mathcal{S}_{n_g}(X_{gv_0})]^T$ and $\mathcal{S}_N(\mathcal{C}) = [\mathcal{S}_{n_1}(\mathcal{C}_1)^T \ldots \mathcal{S}_{n_{g_0}}(\mathcal{C}_{g_0})^T]^T$. From Theorem 8, we have that the sample means, $\mathcal{S}_N(\mathcal{C})$, are Bayes linear sufficient for the full sample $\mathcal{C}(N)$ for adjusting the mean components collection $\mathcal{M}(\mathcal{C})$.

Lemma 6 *The second-order specifications for the mean components, $\mathcal{M}(\mathcal{C})$, and the sample means, $\mathcal{S}_N(\mathcal{C})$, may be expressed as*

$$\begin{aligned} Var(\mathcal{M}(\mathcal{C})) &= A \otimes C; && (2.92)\\ Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) &= A \otimes C; && (2.93)\\ Var(\mathcal{S}_N(\mathcal{C})) &= (A \otimes C) + (N^{-1}B \otimes E). && (2.94) \end{aligned}$$

Proof - The results follow immediately from equations (2.67), (2.68), (2.69), (2.17) and (2.18). □

We now consider separately the analysis of variables and of groups, as follows.

2.4.1 Underlying canonical variable problem

Definition 12 *The underlying canonical variable directions are defined as the columns of the matrix $Y = [Y_1 \dots Y_{v_0}]$ solving the generalised eigenvalue problem*

$$CY = (C+E)Y\Phi, \tag{2.95}$$

where $\Phi = diag(\phi_1, \dots, \phi_{v_0})$ is the matrix of eigenvalues. Y is chosen so that $Y^TCY = I_{v_0}$, $Y^T(C+E)Y\Phi = I_{v_0}$. The ordered eigenvalues $1 > \phi_1 \geq \dots \geq \phi_{v_0} > 0$ are termed the underlying canonical variable resolutions.

We are able to choose Y in the stated form through standard results on simultaneous diagonalisation of matrices, see for example Theorem VI.1.15 of Stewart & Sun (1990). To motivate this definition, consider adjusting the mean components of a single group, g, from a sample, of size n_g, drawn purely from that group. Since we are dealing with a single group, then we have a second-order exchangeable sequence. This is the problem tackled in Goldstein & Wooff (1998). Here they show that irrespective of the sample size the resolution transform for the underlying population structure induced by a second-order exchangeable sample has essentially the same form. The canonical directions are the same for all sample sizes, with simple modifications of the canonical resolutions. Thus, to find the canonical structure for all sample sizes, we are required to solve a single eigenvector/value problem. For our specifications, this problem is that given by equation (2.95). This underlying canonical variable problem should also be compared with equation (1.78), and the corresponding resolution transform given in equation (1.77). Denoting the resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C}_g)]$ by $\mathcal{C}_g(n_g)$ as $T_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]}$, we may rewrite Theorem 3 of Goldstein & Wooff (1998) as:

Theorem 11 *The resolution transform matrix, $T_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]}$, is calculated as*

$$T_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]} = \{\alpha_{gg}C + (1/n_g)\beta_g E\}^{-1}(\alpha_{gg}C). \tag{2.96}$$

For each $s = 1, \dots, v_0$, the canonical directions for the adjustment of $[\mathcal{M}(\mathcal{C}_g)]$ by $\mathcal{C}_g(n_g)$ are given by

$$Z_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} = \sqrt{\frac{1}{\alpha_{gg}}}Y_s^T\mathcal{M}(\mathcal{C}_g), \tag{2.97}$$

with the corresponding canonical resolutions given by

$$\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} = \frac{n_g\alpha_{gg}\phi_s}{n_g\alpha_{gg}\phi_s + \beta_g(1-\phi_s)}. \tag{2.98}$$

Notice that we also have similarly strong coherence properties across the groups. Irrespective of sample size and the specific group in question, up to a scale factor to ensure a prior variance of one, the co-ordinate representation of the canonical resolutions is the same and found by solving the eigenvector/value problem given by equation (2.95). From the eigenvalues of this problem, we may easily modify to obtain the $\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ for any g and any n_g. Thus, the qualitative information we learn about a group from observing a sample drawn from that group is the same for each group. Quantitatively, if we observe the same size in each group, then we would learn more in the groups with high values of α_{gg}/β_g. In terms of interpreting the mean components as the prior and the residual as the likelihood, we learn most in the groups with the highest ratio of prior to likelihood variance.

The underlying canonical variable structure thus plays an important role in the simplification of the adjustment of exchangeable sequences as shown in Goldstein & Wooff (1998). The solution of the underlying canonical variable problem and its use in the modelling of proportionality in co-exchangeable models may also be found in Section 1.6 of Wooff & Goldstein (1994), which also contains some early results in the area. Throughout this thesis, we shall show that the underlying canonical variable structure plays a strong role in many other design problems. We shall find the following lemma useful; $\epsilon_{v_0 v}$ represents the vth column of the $v_0 \times v_0$ identity matrix.

Lemma 7 *Let S, T be arbitrary $p \times q$ matrices and each X_s, for $s = 1, \dots, v_0$, be an arbitrary $q \times r$ matrix. Then*

$$\begin{bmatrix} I_p \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_p \otimes \epsilon_{v_0 v_0}^T \end{bmatrix} \left[I_p \otimes Y^{-1}(C+E)^{-1}\right](S \otimes C)\left[(X_1 \otimes Y_1)\dots(X_{v_0} \otimes Y_{v_0})\right] = \oplus_{s=1}^{v_0} \phi_s S X_s; \quad (2.99)$$

$$\begin{bmatrix} I_p \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_p \otimes \epsilon_{v_0 v_0}^T \end{bmatrix} \left[I_p \otimes Y^{-1}(C+E)^{-1}\right](T \otimes E)\left[(X_1 \otimes Y_1)\dots(X_{v_0} \otimes Y_{v_0})\right] = \oplus_{s=1}^{v_0} (1-\phi_s) T X_s. \quad (2.100)$$

Proof - By using equation (2.95) we find

$$\begin{aligned} \left[I_p \otimes Y^{-1}(C+E)^{-1}\right](S \otimes C) &= S \otimes Y^{-1}(C+E)^{-1}C && (2.101) \\ &= S \otimes \Phi Y^{-1}, && (2.102) \end{aligned}$$

and

$$\left[I_p \otimes Y^{-1}(C+E)^{-1}\right](T \otimes E) = T \otimes Y^{-1}(C+E)^{-1}E \tag{2.103}$$
$$= T \otimes (I_{v0} - \Phi)Y^{-1}. \tag{2.104}$$

Post multiplying equations (2.102) and (2.104) by $(X_s \otimes Y_s)$ gives

$$(S \otimes \Phi Y^{-1})(X_s \otimes Y_s) = SX_s \otimes \Phi Y^{-1} Y_s \tag{2.105}$$
$$= SX_s \otimes \Phi \epsilon_{v_0 s} \tag{2.106}$$
$$= \phi_s SX_s \otimes \epsilon_{v_0 s}; \tag{2.107}$$
$$(T \otimes (I_{v_0} - \Phi)Y^{-1})(X_s \otimes Y_s) = TX_s \otimes (I_{v_0} - \Phi)Y^{-1}Y_s \tag{2.108}$$
$$= TX_s \otimes (I_{v_0} - \Phi)\epsilon_{v_0 s} \tag{2.109}$$
$$= (1-\phi_s)TX_s \otimes \epsilon_{v_0 s}. \tag{2.110}$$

Pre multiplying equations (2.107) and (2.110) by $(I_p \otimes \epsilon_{v_0 t}^T)$ gives

$$(I_p \otimes \epsilon_{v_0 t}^T)(\phi_s SX_s \otimes \epsilon_{v_0 s}) = \delta_{ts}\phi_s SX_s; \tag{2.111}$$
$$(I_p \otimes \epsilon_{v_0 t}^T)((1-\phi_s)TX_s \otimes \epsilon_{v_0 s}) = \delta_{ts}(1-\phi_s)TX_s, \tag{2.112}$$

and equations (2.99) and (2.100) follow. □

The notation $A \oplus B$ is used to denote the direct sum of two matrices A and B and is defined to be

$$A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}. \tag{2.113}$$

This easily extends for more than two matrices, and

$$\oplus_{i=1}^{k} A_i = A_1 \oplus A_2 \oplus \cdots \oplus A_k. \tag{2.114}$$

For further details on the direct sum, and the matrix representation of $\oplus_{i=1}^{k} A_i$, see Searle (1982; p264).

2.4.2 The underlying canonical variable problem for the examiner

Recall the examiner example. From equations (2.47) and (2.48) we see that he has

$$Var(\mathcal{M}(\mathcal{C})) = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}; \tag{2.115}$$
$$Cov(\mathcal{R}_i(\mathcal{C}_g), \mathcal{R}_j(\mathcal{C}_h)) = \begin{cases} \begin{pmatrix} 4 & 3.5 \\ 3.5 & 4 \end{pmatrix} & \text{if } g = h,\ i = j; \\ 0 & \text{otherwise,} \end{cases} \tag{2.116}$$

so that

$$A = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix};\ C = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix};\ B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ E = \begin{pmatrix} 4 & 3.5 \\ 3.5 & 4 \end{pmatrix}. \quad (2.117)$$

He considers taking a sample of size n in each group so that $N = nI_2$. The examiner solves the underlying canonical variable problem corresponding to C, E of equation (2.117). This gives

$$Y = \begin{pmatrix} \sqrt{\frac{2}{3}} & \sqrt{\frac{2}{5}} \\ -\sqrt{\frac{2}{3}} & \sqrt{\frac{2}{5}} \end{pmatrix};\Phi = \begin{pmatrix} \frac{3}{5} & 0 \\ 0 & \frac{1}{7} \end{pmatrix}. \quad (2.118)$$

He then applies Theorem 11 to investigate the adjustment of $[\mathcal{M}(\mathcal{C}_g)]$ by $\mathcal{C}_g(n)$. He has that the canonical directions for this adjustment are given by:

$$Z_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n)]1} = \sqrt{\frac{2}{3}}\{\mathcal{M}(X_{g1}) - \mathcal{M}(X_{g2})\}; \quad (2.119)$$

$$Z_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n)]2} = \sqrt{\frac{2}{5}}\{\mathcal{M}(X_{g1}) + \mathcal{M}(X_{g2})\}. \quad (2.120)$$

Equation (2.119) should be immediately compared to equation (2.50) and equation (2.120) to equation (2.49). The corresponding canonical resolutions are:

$$\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n)]1} = \frac{n(3/5)}{n(3/5) + (1 - (3/5))}; \quad (2.121)$$

$$\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n)]1} = \frac{n(1/7)}{n(1/7) + (1 - (1/7))}. \quad (2.122)$$

Equation (2.121) should be compared with equation (2.52) and equation (2.122) with (2.51).

2.4.3 Underlying canonical group problem

Definition 13 *The underlying canonical group directions are defined as the columns of the matrix $W = [W_1 \dots W_{g_0}]$ solving the generalised eigenvalue problem*

$$AW = (A + N^{-1}B)W\Psi, \quad (2.123)$$

where $\Psi = diag(\psi_1, \dots, \psi_{g_0})$ is the matrix of eigenvalues. W is chosen so that $W^TAW = I_{g_0}$, $W^T(A + N^{-1}B)W\Psi = I_{g_0}$. The ordered eigenvalues $1 > \psi_1 \geq \dots \geq \psi_{g_0} > 0$ are termed the underlying canonical group resolutions.

To motivate this definition, for each $s = 1, \dots v_0$, form the collection

$$\mathcal{Z}_{(N)s} = \{Z_{[\mathcal{M}(\mathcal{C}_1)/\mathcal{C}_1(n_1)]s}, \dots, Z_{[\mathcal{M}(\mathcal{C}_{g_0})/\mathcal{C}_{g_0}(n_{g_0})]s}\}. \quad (2.124)$$

For each group g, $Z_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ is the sth canonical direction for the adjustment of $[\mathcal{M}(\mathcal{C}_g)]$ by $\mathcal{C}_g(n_g)$, as given by equation (2.97). The collection $\mathcal{Z}_{(N)s}$ thus consists of the canonical directions from each group adjustment that have the same co-ordinate representation. Notice that from the underlying canonical variable problem given by equation (2.95), if we have repeated canonical resolutions, then we will always choose the same basis co-ordinate representation in each group (up to a scale factor) and so the formation of each $\mathcal{Z}_{(N)s}$ is always clear-cut. Letting D be the $g_0 \times g_0$ matrix

$$D = diag\left(\sqrt{\frac{1}{\alpha_{11}}}, \ldots, \sqrt{\frac{1}{\alpha_{g_0 g_0}}}\right), \tag{2.125}$$

then in vector notation, we also represent $\mathcal{Z}_{(N)s}$ as the $g_0 \times 1$ vector

$$\mathcal{Z}_{(N)s} = D(I_{g_0} \otimes Y_s^T)\mathcal{M}(\mathcal{C}). \tag{2.126}$$

Lemma 8 *The second-order relationships between $\mathcal{Z}_{(N)s}$ and $\mathcal{S}_N(\mathcal{C})$ may be expressed as follows*

$$\begin{aligned}
Var(\mathcal{Z}_{(N)s}) &= DAD; && (2.127)\\
Cov(\mathcal{Z}_{(N)s}, \mathcal{S}_N(\mathcal{C})) &= D[\epsilon_{v_0 s}^T \otimes A][(I_{g_0} \otimes Y_1)\ldots(I_{g_0} \otimes Y_{v_0})]^{-1}; && (2.128)\\
Var(\mathcal{S}_N(\mathcal{C})) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0}\{\phi_s A + (1-\phi_s)N^{-1}B\}][(I_{g_0} \otimes Y_1)\ldots(I_{g_0} \otimes Y_{v_0})]^{-1}; && (2.129)\\
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N)s}) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\epsilon_{v_0 s} \otimes \{\phi_s A\}]D. && (2.130)
\end{aligned}$$

Proof - Equation (2.127) follows from equation (2.92) since

$$\begin{aligned}
Var(\mathcal{Z}_{(N)s}) &= D(I_{g_0} \otimes Y_s^T)Var(\mathcal{M}(\mathcal{C}))(I_{g_0} \otimes Y_s)D && (2.131)\\
&= D(A \otimes Y_s^T C Y_s)D && (2.132)\\
&= DAD, && (2.133)
\end{aligned}$$

where equation (2.133) follows by the choice of Y in Definition 12. Equation (2.128) follows from equation (2.93) since

$$\begin{aligned}
Cov(\mathcal{Z}_{(N)s}, \mathcal{S}_N(\mathcal{C})) &= D(I_{g_0} \otimes Y_s^T) Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) && (2.134)\\
&= D(A \otimes Y_s^T C) && (2.135)\\
&= D(A \otimes Y_s^T C)[(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})][(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})]^{-1} && (2.136)\\
&= D[(A \otimes Y_s^T C Y_1) \dots (A \otimes Y_s^T C Y_{v_0})][(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})]^{-1} && (2.137)\\
&= D[(A \otimes \delta_{s1}) \dots (A \otimes \delta_{sv_0})][(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})]^{-1} && (2.138)\\
&= D[\epsilon_{v_0 s}^T \otimes A][(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})]^{-1}, && (2.139)
\end{aligned}$$

where the choice of Y in Definition 12 yields equation (2.138). Equation (2.129) follows by applying Lemma 7 to equation (2.94) with $S = A$, $T = N^{-1}B$ and $X_s = I_{g_0}$ for all $s = 1, \dots, v_0$. To obtain equation (2.130), note that by using the transpose of equation (2.93) we have

$$\begin{aligned}
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N)s}) &= Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}))(I_{g_0} \otimes Y_s) D && (2.140)\\
&= (A \otimes C)(I_{g_0} \otimes Y_s) D. && (2.141)
\end{aligned}$$

We apply Lemma 7 to equation (2.93) with $S = A$, $X_s = I_{g_0}$ for all $s = 1, \dots, v_0$ and substituting into equation (2.140) gives

$$\begin{aligned}
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N)s}) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{t=1}^{v_0}\{\phi_t A\}][(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})]^{-1}(I_{g_0} \otimes Y_s) D && (2.142)\\
&= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{t=1}^{v_0}\{\phi_t A\}][\epsilon_{v_0 s} \otimes I_{g_0}] D && (2.143)\\
&= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\epsilon_{v_0 s} \otimes \{\phi_s A\}] D. && (2.144)
\end{aligned}$$

□

The resolution transform for the adjustment of $[\mathcal{Z}_{(N)s}]$ by $\mathcal{C}(N)$ is denoted by $T_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]}$. We have the following theorem.

Theorem 12 *For the adjustment of $[\mathcal{Z}_{(N)s}]$ by $\mathcal{C}(N)$, the resolution transform matrix is calculated as*

$$T_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]} = D^{-1}\{\phi_s A + (1-\phi_s)N^{-1}B\}^{-1}\{\phi_s A\} D. \qquad (2.145)$$

The canonical directions are given by

$$Z_{(N)ds} = (D^{-1}W_d)^T \mathcal{Z}_{(N)s} = (W_d \otimes Y_s)^T \mathcal{M}(\mathcal{C}), \tag{2.146}$$

for each $d = 1, \ldots, g_0$ with corresponding canonical resolutions given by

$$\lambda_{(N)ds} = \frac{\psi_d \phi_s}{\psi_d \phi_s + (1-\psi_d)(1-\phi_s)}. \tag{2.147}$$

W_d is the dth underlying canonical group direction as given in Definition 13; ψ_d is the corresponding dth underlying group resolution.

Proof - Notice that for each s, $[\mathcal{Z}_{(N)s}] \subset [\mathcal{M}(\mathcal{C})]$. It thus follows immediately that $[\mathcal{Z}_{(N)s}] \perp [\mathcal{T}(\mathcal{C}(N))]$, so that $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting $\mathcal{Z}_{(N)s}$ (see Theorem 8). Thus, $T_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]} = T_{[\mathcal{Z}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}$, where $T_{[\mathcal{Z}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}$ is the resolution transform for the adjustment of $[\mathcal{Z}_{(N)s}]$ by $\mathcal{S}_N(\mathcal{C})$. From equation (1.74), $T_{[\mathcal{Z}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}$ may be computed as

$$T_{[\mathcal{Z}_{(N)s}/\mathcal{S}_N(\mathcal{C})]} = \{Var(\mathcal{Z}_{(N)s})\}^{-1} Cov(\mathcal{Z}_{(N)s}, \mathcal{S}_N(\mathcal{C}))\{Var(\mathcal{S}_N(\mathcal{C}))\}^{-1} Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N)s}). \tag{2.148}$$

Thus, by inverting equation (2.127), post multiplying by equation (2.128) and then the inversion of (2.129) and finally by (2.130) we may obtain equation (2.145).

From the solution of the underlying group problem as given by equation (2.123), it is straightforward to see that

$$T_{[\mathcal{Z}_{(N)s}/\mathcal{S}_N(\mathcal{C})]} = D^{-1}W\{\phi_s\Psi\}\{\phi_s\Psi + (1-\phi_s)(I_{g_0} - \Psi)\}^{-1}W^{-1}D, \tag{2.149}$$

so that,

$$T_{[\mathcal{Z}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}(D^{-1}W) = D^{-1}W\{\phi_s\Psi\}\{\phi_s\Psi + (1-\phi_s)(I_{g_0} - \Psi)\}^{-1}. \tag{2.150}$$

Hence, $\Lambda_{(N)s} = \{\phi_s\Psi\}\{\phi_s\Psi + (1-\phi_s)(I_{g_0} - \Psi)\}^{-1}$ is the matrix whose diagonal elements are the canonical resolutions of the adjustment. To confirm that the $(D^{-1}W_d)^T\mathcal{Z}_{(N)s}$ are the corresponding canonical resolutions, we verify that they are mutually uncorrelated with prior variance one. Notice from equation (2.146) we have

$$\begin{aligned} Cov((D^{-1}W_d)^T\mathcal{Z}_{(N)s}, (D^{-1}W_{d'})^T\mathcal{Z}_{(N)s'}) &= \\ (W_d^T \otimes Y_s^T) Var(\mathcal{M}(\mathcal{C}))(W_{d'} \otimes Y_{s'}) & \quad (2.151) \\ = W_d^T A W_{d'} \otimes Y_s^T C Y_{s'} & \quad (2.152) \\ = \delta_{dd'}\delta_{ss'}, & \quad (2.153) \end{aligned}$$

where equation (2.153) follows by the choice of Y in Definition 12 and the choice of W in Definition 13. Setting $s = s'$ completes the verification. Notice that equation (2.153) also shows that since the $Z_{(N)ds}$ form a basis for $[\mathcal{Z}_{(N)s}]$, then if $s \neq s'$, $[\mathcal{Z}_{(N)s}] \perp [\mathcal{Z}_{(N)s'}]$. □

The underlying canonical group structure thus relates to an identifiable inference problem; that of learning about the relationships of the sth most important variable directions when considered across the groups. Note that the adjustment in each $[\mathcal{Z}_{(N)s}]$, for $s = 1, \dots, v_0$ shares the same qualitative information but, this information does, in general, vary with the sample sizes.

We may use the results of this theorem to investigate the affect of the adjustment for any quantity in $[\mathcal{Z}_{(N)s}]$, since by applying equation (1.62) we have for any $\mathcal{X} \in [\mathcal{Z}_{(N)s}]$,

$$Var_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]}(\mathcal{X}) = \sum_{d=1}^{g_0}(1 - \lambda_{(N)ds})Cov(\mathcal{X}, \mathcal{Z}_{(N)ds})^2, \tag{2.154}$$

where $Var_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]}(\mathcal{X})$ is the adjusted variance of $\mathcal{X}$.

2.4.4 The underlying canonical group problem for the examiner

By using equation (2.119), the examiner forms the collection

$$\begin{aligned}\mathcal{Z}_{(nI_2)1} &= \\ &\{\sqrt{(2/3)}(\mathcal{M}(X_{11}) - \mathcal{M}(X_{12})), \sqrt{(2/3)}(\mathcal{M}(X_{21}) - \mathcal{M}(X_{22}))\}.\end{aligned} \tag{2.155}$$

The inner product space $[\mathcal{Z}_{(nI_2)1}]$ may then be formed, and is the same as $[\mathcal{M}(\mathcal{C}_-)]$, with the adjustment summarised in Table 2.2. Similarly, the examiner may form the collection $\mathcal{Z}_{(nI_2)2}$. The corresponding inner product space, $[\mathcal{Z}_{(nI_2)2}]$, is the same as $[\mathcal{M}(\mathcal{C}_+)]$, with the adjustment summarised in Table 2.1. From Theorem 12, the role of $\mathcal{M}(X_{++})$ (see (2.53)), $\mathcal{M}(X_{-+})$ (see (2.55)), $\mathcal{M}(X_{+-})$ (see (2.57)) and $\mathcal{M}(X_{--})$ (see (2.59)) follow immediately from the solution of the canonical group problem corresponding to the matrices A, N, B (as given by equation (2.117)). The solution of this problem gives:

$$W = \begin{pmatrix} \sqrt{\frac{1}{2(1+\gamma)}} & \sqrt{\frac{1}{2(1-\gamma)}} \\ \sqrt{\frac{1}{2(1+\gamma)}} & -\sqrt{\frac{1}{2(1-\gamma)}} \end{pmatrix}; \Psi = \begin{pmatrix} \frac{1+\gamma}{(1/n)+1+\gamma} & 0 \\ 0 & \frac{1-\gamma}{(1/n)+1-\gamma} \end{pmatrix}, \tag{2.156}$$

so that, for example

$$Z_{(nI_2)11} = (W_1 \otimes Y_1)^T \mathcal{M}(\mathcal{C}) \propto \mathcal{M}(X_{+-}); \quad (2.157)$$

$$Z_{(nI_2)21} = (W_2 \otimes Y_1)^T \mathcal{M}(\mathcal{C}) \propto \mathcal{M}(X_{--}), \quad (2.158)$$

and

$$\lambda_{(nI_2)11} = \frac{\frac{3}{5}\left(\frac{1+\gamma}{(1/n)+1+\gamma}\right)}{\frac{3}{5}\left(\frac{1+\gamma}{(1/n)+1+\gamma}\right)+\frac{2}{5}\left(\frac{(1/n)}{(1/n)+1+\gamma}\right)} = \lambda_{(n)+-}; \quad (2.159)$$

$$\lambda_{(nI_2)21} = \frac{\frac{3}{5}\left(\frac{1-\gamma}{(1/n)+1-\gamma}\right)}{\frac{3}{5}\left(\frac{1-\gamma}{(1/n)+1-\gamma}\right)+\frac{2}{5}\left(\frac{(1/n)}{(1/n)+1-\gamma}\right)} = \lambda_{(n)--}. \quad (2.160)$$

Equations (2.159) and (2.160) are the two canonical resolutions we found in Table 2.2. Notice that in this balanced design the canonical directions $Z_{(nI_2)ds}$ do not depend upon the sample size, n, observed in each group.

In equation (2.63) we illustrated that the canonical directions of the adjustment of each $[\mathcal{Z}_{(nI_2)s}]$ when collected together formed the set of canonical resolutions for the adjustment of $[\mathcal{M}(\mathcal{C})]$. We shall now show that this is not a special feature of this example, nor of the balance, but that the canonical variable and group analysis completely determine the adjustment of the full collection $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$, where N is any sample.

2.4.5 The adjustment of the mean components by the observed sample

Theorem 13 *The adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$ satisfies the following properties*

1. There exist v_0 g_0-dimensional orthogonal subspaces, $[\mathcal{Z}_{(N)1}]$, ..., $[\mathcal{Z}_{(N)v_0}]$ of $[\mathcal{M}(\mathcal{C})]$. For each s,

$$\mathcal{Z}_{(N)s} = \{Z_{[\mathcal{M}(\mathcal{C}_1)/\mathcal{C}_1(n_1)]s}, \cdots, Z_{[\mathcal{M}(\mathcal{C}_{g_0})/\mathcal{C}_{g_0}(n_{g_0})]s}\}. \quad (2.161)$$

2. The canonical directions for the adjustment in each $[\mathcal{Z}_{(N)s}]$ share the same *co-ordinate representation, and are given by the columns of the matrix* $Z_{(N)s} = [Z_{(N)1s} \dots Z_{(N)g_0 s}]$, *where*

$$Z_{(N)ds} \;=\; W_d^T \mathcal{Z}_{(N)s} \;=\; (W_d \otimes Y_s)^T \mathcal{M}(\mathcal{C}). \tag{2.162}$$

3. The resolution transform matrix for the adjustment is calculated as

$$T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]} \;=\; \{(A \otimes C) + (N^{-1}B \otimes E)\}^{-1}(A \otimes C). \tag{2.163}$$

4. The collection $Z_{(N)} = \{Z_{(N)ds}\}$ *for* $d = 1, \dots, g_0$, $s = 1, \dots, v_0$, *are the canonical directions of the adjustment with canonical resolutions given by*

$$\lambda_{(N)ds} = \frac{\psi_d \phi_s}{\psi_d \phi_s + (1-\psi_d)(1-\phi_s)}. \tag{2.164}$$

5. The resolution ratio for the canonical directions is given by:

$$RR_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]}(Z_{(N)ds}) \;=\; \frac{\lambda_{ds}}{1-\lambda_{ds}} \;=\; \frac{\psi_d}{1-\psi_d} \times \frac{\phi_s}{1-\phi_s}. \tag{2.165}$$

Proof - Statements 1. and 2. follow immediately from Theorem 12, with the orthogonality between $[\mathcal{Z}_{(N)s}]$ and $[\mathcal{Z}_{(N)s'}]$ for $s \neq s'$ following from equation (2.153).

From the Bayes linear sufficiency of $\mathcal{S}_N(\mathcal{C})$ for $\mathcal{C}(N)$, as given by Theorem 8, we have that $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]} = T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$. By making use of equation (1.74), we have that $T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$ may be computed as

$$\begin{aligned} &T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]} = \\ &\quad Var(\mathcal{M}(\mathcal{C}))^{-1}Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C}))Var(\mathcal{S}_N(\mathcal{C}))^{-1}Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C})) \end{aligned} \tag{2.166}$$

From (2.92) and (2.93) we have

$$Var(\mathcal{M}(\mathcal{C})) \;=\; Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) \;=\; (A \otimes C). \tag{2.167}$$

Since A, C are positive definite then $(A \otimes C)$ is invertible and hence

$$Var(\mathcal{M}(\mathcal{C}))^{-1}Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) \;=\; I_{g_0 v_0}. \tag{2.168}$$

From (2.94) we have that

$$Var(\mathcal{S}_N(\mathcal{C})) \;=\; \{(A \otimes C) + (N^{-1}B \otimes E)\}. \tag{2.169}$$

Since A, C, N, B, E are positive definite, then we have invertibility of $\{(A \otimes C) + (N^{-1}B \otimes E)\}$. The representation of $T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$ given in Statement 3. thus follows. Having established Statement 2., Statement 4. requires that we show that $T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$'s matrix of eigenvalues is given by

$$\Lambda_{(N)} = (\Psi \otimes \Phi)\{(\Psi \otimes \Phi) + (I_{g_0} - \Psi) \otimes (I_{v_0} - \Phi)\}^{-1}, \tag{2.170}$$

with matrix of eigenvalues given by $Z = (W \otimes Y)$ where (Φ, Y), (Ψ, W) are the pairs of matrices of eigenvalues and eigenvectors respectively solving the two generalised eigenvalue problems given by equations (2.95) and (2.123). We show that they solve the equivalent generalised eigenvalue problem

$$(A \otimes C)Z = \{(A \otimes C) + (N^{-1}B \otimes E)\}Z\Lambda_{(N)}. \tag{2.171}$$

Consider

$$(A \otimes C)(W \otimes Y)\{(I_{g_0} - \Psi) \otimes (I_{v_0} - \Phi)\} = AW(I_{g_0} - \Psi) \otimes CY(I_{v_0} - \Phi) \tag{2.172}$$

and notice that from the solution of our two generalised problems (2.95) and (2.123), we may write $CY(I_{v_0} - \Phi) = EY\Phi$ and $AW(I_{g_0} - \Psi) = N^{-1}BW\Psi$. Substituting these into (2.172), we find that:

$$(A \otimes C)(W \otimes Y)\{(I_{g_0} - \Psi) \otimes (I_{v_0} - \Phi)\} = N^{-1}BW\Psi \otimes EY\Phi \tag{2.173}$$
$$= (N^{-1}B \otimes E)(W \otimes Y)(\Psi \otimes \Phi). \tag{2.174}$$

Adding $(A \otimes C)(W \otimes Y)(\Psi \otimes \Phi)$ to both sides gives:

$$(A \otimes C)(W \otimes Y)\{(\Psi \otimes \Phi) + (I_{g_0} - \Psi) \otimes (I_{v_0} - \Phi)\} = \{(A \otimes C) + (N^{-1}B \otimes E)\}(W \otimes Y)(\Psi \otimes \Phi). \tag{2.175}$$

Note that since $0 < \psi_d < 1\ \forall d = 1, \dots, g_0$ and $0 < \phi_s < 1\ \forall s = 1, \dots, v_0$, we may invert $\{(\Psi \otimes \Phi) + (I_{g_0} - \Psi) \otimes (I_{v_0} - \Phi)\}$ so that:

$$(A \otimes C)Z = \{(A \otimes C) + (N^{-1}B \otimes E)\}Z\Lambda_{(N)}, \tag{2.176}$$

where $Z = (W \otimes Y)$ and $\Lambda_{(N)} = (\Psi \otimes \Phi)\{(\Psi \otimes \Phi) + (I_{g_0} - \Psi) \otimes (I_{v_0} - \Phi)\}^{-1}$. Statement 5. follows immediately by considering $\Lambda_{(N)}\{I_{g_0 v_0} - \Lambda_{(N)}\}^{-1}$. □

This theorem tells us many things about our adjustment and has been published as Shaw & Goldstein (1999). Most importantly, it illustrates how we may simplify the design problem. There is no requirement of balance in the theorem. Thus, to compare the benefits of different choices of design where our design choice involves sample size selection, we reduce the $v_0 g_0 \times v_0 g_0$ problem into one problem of size $v_0 \times v_0$ and another of size $g_0 \times g_0$. Not only is this a great advantage computationally, but the two problems also have interpretable forms. The $v_0 \times v_0$ problem consists of finding the underlying canonical variable structure; the $g_0 \times g_0$ of finding the underlying canonical group structure. Notice that beyond being positive definite, no other symmetry requirements are placed on either A or B.

Observe that if $\phi_s > \phi_{s'}$, then $\lambda_{ds} > \lambda_{ds'}$. Likewise, if $\psi_d > \psi_{d'}$, then $\lambda_{ds} > \lambda_{d's}$. Changing the sample size only affects W and Ψ, so that the impact of changing the sample size may be easily seen and involves calculations only over $g_0 \times g_0$ problems. Thus, we can see how we may choose the sample sizes to optimise many different design criteria in terms of quantities of interest in the $\mathcal{M}(\mathcal{C})$; for example we may choose N's to learn about the most important group contrasts. In this case, we may be interested in learning about quantities of the form $(H_j^{g_0} \otimes \hat{Y})^T\mathcal{M}(\mathcal{C})$. The $H_j^{g_0}$ are the columns of the transpose of the $g_0 \times g_0$ Helmert matrix, see Searle (1982; p71) for further details. Then,

$$Cov((H_j^{g_0} \otimes \hat{Y})^T\mathcal{M}(\mathcal{C}), Z_{(N)ds}) = (H_j^{g_0} \otimes \hat{Y})^T Var(\mathcal{M}(\mathcal{C}))(W_d \otimes Y_s) \tag{2.177}$$
$$= \{(H_j^{g_0})^T A W_d\}\{\hat{Y}^T C Y_s\}, \tag{2.178}$$

so that, through use of equation (1.62), we have

$$Var_{\mathcal{C}(N)}((H_j^{g_0} \otimes \hat{Y})^T\mathcal{M}(\mathcal{C})) =$$
$$\sum_{d=1}^{g_0}\sum_{s=1}^{v_0}(1 - \lambda_{(N)ds})Cov((H_j^{g_0} \otimes \hat{Y})^T\mathcal{M}(\mathcal{C}), Z_{(N)ds})^2 \tag{2.179}$$
$$= \sum_{s=1}^{v_0}\{\hat{Y}^T C Y_s\}^2 \sum_{d=1}^{g_0}(1 - \lambda_{(N)ds})\{(H_j^{g_0})^T A W_d\}^2. \tag{2.180}$$

Now, observe that

$$Cov((D^{-1}H_j^{g_0})^T\mathcal{Z}_{(N)s}, Z_{(N)ds}) = \{(H_j^{g_0})^T A W_d\}\{Y_s^T C Y_s\} \tag{2.181}$$
$$= \{(H_j^{g_0})^T A W_d\}, \tag{2.182}$$

so that

$$Var_{\mathcal{C}(N)}((H_j^{g_0} \otimes \hat{Y})^T\mathcal{M}(\mathcal{C})) =$$
$$\sum_{s=1}^{v_0}\{\hat{Y}^T C Y_s\}^2 Var_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]}((D^{-1}H_j^{g_0})^T\mathcal{Z}_{(N)s}), \tag{2.183}$$

where $Var_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]}((D^{-1}H_j^{g_0})^T\mathcal{Z}_{(N)s})$, for any $\mathcal{X} \in [\mathcal{Z}_{(N)s}]$, is as given by equation (2.154). Thus, if we want to minimise the adjusted variance of $(H_j^{g_0} \otimes \hat{Y})^T\mathcal{M}(\mathcal{C})$, we would choose N to minimise equation (2.183); the dependence upon N is completely restricted to the $g_0 \times g_0$ variance, $Var_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]}((D^{-1}H_j^{g_0})^T\mathcal{Z}_{(N)s})$.

Notice also that we can easily assess the sensitivity of the design to the proportionality parameters in a similar way by merely looking at the resulting impact on W

and Ψ. Thus, we can examine the effect on the canonical directions and resolutions by treating the generalised problem given by equation (2.123) as a matrix perturbation problem. Techniques for handling such problems may be found in Stewart & Sun (1990).

Suppose that we want to compare the adjustment of $[\mathcal{M}(\mathcal{C})]$ under two different sample sizes. In the first instance, our sample sizes are given by the matrix N and in the second case by the matrix R where $R = \theta N$ for some positive constant θ. The two adjustments provide the same qualitative information and the quantitative information has a similar relationship to that between different sample sizes for the adjustment of a second-order exchangeable sequence as the following corollary reveals.

Corollary 3 *If $R = \theta N$ and $Z_{(N)ds} = (W_d \otimes Y_s)^T \mathcal{M}(\mathcal{C})$ is a canonical direction of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$ with corresponding canonical resolution $\lambda_{(N)ds}$, then $(W_d \otimes Y_s)^T \mathcal{M}(\mathcal{C})$ is a canonical direction of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(R)$ with corresponding canonical resolution*

$$\lambda_{(R)ds} = \frac{\theta \lambda_{(N)ds}}{(\theta - 1)\lambda_{(N)ds} + 1}. \qquad (2.184)$$

For less related samples, we could view changing the sample size as a perturbation problem of the same equation as that for discussing the proportionality parameters. Thus, for example, we could investigate what happens if we move away from the balanced design, for the balanced design has a particularly desirable form as it reproduces the elegant features of the adjustment of exchangeable vectors shown in Goldstein & Wooff (1998). We have the following corollary.

Corollary 4 *If $N = nI_{g_0}$, then the canonical directions are the same for all n, and if $\lambda_{(1)ds} = \psi_{(1)d}\phi_s / \{\psi_{(1)d}\phi_s + (1 - \psi_{(1)d})(1 - \phi_s)\}$ are the canonical resolutions for a sample of size $n = 1$, so $\Psi_{(1)}$ solves $AW = (A + B)W\Psi_{(1)}$, then the canonical resolutions and resolution ratio for general n are given by:*

$$\lambda_{(n)ds} = \frac{n\lambda_{(1)ds}}{(n-1)\lambda_{(1)ds} + 1}; \quad RR_{\mathcal{C}(N)}(Z_{(n)ds}) = n \times \frac{\psi_{(1)d}}{1 - \psi_{(1)d}} \times \frac{\phi_s}{1 - \phi_s}. \qquad (2.185)$$

Thus, in a similar vein, we may use (2.185) to simplify design problems for choosing sample sizes to achieve specific variance reductions. For example, we have the following corollary:

Corollary 5 *The sample size n in each group required to achieve a proportionate variance reduction of κ for $Z_{(nI_{g_0})ds}$, is $n \geq \{\kappa/(1-\kappa)\}\{(1 - \lambda_{ds(1)})/\lambda_{ds(1)}\}$. If* *the minimal canonical resolution for $n = 1$ is $\lambda_{min} = \psi_{min}\phi_{min}/\{\psi_{min}\phi_{min} + (1 - \psi_{min})(1 - \phi_{min})\}$, then a sample size of $\{\kappa/(1 - \kappa)\}\{(1 - \lambda_{min})/\lambda_{min}\} = \{\kappa/(1 - \kappa)\}\{(1 - \psi_{min})/\psi_{min}\}\{(1 - \phi_{min})/\phi_{min}\}$, rounded up, in each group is the minimum sample which is sufficient to achieve a proportionate variance reduction of κ for every element of $[\mathcal{M}(\mathcal{C})]$.*

The advantage of this corollary is that it provides an immediate upper bound to the sample sizes required to reduce the variance of any $\mathcal{M}(\mathcal{X}) \in [\mathcal{M}(\mathcal{C})]$ for any choice of N, which may have advantageous benefits upon budgeting.

2.4.6 The adjustment of $[\mathcal{M}(\mathcal{C})]$ for the examiner example

From Theorem 13, we know that the canonical structure of the adjustment may be found by solving the underlying canonical variable and group problems as accomplished in Subsections 2.4.2 and 2.4.4. The collection $Z_{(nI_2)} = \{Z_{(nI_2)ds}\}$ (for example, see equations (2.157) and (2.158)) is the collection of canonical resolutions for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_2)$; the corresponding collection of canonical resolutions being $\Lambda_{(nI_2)} = \{\lambda_{(nI_2)ds}\}$ (for example, see equations (2.159) and (2.160)). If $\gamma > 0$, then quantities proportional to $Z_{(nI_2)11}$ have the largest resolution, that resolution being $\lambda_{(nI_2)11}$. Similarly, quantities proportional to $Z_{(nI_2)22}$ have the smallest resolution, that resolution being $\lambda_{(nI_2)22}$. Observe that the observation that the $Z_{(nI_2)}ds$'s do not depend upon the sample size is explained by Corollary 4, since we have a balanced design. Similarly, we may use Corollary 5 to find the minimal sample size required in each group to obtain a proportionate variance reduction of κ for every element of $[\mathcal{M}(\mathcal{C})]$. As noted above, when $\gamma > 0$, the smallest resolution corresponds to $\lambda_{(nI_2)22}$, with $\phi_{min} = (1/7)$ and $\psi_{min} = \{(1-\gamma)/(2-\gamma)\}$ (as can be easily seen by setting $n = 1$ in the matrix Ψ of equation (2.156)). Thus, the size of n required is $6\kappa/\{(1-\kappa)(1-\gamma)\}$, rounded up.

2.5 A link with analysis of variance models

Recall that a possible modelling interpretation of our beliefs is that of scaled exchangeability as given by Lemma 4 and Lemma 5. Now, suppose that we judge that the $\mathcal{A}_g$s are second-order exchangeable, so that

$$E(\mathcal{A}_g) = m \, \forall g; \tag{2.186}$$

$$Var(\mathcal{A}_g) + m^2 = \alpha \, \forall g; \tag{2.187}$$

$$Cov(\mathcal{A}_g, \mathcal{A}_h) + m^2 = \gamma \, \forall g \neq h. \tag{2.188}$$

The matrix A may then be written as

$$A = \alpha I_{g_0} + \gamma(J_{g_0} - I_{g_0}). \tag{2.189}$$

Observe that by taking $\alpha = 1$ then this is precisely the specification that the examiner adopts. Notice that if we judge that the $\mathcal{A}_g$s are second-order exchangeable, then this also states that we judge that $\mathcal{M}(X_g)$ is second-order exchangeable over groups. We shall assume that the g_0 groups assessed here are a sample from a potentially infinite sequence of possible groups. It should be noted that this infinite assumption is not needed and that we may proceed by using the finite version of the representation theorem of Goldstein (1986a). Since the $\mathcal{M}(X_g)$ are an infinitely exchangeable sequence then we may use the representation theorem for infinitely exchangeable sequences to write

$$\mathcal{M}(\mathcal{C}_g) = \mathcal{M}(\mathcal{C}_.) + \mathcal{R}_g(\mathcal{C}_.). \tag{2.190}$$

where $[\mathcal{M}(\mathcal{C}_.)] \perp [\mathcal{R}_g(\mathcal{C}_.)]$ for each g. Our full model is then

$$X_{gi} = \mathcal{M}(\mathcal{C}_.) + \mathcal{R}_g(\mathcal{C}_.) + \mathcal{R}_i(X_g), \tag{2.191}$$

which may be compared to a one-way multivariate analysis of variance (MANOVA) model; see, for example, Press (1989, Chapter VI). In a more traditional framework, we could perceive equation (2.191) as being a random effects model (see Searle *et al.* (1992)). Notice that we have

$$Var(\mathcal{R}_g(\mathcal{C}_.)) = (\alpha - \gamma)C, \tag{2.192}$$

so that the hypothesis $Var(\mathcal{R}_g(\mathcal{C}_.)) = 0$ amounts to $\alpha = \gamma$, or that the relationship between individuals in different groups is identical to that of different individuals in the same group; the labelling of groups is superfluous. If we were particularly interested in learning about this model and investigating this hypothesis, then we could use an experiment to learn about the variances, $Var(\mathcal{R}_g(\mathcal{C}_.))$. This would involve the specification of fourth-order moments and is beyond the scope of this thesis; for more details on the adjustment of covariance matrices see Wilkinson (1995).

In the case of $\mathcal{A}_g$s exchangeable, then the underlying group problem, for a balanced design, has just two distinct canonical resolutions. The first, of multiplicity one, corresponds to canonical directions proportional to $H^{g_0}_{g_0} = (1/g_0)1^T_{g_0}$. The second, with multiplicity $g_0 - 1$ corresponds to directions in the space spanned by the collection of directions, $H^{g_0}_d$ for $d = 1, \ldots, g_0 - 1$, where

$$H^{g_0}_d = \frac{1}{\sqrt{d(d+1)}}[1^T_d \; -d\; 0 \ldots 0]^T. \tag{2.193}$$

Thus, we partition the adjusted variance for any quantity $\mathcal{M}(\mathcal{X}) \in [\mathcal{M}(\mathcal{C})]$ into two uncorrelated parts, the first corresponding to the mean across groups; the second, of dimension $g_0 - 1$, corresponds to the differences between groups. This has immediate parallels with the types of variance partition performed in a classical analysis of variance framework. The effect of moving away from the exchangeability of the $\mathcal{A}_g$ is to rotate these directions to the general directions given by the columns of the matrix W.

Notice that the matrix H^{g_0} is a column permutation of the transpose of the Helmert matrix. We choose this ordering, as opposed to that used in Goldstein & Wooff (1997) to maintain consistency with the useage in Definition 22 where the column ordering follows to correspond to the size of the underlying group residual canonical group resolutions. Notice in this present example, the ordering of the canonical resolutions may depend upon the choice of α and γ as is illustrated by equation (2.156).

Moreover, if we judge further that the variables are exchangeable, as we did in the examiner's problem, then Y_s is proportional to $H^{v_0}_s$, where

$$H^{v_0}_s = \begin{cases} (1/v_0)1^T_{v_0} & \text{if } s = v_0; \\ \frac{1}{\sqrt{s(s+1)}}[1^T_s \; -s\; 0 \ldots 0]^T & \text{otherwise.} \end{cases} \tag{2.194}$$

The full set of canonical directions are thus of the form $(H^{g_0}_d \otimes H^{v_0}_s)^T \mathcal{M}(\mathcal{C})$. This kind of structure for the canonical directions is found, and extensively addressed in the context of factorial designs in Goldstein & Wooff (1997).

2.6 Predictive adjustment: predicting in already observed groups

We now want to examine the impact of the observation of the sample of n_g individuals in the gth group for $g = 1, \ldots, g_0$ for predicting the values for a further collection of r_h individuals in the hth group for $h = 1, \ldots, h_0$, where $h_0 \leq g_0$. Thus, we are presently interested in predicting in groups where we have already observed some individuals, and we assume (without loss of generality) that the groups have been labelled so that we predict in the first g_0 groups. We collect the prediction sample sizes together as the $h_0 \times h_0$ diagonal matrix R with (h,h)th entry $(R)_{hh} = r_h$.

For example, in our running examiner example, the examiner may collect samples of the marking from the examiners at different stages of the marking process. He may take an early sample to check that they are marking proficiently and a later sample to monitor their marking again, perhaps to check that they have not become blase. Alternatively, the prediction could be for the remaining papers they have to mark.

Let $\mathcal{C}(n_h; r_h) = \cup_{i=n_h+1}^{n_h+r_h} \mathcal{C}_{hi}$, denote the collection of individuals in the hth group that we would like to predict the values for, having seen the first n_h individuals in that group as part of our sample, $\mathcal{C}(N)$. We denote the complete collection of individuals we would like to predict the values for as $\mathcal{C}(N;\ R) = \cup_{h=1}^{h_0} \mathcal{C}(n_h; r_h)$. For $\mathcal{X}_h \in \langle \mathcal{C}_h \rangle$, let

$$\mathcal{S}_{(n_h;r_h)}(\mathcal{X}_h) \quad = \quad \frac{1}{r_h} \sum_{i=n_h+1}^{n_h+r_h} \mathcal{X}_{hi}, \tag{2.195}$$

and let $\mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h) = \{\mathcal{S}_{(n_h;r_h)}(X_{h1}), \ldots, \mathcal{S}_{(n_h;r_h)}(X_{hv_0})\}$ be the collection of averages of the r_h future observations in the hth group. Initially, we shall restrict attention to adjusting the complete collection of averages of future observations, $\mathcal{S}_{(N;R)}(\mathcal{C}) = \cup_{h=1}^{h_0} \mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h)$, before showing that this study is sufficient to give us the canonical directions with non-zero canonical resolutions for the adjustment of $[\mathcal{C}(N;R)]$. In our usual vector notation, we have $\mathcal{C}_h(n_h;r_h) = [\mathcal{C}_{hn_{h+1}}^T \ldots \mathcal{C}_{hn_h+r_h}^T]^T$; $\mathcal{C}(N;R) = [\mathcal{C}_1(n_1;r_1)^T \ldots \mathcal{C}_{h_0}(n_{h_0};r_{h_0})^T]^T$; $\mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h) = [\mathcal{S}_{(n_h;r_h)}(X_{h1}) \ldots \mathcal{S}_{(n_h;r_h)}(X_{hv_0})]^T$ and $\mathcal{S}_{(N;R)}(\mathcal{C}) = [\mathcal{S}_{(n_1;r_1)}(\mathcal{C}_1)^T \ldots \mathcal{S}_{(n_{h_0};r_{h_0})}(\mathcal{C}_{h_0})^T]^T$.

From the specifications given by equations (2.92) - (2.94), we may derive the specifications for the quantities mentioned above. They are given in the lemma below.

Lemma 9 *The second-order relationships between the $\mathcal{S}_{n_g}(\mathcal{C}_g)$s, the $\mathcal{C}_h(n_h;r_h)$s, and the $\mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h)$s may be expressed, $\forall g = 1, \ldots, g_0$; $\forall h, h' = 1, \ldots, h_0$; $\forall i = 1, \ldots r_h$,* *as*

$$
\begin{aligned}
Cov(\mathcal{S}_{n_g}(\mathcal{C}_g), \mathcal{C}_{hn_h+i}) &= \alpha_{gh} C; & (2.196)\\
Cov(\mathcal{S}_{n_g}(\mathcal{C}_g), \mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h)) &= \alpha_{gh} C; & (2.197)\\
Cov(\mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h), \mathcal{S}_{(n_{h'};r_{h'})}(\mathcal{C}_{h'})) &= \begin{cases} \alpha_{hh}C + \frac{1}{r_h}\beta_h E & \text{if } h = h'; \\ \alpha_{hh'}C & \text{if } h \neq h'; \end{cases} & (2.198)\\
Cov(\mathcal{S}_{(n_{h'};r_{h'})}(\mathcal{C}_{h'}), \mathcal{C}_{hn_h+i}) &= \begin{cases} \alpha_{hh}C + \frac{1}{r_h}\beta_h E & \text{if } h' = h; \\ \alpha_{h'h}C & \text{if } h' \neq h. \end{cases} & (2.199)
\end{aligned}
$$

Denote by A_{pq} the $p \times q$ matrix with (p_1, q_1)th entry $(A_{pq})_{p_1q_1} = \alpha_{p_1q_1}$ and let B_{pp} be the $p \times p$ diagonal matrix with (p_1, p_1)th entry $(B_{pp})_{p_1p_1} = \beta_{p_1}$. Gathering the $\mathcal{S}_{n_g}(\mathcal{C}_g)$s together into the $g_0v_0 \times 1$ vector $\mathcal{S}_N(\mathcal{C})$ and the $\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)$s as the $h_0v_0 \times 1$ vector $\mathcal{S}_{(N;R)}(\mathcal{C})$, then we have the following lemma to express equations (2.94), (2.197) and (2.198).

Lemma 10 *The second-order relationships between $\mathcal{S}_N(\mathcal{C})$ and $\mathcal{S}_{(N;R)}(\mathcal{C})$ may be expressed as:*

$$
\begin{aligned}
Var(\mathcal{S}_N(\mathcal{C})) &= (A_{g_0g_0} \otimes C) + (N^{-1}B_{g_0g_0} \otimes E); & (2.200)\\
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{S}_{(N;R)}(\mathcal{C})) &= A_{g_0h_0} \otimes C; & (2.201)\\
Var(\mathcal{S}_{(N;R)}(\mathcal{C})) &= (A_{h_0h_0} \otimes C) + (R^{-1}B_{h_0h_0} \otimes E). & (2.202)
\end{aligned}
$$

2.6.1 Analysis of groups and variables

Once more, we consider separately the analysis of a variable problem and related group problem. We shall make use of the underlying canonical variable structure defined in Definition 12. We motivate the use of it as follows. Consider the problem of examining the effect of a sample of n_g individuals from the gth group for predicting the values for a further collection of r_g individuals drawn from that group. Since we are dealing with a single group, then the individuals that we wish to predict are second-order exchangeable with those in the sample. This is precisely the problem tackled in Section 8 of Goldstein & Wooff (1998), and reproduced in Section 1.9 in this thesis. They showed that the canonical directions with non-zero resolutions lay in the space $[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)]$, and had the same, up to a scale factor to ensure a prior variance of one, co-ordinate representation as the resolutions for the adjustment over the population structure, $[\mathcal{M}(\mathcal{C}_g)]$. Thus, predictive adjustment shares the same qualitative features as adjustment over the population structure. Denoting the resolution transform for the adjustment of $[\mathcal{C}(n_g; r_g)]$ by $\mathcal{S}_{n_g}(\mathcal{C}_g)$ as $T_{[\mathcal{C}(n_g;r_g)/\mathcal{S}_{n_g}(\mathcal{C}_g)]}$, we rewrite Theorem 4 of Goldstein & Wooff (1998), reproduced as Theorem 6 in this thesis, as:

Theorem 14 *The resolution transform matrix, $T_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]}$ is calculated as*

$$T_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]} = \{\alpha_{gg}C + (1/r_g)\beta_g E\}^{-1}(\alpha_{gg}C)\{\alpha_{gg}C + (1/n_g)\beta_g E\}^{-1}(\alpha_{gg}C). \quad (2.203)$$

For each $s = 1, \ldots, v_0$, the canonical directions for the adjustment of $[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)]$ by $\mathcal{C}_g(n_g)$ are given by

$$Z_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} = \sqrt{\frac{r_g\phi_s}{r_g\alpha_{gg}\phi_s + \beta_g(1-\phi_s)}}Y_s^T\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g), \quad (2.204)$$

with the corresponding canonical resolutions given by

$$\lambda_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} = \lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{S}_{r_g}(\mathcal{C}_g)]s}\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}. \quad (2.205)$$

The collection

$$\Lambda_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]} = \{\lambda_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]1}, \ldots, \lambda_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]v_0}\} \quad (2.206)$$

is the collection of non-zero resolutions for the adjustment of $[\mathcal{C}(n_g;r_g)]$ by $\mathcal{C}_g(n_g)$; the corresponding directions being given by

$$Z_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]} = \{Z_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]1}, \ldots, Z_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]v_0}\}. \quad (2.207)$$

$\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(m_g)]s}$ is as given in equation (2.98). Recall that in Theorem 11 we noted that for our specifications, the adjustment of the population structure $\mathcal{M}(\mathcal{C}_g)$ had the same qualitative features in each group. Hence, as we can see from equations (2.204) and (2.205), we have that prediction in each group has the same qualitative features with simple modifications for the canonical resolutions summarising the quantitative information.

Definition 14 *The sth underlying canonical predictive group directions are defined as the columns of the matrix $W_s = [W_{1s} \ldots W_{h_0 s}]$ solving the generalised eigenvalue problem*

$$(\phi_s A_{h_0 g_0})\{\phi_s A_{g_0 g_0} + (1-\phi_s)N^{-1}B_{g_0 g_0}\}^{-1}(\phi_s A_{g_0 h_0})W_s = \{\phi_s A_{h_0 h_0} + (1-\phi_s)R^{-1}B_{h_0 h_0}\}W_s\Lambda_{(N;R)s}, \quad (2.208)$$

where $\Lambda_{(N;R)s} = diag(\lambda_{(N;R)1s}, \ldots, \lambda_{(N;R)h_0 s})$ is the matrix of eigenvalues. W_s is chosen so that $W_s^T\{A_{h_0h_0} + ((1/\phi_s)-1)R^{-1}B_{h_0h_0}\}W_s = I_{h_0}$, $W_s^T A_{h_0 g_0}\{A_{g_0g_0} + ((1/\phi_s)-1)N^{-1}B_{g_0g_0}\}^{-1}A_{g_0h_0}W_s\Lambda^{-1}_{(N;R)s} = I_{h_0}$. The ordered eigenvalues $1 > \lambda_{(N;R)1s} \geq \ldots \geq \lambda_{(N;R)h_0 s} > 0$ are termed the sth underlying canonical predictive group resolutions.

To motivate this definition, for each $s = 1, \dots v_0$, form the collection:

$$\mathcal{Z}_{(N;R)s} = \{Z_{[\mathcal{S}_{(n_1;r_1)}(\mathcal{C}_1)/\mathcal{C}_1(n_1)]s}, \dots, Z_{[\mathcal{S}_{(n_{h_0};r_{h_0})}(\mathcal{C}_{h_0})/\mathcal{C}_{h_0}(n_{h_0})]s}\}. \quad (2.209)$$

The collection of linear combinations, $\sum_{g=1}^{h_0} a_g \mathcal{Z}_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ of the elements of $\mathcal{Z}_{(N;R)s}$ is denoted by $\langle \mathcal{Z}_{(N;R)s} \rangle$. We consider using the sample to learn (separately) about each of the collections $\mathcal{Z}_{(N;R)s}$. The sample means, $\mathcal{S}_N(\mathcal{C})$, are again sufficient for this adjustment.

Thus, we collect together the directions in each group that correspond to the sth largest resolution for the adjustment of $[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)]$ by $\mathcal{C}_g(n_g)$. It should be noted that $\mathcal{Z}_{(N;R)s}$ thus consists of the directions that have, up to a scale factor, the same co-ordinate representation in each group. This then makes clear what we do in the case where we have two or more directions with the same resolution.

Letting D_{Rs} be the $h_0 \times h_0$ diagonal matrix with (h, h)th entry

$$(D_{Rs})_{hh} = \sqrt{\frac{r_h \phi_s}{r_h \alpha_{hh} \phi_s + \beta_h (1 - \phi_s)}}, \quad (2.210)$$

then in our usual vector notation, we may equally express equation (2.209) as

$$\mathcal{Z}_{(N;R)s} = D_{Rs}(I_{h_0} \otimes Y_s^T) \mathcal{S}_{(N;R)}(\mathcal{C}). \quad (2.211)$$

Notice the similarity of equation (2.211) with equation (2.126).

Lemma 11 *The second-order relationships between $\mathcal{Z}_{(N;R)s}$ and $\mathcal{S}_N(\mathcal{C})$ may be expressed as follows*

$$\begin{aligned}
Var(\mathcal{Z}_{(N;R)s}) &= (1/\phi_s) D_{Rs} \{\phi_s A_{h_0 h_0} + (1 - \phi_s) R^{-1} B_{h_0 h_0}\} D_{Rs}; && (2.212)\\
Cov(\mathcal{Z}_{(N;R)s}, \mathcal{S}_N(\mathcal{C})) &= (1/\phi_s) D_{Rs} [\epsilon_{v_0 s}^T \otimes \{\phi_s A_{h_0 g_0}\}] \times \\
&\quad [(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})]^{-1}; && (2.213)\\
Var(\mathcal{S}_N(\mathcal{C})) &= [I_{g_0} \otimes Y^{-1}(C + E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0} \{\phi_s A_{g_0 g_0} + (1 - \phi_s) N^{-1} B_{g_0 g_0}\}] \times \\
&\quad [(I_{g_0} \otimes Y_1) \dots (I_{g_0} \otimes Y_{v_0})]^{-1}; && (2.214)\\
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N;R)s}) &= [I_{g_0} \otimes Y^{-1}(C + E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\epsilon_{v_0 s} \otimes \{\phi_s A_{g_0 h_0}\}] D_{Rs}. && (2.215)
\end{aligned}$$

Proof - Equation (2.212) follows from equation (2.202) since

$$\begin{aligned}
Var(\mathcal{Z}_{(N;R)s}) &= D_{Rs}(I_{h_0} \otimes Y_s^T)Var(\mathcal{S}_{(N;R)}(\mathcal{C}))(I_{h_0} \otimes Y_s)D_{Rs} && (2.216)\\
&= D_{Rs}\{(A_{h_0h_0} \otimes Y_s^T CY_s) + (R^{-1}B_{h_0h_0} \otimes Y_s^T EY_s)\}D_{Rs} && (2.217)\\
&= D_{Rs}\{(A_{h_0h_0} \otimes 1) + (R^{-1}B_{h_0h_0} \otimes ((1/\phi_s) - 1))\}D_{Rs} && (2.218)
\end{aligned}$$

where equation (2.218) follows by the choice of Y in Definition 12. Equation (2.213) follows from the transpose of equation (2.201) since

$$\begin{aligned}
Cov(\mathcal{Z}_{(N;R)s}, \mathcal{S}_N(\mathcal{C})) &= D_{Rs}(I_{h_0} \otimes Y_s^T)Cov(\mathcal{S}_{(N;R)}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) && (2.219)\\
&= D_{Rs}(A_{h_0g_0} \otimes Y_s^T C) && (2.220)\\
&= D_{Rs}(A_{h_0g_0} \otimes Y_s^T C)[(I_{g_0} \otimes Y_1)\dots(I_{g_0} \otimes Y_{v_0})][(I_{g_0} \otimes Y_1)\dots(I_{g_0} \otimes Y_{v_0})]^{-1} && (2.221)\\
&= D_{Rs}[(A_{h_0g_0} \otimes Y_s^T CY_1)\dots(A_{h_0g_0} \otimes Y_s^T CY_{v_0})][(I_{g_0} \otimes Y_1)\dots(I_{g_0} \otimes Y_{v_0})]^{-1} && (2.222)\\
&= D_{Rs}[(A_{h_0g_0} \otimes \delta_{1s})\dots(A_{h_0g_0} \otimes \delta_{1v_0})][(I_{g_0} \otimes Y_1)\dots(I_{g_0} \otimes Y_{v_0})]^{-1} && (2.223)\\
&= D_{Rs}[\epsilon_{v_0s}^T \otimes A_{h_0g_0}][(I_{g_0} \otimes Y_1)\dots(I_{g_0} \otimes Y_{v_0})]^{-1}, && (2.224)
\end{aligned}$$

where equation (2.223) also follows by the choice of Y in Definition 12. Equation (2.214) was derived in Lemma 8. To obtain equation (2.213) notice that

$$\begin{aligned}
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N;R)s}) &= Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{S}_{(N;R)}(\mathcal{C}))(I_{h_0} \otimes Y_s)D_{Rs} && (2.225)\\
&= (A_{g_0h_0} \otimes C)(I_{h_0} \otimes Y_s)D_{Rs}. && (2.226)
\end{aligned}$$

We apply Lemma 7 to equation (2.201) with $S = A_{g_0h_0}$, $X_s = I_{h_0}$ for each $s = 1, \dots, v_0$. Substituting into equation (2.226) gives

$$\begin{aligned}
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N;R)s}) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_01}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{t=1}^{v_0}\{\phi_t A_{g_0h_0}\}][(I_{h_0} \otimes Y_1)\dots(I_{h_0} \otimes Y_{v_0})]^{-1}(I_{h_0} \otimes Y_s)D_{Rs} && (2.227)\\
&= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_01}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{t=1}^{v_0}\{\phi_t A_{g_0h_0}\}][\epsilon_{v_0s} \otimes I_{h_0}]D_{Rs} && (2.228)\\
&= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_01}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\epsilon_{v_0s} \otimes \{\phi_s A_{g_0h_0}\}]D_{Rs}. && (2.229)
\end{aligned}$$

□

Denote the resolution transform for the adjustment of $[\mathcal{Z}_{(N;R)s}]$ by $\mathcal{C}(N)$ as $T_{[\mathcal{Z}_{(N;R)s}/\mathcal{C}(N)]}$. We have the following theorem.

Theorem 15 *For the adjustment of $[\mathcal{Z}_{(N;R)s}]$ by $\mathcal{C}(N)$, the resolution transform matrix is calculated as*

$$T_{[\mathcal{Z}_{(N;R)s}/\mathcal{C}(N)]} = D_{Rs}^{-1}\{\phi_s A_{h_0h_0} + (1-\phi_s)R^{-1}B_{h_0h_0}\}^{-1}\{\phi_s A_{h_0g_0}\} \times \{\phi_s A_{g_0g_0} + (1-\phi_s)N^{-1}B_{g_0g_0}\}^{-1}\{\phi_s A_{g_0h_0}\}D_{Rs}. \quad (2.230)$$

The canonical directions are given by $(D_{Rs}^{-1}W_{ds})^T\mathcal{Z}_{(N;R)s}$ for each $d = 1,\dots,h_0$, with corresponding canonical resolutions given by $\lambda_{(N;R)ds}$. W_{ds} is the (d,s)th underlying canonical predictive group direction as given in Definition 14; $\lambda_{(N;R)ds}$ the corresponding (d,s)th underlying canonical predictive group resolution.

Proof - From Theorem 9, $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting $\mathcal{C}(N;R)$. For each s, we have $[\mathcal{Z}_{(N;R)s}] \subset [\mathcal{C}(N;R)]$, so that $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting $\mathcal{Z}_{(N;R)s}$. Thus, $T_{[\mathcal{Z}_{(N;R)s}/\mathcal{C}(N)]} = T_{[\mathcal{Z}_{(N;R)s}/\mathcal{S}_N(\mathcal{C})]}$, where $T_{[\mathcal{Z}_{(N;R)s}/\mathcal{S}_N(\mathcal{C})]}$ is the resolution transform for the adjustment of $[\mathcal{Z}_{(N;R)s}]$ by $\mathcal{S}_N(\mathcal{C})$. From equation (1.74), $T_{[\mathcal{Z}_{(N;R)s}/\mathcal{S}_N(\mathcal{C})]}$ may be computed as

$$T_{[\mathcal{Z}_{(N;R)s}/\mathcal{S}_N(\mathcal{C})]} = \{Var(\mathcal{Z}_{(N;R)s})\}^{-1}Cov(\mathcal{Z}_{(N;R)s}, \mathcal{S}_N(\mathcal{C})) \times \{Var(\mathcal{S}_N(\mathcal{C}))\}^{-1}Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N;R)s}). \quad (2.231)$$

Hence, by inverting equation (2.212), post multiplying by equation (2.213) and then the inversion of (2.214) and finally by (2.215) we may achieve equation (2.230). Substituting equation (2.208) into equation (2.230) gives

$$T_{[\mathcal{Z}_{(N;R)s}/\mathcal{S}_N(\mathcal{C})]} = D_{Rs}^{-1}W_s\Lambda_{(N;R)s}W_s^{-1}D_{Rs}, \quad (2.232)$$

so that

$$T_{[\mathcal{Z}_{(N;R)s}/\mathcal{S}_N(\mathcal{C})]}(D_{Rs}^{-1}W_s) = \Lambda_{(N;R)s}W_s^{-1}D_{Rs}. \quad (2.233)$$

Hence, $\Lambda_{(N;R)s}$ is the matrix whose diagonal elements are the canonical resolutions of the adjustment. To confirm that the $(D_{Rs}^{-1}W_{ds})^T\mathcal{Z}_{(N;R)s}$ for each $d = 1,\dots h_0$ are the canonical resolutions, we verify that they are mutually uncorrelated with prior variance one. Notice that by using equation (2.211), we may write

$$(D_{Rs}^{-1}W_{ds})^T\mathcal{Z}_{(N;R)s} = (W_{ds}^T \otimes Y_s^T)\mathcal{S}_{(N;R)}(\mathcal{C}) \quad (2.234)$$

so that

$$\begin{aligned}
&Cov((D_{Rs}^{-1}W_{ds})^T\mathcal{Z}_{(N;R)s}, (D_{Rs}^{-1}W_{d's})^T\mathcal{Z}_{(N;R)s}) \\
&= (W_{ds}^T \otimes Y_s^T)Var(\mathcal{S}_{(N;R)}(\mathcal{C}))(W_{d's} \otimes Y_s) && (2.235)\\
&= \{(W_{ds}^T A_{h_0h_0}W_{d's} \otimes Y_s^T C Y_s) + (W_{ds}^T R^{-1}B_{h_0h_0}W_{d's} \otimes Y_s^T E Y_s)\} && (2.236)\\
&= W_{ds}^T\{A_{h_0h_0} + ((1/\phi_s) - 1)R^{-1}B_{h_0h_0}\}W_{d's} && (2.237)\\
&= \delta_{dd'}. && (2.238)
\end{aligned}$$

Equation (2.237) follows from the choice of Y in Definition 12, and equation (2.238) follows from the choice of W_s in Definition 14. □

Theorem 15 should be compared with Theorem 12. Notice the similarities in the construction of the $\mathcal{Z}_{(N)s}$ and $\mathcal{Z}_{(N;R)s}$ from the same underlying canonical variable problem, but also notice the difference. The adjustment of each $\mathcal{Z}_{(N)s}$ is qualitatively the same for each s, whilst the adjustment of each $\mathcal{Z}_{(N;R)s}$ does depend upon ϕ_s, although the canonical predictive group problems are of similar forms for each ϕ_s.

2.6.2 Adjustment of the full collection

We now consider the adjustment of the full collection, $[\mathcal{C}(N;R)]$. We shall show that in a similar way to inference about the population means, $\mathcal{M}(\mathcal{C})$, the canonical variable and each canonical predictive group analysis completely determine the adjustment of the full collection $[\mathcal{C}(N;R)]$. To proceed, we find it easiest to reexpress equations (2.200) - (2.202) as the following lemma.

Lemma 12 *The relationships between $\mathcal{S}_{(N;R)}(\mathcal{C})$ and $\mathcal{S}_N(\mathcal{C})$ may be expressed as follows*

$$\begin{aligned}
Var(\mathcal{S}_{(N;R)}(\mathcal{C})) &= [I_{h_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{h_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{h_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0}\{\phi_s A_{h_0 h_0} + (1-\phi_s) R^{-1} B_{h_0 h_0} W_s\}] \times \\
&\quad [(W_1 \otimes Y_1) \ldots (W_{v_0} \otimes Y_{v_0})]^{-1}; \qquad (2.239)\\
Cov(\mathcal{S}_{(N;R)}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) &= [I_{h_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{h_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{h_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0} \phi_s A_{h_0 g_0}][(I_{g_0} \otimes Y_1) \ldots (I_{g_0} \otimes Y_{v_0})]^{-1}; \qquad (2.240)\\
Var(\mathcal{S}_N(\mathcal{C})) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0}\{\phi_s A_{g_0 g_0} + (1-\phi_s) N^{-1} B_{g_0 g_0}\}] \times \\
&\quad [(I_{g_0} \otimes Y_1) \ldots (I_{g_0} \otimes Y_{v_0})]^{-1}; \qquad (2.241)\\
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{S}_{(N;R)}(\mathcal{C})) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g_0} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0}\{\phi_s A_{g_0 h_0}\} W_s] \times \\
&\quad [(W_1 \otimes Y_1) \ldots (W_{v_0} \otimes Y_{v_0})]^{-1}. \qquad (2.242)
\end{aligned}$$

where W_s is the matrix whose columns are the sth underlying canonical predictive group directions.

Proof - Applying Lemma 7 to equation (2.202) with $S = A_{h_0h_0}$; $T = R^{-1}B_{h_0h_0}$; $X_s = W_s$ for each $s = 1,\dots,v_0$ and rearranging gives equation (2.239). Equation (2.240) may by derived from equation (2.201) by transposing, applying Lemma 7 with $S = A_{h_0g_0}$; $X_s = I_{g_0}$ for all $s = 1,\dots,v_0$, and then rearranging. Similar use of Lemma 7 with $S = A_{g_0g_0}$; $T = N^{-1}B_{g_0g_0}$; $X_s = I_{g_0}$ for all $s = 1,\dots,v_0$ will see equation (2.241) obtained from equation (2.200). Taking $S = A_{g_0h_0}$; $X_s = W_s$ for each $s = 1,\dots,v_0$ in Lemma 7 and rearranging will link equation (2.201) with equation (2.242). □

It should be emphasied that equation (2.242) is the transpose of equation (2.240). For convenience, we choose the two different representations given in Lemma 12.

The resolution transform for the adjustment of $[\mathcal{S}_{(N;R)}(\mathcal{C})]$ by $\mathcal{C}(N)$ is denoted by $T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{C}(N)]}$. We have the following theorem.

Theorem 16 *The adjustment of $[\mathcal{S}_{(N;R)}(\mathcal{C})]$ by $\mathcal{C}(N)$ satisfies the following properties*

1. There exist v_0 h_0-dimensional orthogonal subspaces, $[\mathcal{Z}_{(N;R)1}]$, ..., $[\mathcal{Z}_{(N;R)v_0}]$ of $[\mathcal{S}_{(N;R)}(\mathcal{C})]$. For each s,

$$\mathcal{Z}_{(N;R)s} = \{Z_{[\mathcal{S}_{(n_1;r_1)}(\mathcal{C}_1)/\mathcal{C}_1(n_1)]s}, \dots, Z_{[\mathcal{S}_{(n_{h_0};r_{h_0})}(\mathcal{C}_{h_0})/\mathcal{C}_{h_0}(n_{h_0})]s}\}. \tag{2.243}$$

2. The canonical directions for the adjustment in each $[\mathcal{Z}_{(N;R)s}]$ are given by the columns of the matrix $Z_{(N;R)s} = [Z_{(N;R)1s} \dots Z_{(N;R)h_0s}]$, where

$$Z_{(N;R)ds} = (D_{Rs}^{-1}W_{ds})^T\mathcal{Z}_{(N;R)s} = (W_{ds}\otimes Y_s)^T\mathcal{S}_{(N;R)}(\mathcal{C}). \tag{2.244}$$

3. The resolution transform matrix of the full collection is calculated as

$$\begin{aligned} T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{C}(N)]} &= \{(A_{h_0h_0}\otimes C) + (R^{-1}B_{h_0h_0}\otimes E)\}^{-1}(A_{h_0g_0}\otimes C) \\ &\quad \{(A_{g_0g_0}\otimes C) + (N^{-1}B_{g_0g_0})\}^{-1}(A_{g_0h_0}\otimes C) \qquad (2.245)\\ &= [(W_1\otimes Y_1)\dots(W_{v_0}\otimes Y_{v_0})][\oplus_{s=1}^{v_0} W_s^{-1}\{\phi_s A_{h_0h_0} + (1-\phi_s) \\ &\quad R^{-1}B_{h_0h_0}\}^{-1}\{\phi_s A_{h_0g_0}\}\{\phi_s A_{g_0g_0} + (1-\phi_s)N^{-1}B_{g_0g_0}\}^{-1} \\ &\quad \{\phi_s A_{g_0h_0}\}W_s][(W_1\otimes Y_1)\dots(W_{v_0}\otimes Y_{v_0})]^{-1}. \qquad (2.246)\end{aligned}$$

4. The collection $Z_{(N;R)} = \{Z_{(N;R)ds}\}$ for $d = 1,\dots,h_0$, $s = 1,\dots,v_0$ are the canonical directions of the adjustment with the corresponding canonical resolutions given *by* $\lambda_{(N;R)ds}$.

5. *The collection* $Z_{(N;R)} = \{Z_{(N;R)ds}\}$ *for* $d = 1, \ldots, h_0$, $s = 1, \ldots, v_0$ *are also canonical directions of the adjustment of* $[\mathcal{C}(N;R)]$ *by* $\mathcal{C}(N)$ *with canonical resolutions given by* $\lambda_{(N;R)ds}$. *The* $Z_{(N;R)ds}$ *are the only canonical directions with non-zero resolutions.*

Proof - Statement 1. follows by verifying that for $s \neq t$, $[\mathcal{Z}_{(N;R)s}]$ and $[\mathcal{Z}_{(N;R)t}]$ are orthogonal. Notice that

$$\begin{aligned} & Cov(Z_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}, Z_{[\mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h)/\mathcal{C}_h(n_h)]t}) = \\ & \quad (D_{Rs})_{gg} Y_s^T Cov(\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g), \mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h)) Y_t (D_{Rs})_{hh} \qquad (2.247) \\ & = (D_{Rs})_{gg} Y_s^T \{\alpha_{gh} C + \delta_{gh}(\beta_g/r_g) E\} Y_t^T (D_{Rs})_{hh} \qquad (2.248) \\ & = \delta_{st}[(D_{Rs})_{gg}\{\alpha_{gh} + \delta_{gh}(\beta_g/r_g)((1/\phi_s) - 1\}(D_{Rs})_{hh}], \qquad (2.249) \end{aligned}$$

where equation (2.248) follows from equation (2.198) and equation (2.249) follows from the choice of Y in Definition 12. Hence, when $s \neq t$,

$$Cov(Z_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}, Z_{[\mathcal{S}_{(n_h;r_h)}(\mathcal{C}_h)/\mathcal{C}_h(n_h)]t}) = 0, \qquad (2.250)$$

so that $[\mathcal{Z}_{(N;R)s}]$ and $[\mathcal{Z}_{(N;R)t}]$ are orthogonal.

Statement 2. follows immediately from Theorem 15.

To obtain Statement 3., we use the Bayes linear sufficiency of $\mathcal{S}_N(\mathcal{C})$ for $\mathcal{C}(N)$ for adjusting $\mathcal{S}_{(N;R)}(\mathcal{C})$ which follows from Theorem 9 and the observation that $[\mathcal{S}_{(N;R)}(\mathcal{C})] \subset [\mathcal{C}(N;R)]$. Letting $T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$ denote the resolution transform for the adjustment of $[\mathcal{S}_{(N;R)}(\mathcal{C})]$ by $\mathcal{S}_N(\mathcal{C})$, the Bayes linear sufficiency means we have that $T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{C}(N)]} = T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$. Notice from equation (1.74) that

$$\begin{aligned} T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]} &= \{Var(\mathcal{S}_{(N;R)}(\mathcal{C}))\}^{-1} Cov(\mathcal{S}_{(N;R)}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) \times \\ & \quad \{Var(\mathcal{S}_N(\mathcal{C}))\}^{-1} Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{S}_{(N;R)}(\mathcal{C})). \qquad (2.251) \end{aligned}$$

Then we may easily verify equation (2.245) by using equations (2.200) - (2.202) in equation (2.251). To verify equation (2.246), we use equations (2.239) - (2.242) in equation (2.251).

Using equation (2.246) and substituting in equation (2.208) gives

$$\begin{aligned} & T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]} = \\ & [(W_1 \otimes Y_1) \ldots (W_{v_0} \otimes Y_{v_0})][\oplus_{s=1}^{v_0} \Lambda_{(N;R)} s][(W_1 \otimes Y_1) \ldots (W_{v_0} \otimes Y_{v_0})]^{-1} \qquad (2.252) \end{aligned}$$

so that

$$T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}[(W_1 \otimes Y_1)\dots(W_{v_0} \otimes Y_{v_0})] = [(W_1 \otimes Y_1)\dots(W_{v_0} \otimes Y_{v_0})][\oplus_{s=1}^{v_0}\Lambda_{(N;R)}s]. \quad (2.253)$$

Hence $(Z_{(N:R)ds}, \lambda_{(N:R)ds})$ is an eigenvector/value pair for $T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$ for all $d = 1, \dots, h_0$, $s = 1, \dots, v_0$. Statement 4. thus follows.

A proper canonical resolution is one that is positive. Denote the resolution transform for the adjustment of $[\mathcal{C}(N;R)]$ by $\mathcal{S}_N(\mathcal{C})$ as $T_{[\mathcal{C}(N;R)/\mathcal{S}_N(\mathcal{C})]}$ and the transform for the adjustment of $[\mathcal{S}_N(\mathcal{C})]$ by $\mathcal{C}(N;R)$ as $T_{[\mathcal{S}_N(\mathcal{C})/\mathcal{C}(N;R)]}$. Then, as Goldstein (1990; p153) states, the proper canonical resolutions of $T_{[\mathcal{C}(N;R)/\mathcal{S}_N(\mathcal{C})]}$ and $T_{[\mathcal{S}_N(\mathcal{C})/\mathcal{C}(N;R)]}$ are the same. Now for each $\mathcal{X}_g \in \langle \mathcal{C}_g \rangle$, where $1 \leq g \leq h_0$, we may define for each $i = 1, \dots, n_g$

$$\mathcal{T}_i(\mathcal{X}_g) = \mathcal{X}_{gi} - \mathcal{S}_{(n_g;r_g)}(\mathcal{X}_g), \quad (2.254)$$

where we recall that $\mathcal{S}_{(n_g;r_g)}(\mathcal{X}_g)$ is as given by equation (2.195) and let $\mathcal{T}_i(\mathcal{C}_g(n_g;r_g)) = \{\mathcal{T}_i(X_{g1}), \dots, \mathcal{T}_i(X_{gv_0})\}$. Then subtracting equation (2.197) from equation (2.196) we have, for all $g = 1, \dots, g_0$; $h = 1, \dots, h_0$ and $i = 1, \dots, n_h + r_h$ that

$$Cov(\mathcal{S}_{n_g}(\mathcal{C}_g), \mathcal{T}_i(\mathcal{C}_h(n_h;r_h))) = 0, \quad (2.255)$$

and by subtracting equation (2.198) from equation (2.199) that

$$Cov(\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g), \mathcal{T}_i(\mathcal{C}_h(n_h;r_h))) = 0, \quad (2.256)$$

for all $g = 1, \dots, g_0$; $h = 1, \dots, h_0$ and $i = 1, \dots, n_h + r_h$. Putting $\mathcal{T}_{(N;R)}(\mathcal{C}) = \cup_{g=1}^{h_0} \cup_{i=n_g+1}^{n_g+r_g} \mathcal{T}_i(\mathcal{C}_g(n_g;r_g))$ then we have $[\mathcal{C}(N;R)] \subseteq [\mathcal{S}_{(N;R)}(\mathcal{C})] \cup [\mathcal{T}_{(N;R)}(\mathcal{C})]$ where from equation (2.256) we have that $[\mathcal{T}_{(N;R)}(\mathcal{C})] \perp [\mathcal{S}_{(N;R)}(\mathcal{C})]$ and from equation (2.255) we have that $[\mathcal{T}_{(N;R)}(\mathcal{C})] \perp [\mathcal{S}_N(\mathcal{C})]$. Hence, to follow the notation of Goldstein (1988a), we have

$$\begin{aligned}
[\mathcal{S}_N(\mathcal{C})/\mathcal{C}(N;R)] &= [\mathcal{S}_N(\mathcal{C})/(\mathcal{S}_{(N;R)}(\mathcal{C}) + \mathcal{T}_{(N;R)}(\mathcal{C}))] && (2.257)\\
&= [[\mathcal{S}_N(\mathcal{C})/\mathcal{T}_{(N;R)}(\mathcal{C})]/[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{T}_{(N;R)}(\mathcal{C})]] && (2.258)\\
&= [\mathcal{S}_N(\mathcal{C})/\mathcal{S}_{(N;R)}(\mathcal{C})], && (2.259)
\end{aligned}$$

where equation (2.258) follows from Property 5 of Goldstein (1988a) and equation (2.259) follows by orthogonality and Property 1 of Goldstein (1988a). Thus, from Goldstein (1990, p153), we have that

$$T_{[\mathcal{S}_N(\mathcal{C})/\mathcal{C}(N;R)]} = T_{[\mathcal{S}_N(\mathcal{C})/\mathcal{S}_{(N;R)}(\mathcal{C})]} \quad (2.260)$$

where $T_{[\mathcal{S}_N(\mathcal{C})/\mathcal{S}_{(N;R)}(\mathcal{C})]}$ denotes the resolution transform for the adjustment of $[\mathcal{S}_N(\mathcal{C})]$ by $\mathcal{S}_{(N;R)}(\mathcal{C})$. Thus, the proper canonical resolutions for $T_{[\mathcal{C}(N;R)/\mathcal{S}_N(\mathcal{C})]}$ are the same as those of $T_{[\mathcal{S}_N(\mathcal{C})/\mathcal{S}_{(N;R)}(\mathcal{C})]}$, which are the same as those of $T_{[\mathcal{S}_{(N;R)}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$, which are $\lambda_{(N;R)ds}$ for $d = 1, \dots, h_0$, $s = 1, \dots, v_0$. Noting that $Z_{(N;R)ds} \in [\mathcal{C}(N;R)]$ gives that $(Z_{(N;R)ds}, \lambda_{(N;R)ds})$ are the pairs of canonical directions and resolutions for the adjustment of $[\mathcal{C}(N;R)]$ by $\mathcal{S}_N(\mathcal{C})$. Statement 5. follows by the Bayes Linear sufficiency of $\mathcal{S}_N(\mathcal{C})$ for $\mathcal{C}(N)$. □

This theorem illustrates how we may simplify the design and should be compared with Theorem 13. Notice the similarities. We break down the $v_0h_0 \times v_0h_0$ problem into one problem of size $v_0 \times v_0$ and then v_0 problems of size $h_0 \times h_0$. The $v_0 \times v_0$ problem is identical to the one where we considered adjustment over the population structure and does not depend upon the sample sizes. The predictive adjustment and adjustment over the population structure do differ in the types of qualitative information they provide. This is due to the difference in the group problem that we solve; it provides different qualitative information both in the difference between the population structure inference and the predictive case, but also across the v_0 problems. It should however be noted that the motivational justification of each of the v_0 problems remains the same as that for the population structure inference. Thus, although the qualitative information changes, our understanding of the interpretative forms of the design simplification remains the same.

The v_0 $h_0 \times h_0$ problems as given by Definition 14 have a similar form and we need only to solve a single one in terms of ϕ_s to have solved all v_0 problems. Indeed, there may be situations where the v_0 problems all have the same form. Once such case is when $h_0 = g_0$ and $R \propto N$. In this case, it is straightforward to see from equation (2.208) that $W_s = D_{Rs}W$, where W is the matrix of underlying canonical group directions as given in Definition 13. Thus, the adjustment in each $[Z_{(N;R)s}]$ is qualitatively the same and we immediately have the following corollary to Theorem 16.

Corollary 6 *If $g_0 = h_0$ and $R \propto N$, then up to scaling factors the canonical directions, with non-zero resolutions, of $T_{[\mathcal{C}(N;R)/\mathcal{C}(N)]}$ share the same co-ordinate representation as the canonical directions of $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]}$. If $\lambda_{(N)ds}$ is the canonical resolution of $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]}$ corresponding to canonical direction $Z_{(N)ds}$, then the canonical resolution of $T_{[\mathcal{C}(N;R)/\mathcal{C}(N)]}$ corresponding to canonical direction $Z_{(N;R)ds}$ is given by*

$$\lambda_{(N;R)ds} = \lambda_{(N)ds}\lambda_{(R)ds}. \tag{2.261}$$

Notice that in this case, the relationship between predictive adjustment and adjustment over the population structure shares the same features as for the case of a single exchangeable sequence given by Goldstein & Wooff (1998), see equation (1.71). Since $\lambda_{(N;R)ds}$ are the only non-zero canonical resolutions of the adjustment of $[\mathcal{C}(N;R)]$ by $\mathcal{C}(N)$, then the full predictive adjustment shares the same qualitative features as the adjustment over the population structure and hence the implications for interpretation and design are analogous.

2.6.3 Example: the examiner uses prediction

Suppose that having observed the marks on n papers from each of the two markers, the examiner wishes to consider the effect of these marks for predicting the marks for the next r_1 students in the first group and the next r_2 students in the second group. Having already calculated the canonical structure for the adjustment of the population collection by the sample, as given by Theorem 13, then to calculate the canonical structure for this predictive problem the examiner knows from Theorem 16 that all he needs to calculate to find this is the sth underlying canonical predictive group problem as given by Definition 14, for each $s = 1, 2$, since from equation (2.118) that the solution of the underlying canonical variable problem is

$$Y = \begin{pmatrix} \sqrt{\frac{2}{3}} & \sqrt{\frac{2}{5}} \\ -\sqrt{\frac{2}{3}} & \sqrt{\frac{2}{5}} \end{pmatrix}; \Phi = \begin{pmatrix} \frac{3}{5} & 0 \\ 0 & \frac{1}{7} \end{pmatrix}. \tag{2.262}$$

In the sth underlying canonical predictive group problem, the examiner has

$$A = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix}; B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; N = \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix}; R = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix}. \tag{2.263}$$

He thus solves the sth underlying canonical predictive group problem to obtain

$$\begin{aligned}
\lambda_{(nI_2;R)1s} &= \frac{\lambda_{(nI_2)1s}\lambda_{(r_2I_2)1s}}{1+((1/r_1)-(1/r_2))f_1(\phi_s)} + ((1/r_1)-(1/r_2))f_2(\phi_s) && (2.264)\\
W_{1s} &\propto [1 \;\; 1+((1/r_1)-(1/r_2))f_3(\phi_s)]^T && (2.265)\\
\lambda_{(nI_2;R)2s} &= \frac{\lambda_{(nI_2)2s}\lambda_{(r_2I_2)2s}}{1+((1/r_1)-(1/r_2))f_1(\phi_s)} + ((1/r_1)-(1/r_2))f_4(\phi_s) && (2.266)\\
W_{2s} &\propto [1 \;\; -1+((1/r_1)-(1/r_2))f_5(\phi_s)]^T && (2.267)
\end{aligned}$$

where $\lambda_{(mI_2)ds}$ is the (d,s)th canonical resolution for the adjustment of $\mathcal{M}(\mathcal{C})$ by $\mathcal{C}(mI_2)$. For the precise form, see for example, equations (2.159) and (2.160). $f_1(\phi_s), \dots, f_5(\phi_s)$ are functions of ϕ_s obtained by regarding the solution of the sth underlying canonical predictive group problem as given by Definition 14 as a function of ϕ_s. From Theorem 16, the canonical directions with non-zero resolutions for the adjustment of $[\mathcal{C}(nI_2; R)]$ by $\mathcal{C}(nI_2)$ are $Z_{(nI_2;R)ds} = (W_{ds} \otimes Y_s)^T \mathcal{S}_{(nI_2;R)}(\mathcal{C})$, the corresponding canonical resolutions being $\lambda_{(nI_2;R)ds}$.

Observe that when $r_1 = r_2$, so that $R \propto N$, then the canonical resolutions do indeed have the form of Corollary 6. Notice that in this case, up to a scale factor to ensure a prior variance of one, the $Z_{(nI_2;R)ds}$ do have the same co-ordinate representation as the $Z_{(nI_2)ds}$ and that this co-ordinate representation is fixed for all choices of $r_1 = r_2$ and n.

2.7 Predictive adjustment: predicting in unobserved groups

We wish to examine the effect of the observation of the sample of n_g individuals in the gth group for $g = 1, \dots, g_0$ for predicting the values for the first r_h individuals in the hth group for $h = g_0 + 1, \dots, g_0 + h_0$. Thus, we are interested in predicting in groups where we have not previously observed any individuals. We collect the prediction sample sizes together as the $h_0 \times h_0$ diagonal matrix R with (h, h)th entry $(R)_{hh} = r_{g_0+h}$.

Let $\mathcal{C}_h(r_h) = \cup_{i=1}^{r_h} \mathcal{C}_{hi}$ denote the collection of individuals in the hth group that we would like to predict the value for having observed our sample $\mathcal{C}(N)$. We denote the complete collection of individuals that we would like to predict the values of as $\mathcal{C}(R/N) = \cup_{h=1}^{h_0} \mathcal{C}_h(r_h)$. Form the collection $\mathcal{S}_{r_h}(\mathcal{C}_h) = \{\mathcal{S}_{r_h}(X_{h1}), \dots, \mathcal{S}_{r_h}(X_{hv_0})\}$ of averages of the r_h future individuals in the hth group.

In a similar way to how we developed prediction in already observed groups, we shall initially restrict attention to adjusting the complete collection of averages of future observations, $\mathcal{S}_{(R/N)}(\mathcal{C}) = \cup_{h=1}^{h_0} \mathcal{S}_{r_h}(\mathcal{C}_h)$, before showing that this study is sufficient to give us all the canonical directions with non-zero canonical resolutions. In a similar way to Lemma 9 we have the following lemma.

Lemma 13 *The second-order relationships between the $\mathcal{S}_{n_g}(\mathcal{C}_g)$s, the $\mathcal{C}_h(r_h)$s and the $\mathcal{S}_{r_h}(\mathcal{C}_h)$s may be expressed, $\forall g = 1, \dots, g_0$; $\forall h, h' = 1, \dots, h_0$; $\forall i = 1, \dots r_{g_0+h}$*

as

$$
\begin{aligned}
Cov(\mathcal{S}_{n_g}(\mathcal{C}_g), \mathcal{C}_{g_0+hi}) &= \alpha_{gg_0+h}C; & (2.268)\\
Cov(\mathcal{S}_{n_g}(\mathcal{C}_g), \mathcal{S}_{r_{g_0+h}}(\mathcal{C}_{g_0+h})) &= \alpha_{gg_0+h}C; & (2.269)\\
Cov(\mathcal{S}_{r_{g_0+h}}(\mathcal{C}_{g_0+h}), \mathcal{S}_{r_{g_0+h'}}(\mathcal{C}_{g_0+h'})) &= \begin{cases} \alpha_{g_0+hg_0+h}C + \frac{1}{r_{g_0+h}}\beta_{g_0+h}E & h = h';\\ \alpha_{g_0+hg_0+h'h}C & h \neq h'; \end{cases} & (2.270)\\
Cov(\mathcal{S}_{r_{g_0+h'}}(\mathcal{C}_{g_0+h'}), \mathcal{C}_{g_0+hi}) &= \begin{cases} \alpha_{g_0+hg_0+h}C + \frac{1}{r_{g_0+h}}\beta_{g_0+h}E & h = h';\\ \alpha_{g_0+h'g_0+h}C & h' \neq h. \end{cases} & (2.271)
\end{aligned}
$$

We denote by $A_{p_0+pq_0+q}$ the $p \times q$ matrix with (p_1, q_1)th entry $(A_{p_0+pq_0+q})_{p_1q_1} = \alpha_{p_0+p_1q_0+q_1}$. If either p_0 or q_0 is zero, we omit the addition. Thus, for example, A_{pq_0+q} is the $p \times q$ matrix with (p_1, q_1)th entry $(A_{pq_0+q})_{p_1q_1} = \alpha_{p_1q_0+q_1}$. Notice that this conforms with the definition of A_{pq} given in the previous section. In a similar way, we let $B_{p_0+pp_0+p}$ be the $p \times p$ diagonal matrix with (p_1, p_1)th entry $(B_{p_0+pp_0+p})_{p_1p_1} = \beta_{p_0+p_1}$. Gathering the $\mathcal{S}_{r_h}(\mathcal{C}_h)$s together into the $h_0v_0 \times 1$ vector $\mathcal{S}_{(R/N)}(\mathcal{C})$, then we have the following lemma to express equations (2.200), (2.269) and (2.270).

Lemma 14 *The second-order relationships between $\mathcal{S}_N(\mathcal{C})$ and $\mathcal{S}_{(R/N)}(\mathcal{C})$ may be expressed as:*

$$
\begin{aligned}
Var(\mathcal{S}_N(\mathcal{C})) &= (A_{g_0g_0} \otimes C) + (N^{-1}B_{g_0g_0} \otimes E); & (2.272)\\
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{S}_{(R/N)}(\mathcal{C})) &= A_{g_0g_0+h_0} \otimes C; & (2.273)\\
Var(\mathcal{S}_{(R/N)}(\mathcal{C})) &= (A_{g_0+h_0g_0+h_0} \otimes C) + (R^{-1}B_{g_0+h_0g_0+h_0} \otimes E) & (2.274)
\end{aligned}
$$

2.7.1 Analysis of groups and variables

We consider separately the analysis of a variable problem and related group problem. Again, we make use of the underlying canonical variable structure defined in Definition 12. We motivate the use of it as follows. Consider the problem of examining the effect of a sample of n_g individuals from the gth group, for some $g \in \{1, \dots, g_0\}$, for predicting the values for a collection of r_h individuals in the hth group, for some $h \in \{g_0+1, \dots, h_0+g_0\}$. We denote the resolution transform for the adjustment of $[\mathcal{C}_h(r_h)]$ by $\mathcal{C}_g(n_g)$ as $T_{[\mathcal{C}_h(r_h)/\mathcal{C}_g(n_g)]}$ and the transform for the adjustment of $[\mathcal{S}_{r_h}(\mathcal{C}_h)]$ by $\mathcal{C}_g(n_g)$ as $T_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]}$. We have the following theorem.

Theorem 17 *The resolution transform matrix, $T_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]}$, is calculated as*

$$
T_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]} = \{\alpha_{hh}C + (1/r_h)\beta_h E\}^{-1}(\alpha_{hg}C)\{\alpha_{gg}C + (1/n_g)\beta_g E\}^{-1}(\alpha_{gg}C). \quad (2.275)
$$

For each $s = 1, \dots, v_0$, the canonical directions for the adjustment of $[\mathcal{S}_{r_h}(\mathcal{C}_h)]$ by $\mathcal{C}_g(n_g)$ are given by

$$Z_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]s} = \sqrt{\frac{r_h\phi_s}{r_h\alpha_{hh}\phi_s + \beta_h(1-\phi_s)}} Y_s^T \mathcal{S}_{r_h}(\mathcal{C}_h), \tag{2.276}$$

with the corresponding canonical resolutions given by

$$\lambda_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]s} = \frac{\alpha_{gh}^2}{\alpha_{gg}\alpha_{hh}} \lambda_{[\mathcal{M}(\mathcal{C}_h)/\mathcal{C}_h(r_h)]s} \lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}, \tag{2.277}$$

where $\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(m_g)]s}$ is as given by equation (2.123). The collection

$$\Lambda_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]} = \{\lambda_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]1}, \dots, \lambda_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]v_0}\} \tag{2.278}$$

is the collection of non-zero canonical resolutions for the adjustment of $[\mathcal{C}_h(r_h)]$ by $\mathcal{C}_g(n_g)$; the corresponding canonical directions being given by

$$Z_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]} = \{Z_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]1}, \dots, Z_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]v_0}\}. \tag{2.279}$$

Proof - Notice that since the collection $\mathcal{C}_h(r_h)$ respects exchangeability with $\mathcal{C}_g(n_g)$, then from Goldstein & Wooff (1998), we have that the sample means, $\mathcal{S}_{(n_g)}(\mathcal{C}_g)$ are Bayes linear sufficient for $\mathcal{C}_g(n_g)$ for the adjustment of $[\mathcal{C}_h(r_h)]$. Hence $T_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]}$ $= T_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{S}_{(n_g)}(\mathcal{C}_g)]}$. Equation (2.275) follows by using equations (2.94), (2.269) and (2.270) in equation (1.74). The eigenvalues and eigenvectors of this then follow immediately from the solution of the canonical variable problem given in Definition 12. That these are the only non-zero canonical resolutions follows in an identical manner to the proof of Theorem 16, in this case we have that

$$T_{[\mathcal{S}_{n_g}(\mathcal{C}_g)/\mathcal{C}_h(r_h)]} = T_{[\mathcal{S}_{n_g}(\mathcal{C}_g)/\mathcal{S}_{r_h}(\mathcal{C}_h)]}, \tag{2.280}$$

and the result follows. □

The crucial points to note here are as follows. Each $Z_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]s}$ does not depend upon the observation group g, either through the relationships between the group h and the group g nor on the sample size, n_g, observed in the group. Thus, the qualitative form of the adjustment remains the same for all choices of g and all choices of n_g. Note further that each $Z_{[\mathcal{S}_{r_h}(\mathcal{C}_h)/\mathcal{C}_g(n_g)]s}$ has the same form as $Z_{[\mathcal{S}_{(n_g;r_g)}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ so that prediction in the observed group has the same qualitative information as prediction in the unobserved group.

Definition 15 *The sth underlying canonical predictive unobserved group directions are defined as the columns of the matrix $W_s = [W_{1s} \dots W_{h_0 s}]$ solving the generalised* *eigenvalue problem*

$$(\phi_s A_{g_0+h_0g_0})\{\phi_s A_{g_0g_0} + (1-\phi_s)N^{-1}B_{g_0g_0}\}^{-1}(\phi_s A_{g_0g_0+h_0})W_s \;=\; \{\phi_s A_{g_0+h_0g_0+h_0} + (1-\phi_s)R^{-1}B_{g_0+h_0g_0+h_0}\}W_s\Lambda_{(R/N)s}, \quad (2.281)$$

where $\Lambda_{(R/N)s} = diag(\lambda_{(R/N)1s}, \dots, \lambda_{(R/N)h_0s})$ *is the matrix of eigenvalues.* W_s *is chosen so that* $W_s^T\{A_{g_0+h_0g_0+h_0} + ((1/\phi_s)-1)R^{-1}B_{g_0+h_0g_0+h_0}\}W_s = I_{h_0}$, $W_s^T A_{g_0g_0}$ $\{A_{g_0g_0} + ((1/\phi_s)-1)N^{-1}B_{g_0g_0}\}^{-1}A_{g_0g_0+h_0}W_s = \Lambda_{(R/N)s}$. *The ordered eigenvalues* $1 > \lambda_{(R/N)1s} \geq \dots \geq \lambda_{(R/N)h_0s} \geq 0$ *are termed the sth underlying canonical predictive unobserved group resolutions.*

Notice that Definition 15 has almost the identical form as Definition 14. We motivate the definition in identically the same way. For each $s = 1, \dots, v_0$ form the collection:

$$\mathcal{Z}_{(R/N)s} \;=\; \{Z_{[\mathcal{S}_{r_{g_0+1}}(\mathcal{C}_{g_0+1})/\mathcal{C}_g(n_g)]s}, \dots, Z_{[\mathcal{S}_{r_{g_0+h_0}}(\mathcal{C}_{g_0+h_0})/\mathcal{C}_g(n_g)]s}\}. \quad (2.282)$$

As we have already commented, $\mathcal{Z}_{(R/N)s}$ does not depend on g. Letting D_{Rs} be the $h_0 \times h_0$ diagonal matrix with (h,h)th entry

$$(D_{Rs})_{hh} \;=\; \sqrt{\frac{r_{g_0+h}\phi_s}{r_{g_0+h}\alpha_{g_0+hg_0+h}\phi_s + \beta_{g_0+h}(1-\phi_s)}}, \quad (2.283)$$

then in our usual vector notation, we may express equation (2.282) as

$$\mathcal{Z}_{(R/N)s} \;=\; D_{Rs}(I_{h_0} \otimes Y_s^T)\mathcal{S}_{(R/N)}(\mathcal{C}). \quad (2.284)$$

We consider the adjustment of $[\mathcal{Z}_{(R/N)s}]$ by the observed sample, $\mathcal{C}(N)$. We denote the resolution transform for this adjustment by $T_{[\mathcal{Z}_{(R/N)s}/\mathcal{C}(N)]}$. We have the following theorem; the proof is identical to that of Theorem 15.

Theorem 18 *For the adjustment of* $[\mathcal{Z}_{(R/N)s}]$ *by* $\mathcal{C}(N)$, *the resolution transform matrix is calculated as*

$$T_{[\mathcal{Z}_{(R/N)s}/\mathcal{C}(N)]} \;=\; D_{Rs}^{-1}\{\phi_s A_{g_0+h_0g_0+h_0} + (1-\phi_s)R^{-1}B_{g_0+h_0g_0+h_0}\}^{-1} \{\phi_s A_{g_0+h_0g_0}\}\{\phi_s A_{g_0g_0} + (1-\phi_s)N^{-1}B_{g_0g_0}\}^{-1}\{\phi_s A_{g_0g_0+h_0}\}D_{Rs}. \quad (2.285)$$

The canonical directions are given by $(D_{Rs}^{-1}W_{ds})^T\mathcal{Z}_{(R/N)s}$ *for each* $d = 1, \dots, h_0$, *with corresponding canonical resolutions given by* $\lambda_{(R/N)ds}$. W_{ds} *is the* (d,s)*th underlying canonical predictive unobserved group direction as given in Definition 15;* $\lambda_{(R/N)ds}$ *the corresponding* (d,s)*th underlying canonical predictive unobserved group resolution.*

2.7.2 Adjustment of the full collection

We consider the adjustment of the full collection $[\mathcal{C}(R/N)]$. In an identical way to our previously considered problems, the canonical variable and each canonical predictive unobserved group analysis completely determine the adjustment of the full collection. We denote by $T_{[\mathcal{S}_{(R/N)}(\mathcal{C})/\mathcal{C}(N)]}$ the resolution transform for the adjustment of $[\mathcal{S}_{(R/N)}(\mathcal{C})]$ by $\mathcal{C}(N)$. We have the following theorem; the proof follows in an identical fashion to Theorem 16.

Theorem 19 *The adjustment of $[\mathcal{S}_{(R/N)}(\mathcal{C})]$ by $\mathcal{C}(N)$ satisfies the following properties*

1. There exist v_0 h_0-dimensional orthogonal subspaces, $[\mathcal{Z}_{(R/N)1}], \ldots, [\mathcal{Z}_{(R/N)v_0}]$ of $[\mathcal{S}_{(R/N)}(\mathcal{C})]$. For each s,

$$\mathcal{Z}_{(R/N)s} = \{Z_{[\mathcal{S}_{r_{g_0+1}}(\mathcal{C}_{g_0+1})/\mathcal{C}_g(n_g)]s}, \ldots, Z_{[\mathcal{S}_{r_{g_0+h_0}}(\mathcal{C}_{g_0+h_0})/\mathcal{C}_g(n_g)]s}\}. \quad (2.286)$$

2. The canonical directions for the adjustment in each $[\mathcal{Z}_{(R/N)s}]$ are given by the columns of the matrix $Z_{(R/N)s} = [Z_{(R/N)1s} \ldots Z_{(R/N)v_0 s}]$, where

$$Z_{(R/N)ds} = (D_{Rs}^{-1}W_{ds})^T \mathcal{Z}_{(R/N)s} = (W_{ds}^T \otimes Y_s^T)\mathcal{S}_{(R/N)}(\mathcal{C}). \quad (2.287)$$

3. The resolution transform matrix of the full collection is calculated as

$$\begin{aligned} T_{[\mathcal{S}_{(R/N)}(\mathcal{C})/\mathcal{C}(N)]} &= \{(A_{g_0+h_0g_0+h_0} \otimes C) + (R^{-1}B_{g_0+h_0g_0+h_0} \otimes E)\}^{-1} \\ &\quad (A_{g_0+h_0g_0} \otimes C)\{(A_{g_0g_0} \otimes C) + (N^{-1}B_{g_0g_0} \otimes E)\}^{-1}(A_{g_0g_0+h_0} \otimes C) \quad (2.288) \\ &= [(W_1 \otimes Y_1) \ldots (W_{v_0} \otimes Y_{v_0})][\oplus_{s=1}^{v_0} W_s^{-1}\{\phi_s A_{g_0+h_0g_0+h_0} + (1-\phi_s) \\ &\quad R^{-1}B_{g_0+h_0g_0+h_0}\}^{-1}\{\phi_s A_{g_0+h_0g_0}\}\{\phi_s A_{g_0g_0} + (1-\phi_s)N^{-1}B_{g_0g_0}\}^{-1} \\ &\quad \{\phi_s A_{g_0g_0+h_0}\}W_s][(W_1 \otimes Y_1) \ldots (W_{v_0} \otimes Y_{v_0})]^{-1}. \quad (2.289) \end{aligned}$$

4. The collection $Z_{(R/N)} = \{Z_{(R/N)ds}\}$ for $d = 1, \ldots, h_0$, $s = 1, \ldots, v_0$ are the canonical directions of the adjustment with the corresponding canonical resolutions given by $\lambda_{(R/N)ds}$.

5. The collection $Z_{(R/N)} = \{Z_{(R/N)ds}\}$ for $d = 1, \ldots, h_0$, $s = 1, \ldots, v_0$ are also canonical directions of the adjustment of $[\mathcal{C}(R/N)]$ by $\mathcal{C}(N)$ with corresponding canonical resolutions given by $\lambda_{(R/N)ds}$. The $Z_{(R/N)ds}$ are the only canonical directions with (potentially) non-zero resolutions.

Thus, the theorem tells us how we may calculate the canonical directions and resolutions for the adjustment of $[\mathcal{C}(R/N)]$ by $\mathcal{C}(N)$ by solving a series of lower dimensional problems. As we have emphasised, we may use the resulting canonical structure to, for example, make decisions about the sample sizes require to achieve desired levels of variance reduction for quantities of interest within $[\mathcal{C}(R/N)]$, whilst also providing insights about how we expect to learn for the given design by supplying us with the orthogonal grid of most important directions. The reduction into interpretable subproblems remains the same as in the adjustment of the population structure and also for the prediction in already observed groups. In each of the three cases, our understanding of how we may simplify the design, and how we interpret the effect of the groups remains the same.

Notice that if we let r_h tend to infinity in each of the h_0 previously unobserved groups, then $\mathcal{S}_{r_h}(\mathcal{C}_h) \to \mathcal{M}(\mathcal{C}_h)$, the underlying mean components in the hth group. By letting $R \to 0$ in all of our results above, we derive the theory for the adjustment of the previously unobserved collection of mean components $\mathcal{M}_{(R/N)}(\mathcal{C}) = \cup_{h=g_0+1}^{g_0+h_0} \mathcal{M}(\mathcal{C}_h)$ by $\mathcal{C}(N)$.

2.7.3 Unobserved prediction for the examiner

To link with our running example, suppose that the examiner has a total of $h_0 + 2$ markers and he asks for the first n marks from each of the markers. These are sent to him via the post. On the first day, he receives the collection of marks from the first two markers and he wants to use these to predict the marks for the collections he expects to receive from the remaining h_0 markers. From Theorem 19, he knows that all he is additionally required to do in order to calculate the canonical resolutions and directions for the (unobserved) predictive adjustment is to solve the sth underlying canonical predictive group problem for each $s = 1, 2$. He has $A_{2+h_0 2+h_0} = I_{h_0} + \gamma(J_{h_0} - I_{h_0})$; $R = nI_{h_0}$; $B_{2+h_0 2+h_0} = I_{h_0}$; and $A_{2+h_0 2} = \gamma J_{2+h_0 2}$. It is then straightforward to see that the sth underlying canonical predictive group problem reduces to finding the eigenstructure of the matrix

$$T = \frac{2\{n\phi_s\gamma\}^2}{\{n\phi_s(1+(h_0-1)\gamma)+(1-\phi_s)\}\{n\phi_s(1+\gamma)+(1-\phi_s)\}} J_{h_0}, \qquad (2.290)$$

so that, as we would expect if $h_0 > g_0$, there will be zero resolutions.

Chapter 3

Multifactor multivariate exchangeable systems

SUMMARY

We consider a special case to the beliefs considered in Chapter 2 by considering cases where the groups can be thought of as convenient labellings for more complicated structures, namely particular combinations in a factorial design. We examine a particular set of beliefs, as described in Section 3.1. These maintain the beliefs in the same form as Chapter 2 and so the solution of the mean components is found by solving the underlying canonical variable and group problems discussed in Subsections 2.4.1 and 2.4.3. However, our beliefs also allow us to decompose the underlying canonical group problem into a series of underlying canonical factor problems, as described in Section 3.3. The solution of these enables us to solve the underlying canonical group problem and hence for the full adjustment of the mean components, as given in Section 3.4. Having considered the groups as particular combinations in a factorial design, we consider, in Section 3.6, marginalising over some of the factors, and in Section 3.7 of taking slices of the table. In both cases, the adjustment may be deduced from the solution of the underlying canonical variable problem of Subsection 2.4.1 and each underlying canonical factor problem of Section 3.3. In the marginal case, in Subsection 3.6.1, we show that the canonical directions may be deduced from the underlying canonical variable directions and the unmarginalised

underlying canonical factor directions. The corresponding canonical resolutions are the weighted sum of the canonical resolutions for the full adjustment over the marginalised factors. In the sliced case, in Subsection 3.7.1, we show that the canonical directions may be deduced from the underlying canonical variable directions and the unrestrained underlying canonical factor directions. The corresponding canonical resolutions are the weighted sum of the canonical resolutions for the full adjustment over the restrained factors. Once more, we illustrate the results by a simple extension of the examiner example.

3.1 Specifications of the model

In the previous chapter we considered grouped multivariate exchangeable systems. We now wish to extend this theory for the case when the 'groups' may be viewed as a convenient indexing of more complicated structures, for example, each 'group' could correspond to a particular combination in a factorial design.

Consider a factorial experiment where the effects of a number of different factors may be investigated simultaneously. We consider that we have a finite set, $\Delta = \{1, \ldots, k\}$, of classification criteria, or factors, with the qth factor having $l_{q,o}$ levels for each $q \in \Delta$. So, to follow Lauritzen (1996), for each $q = 1, \ldots, k$, $\mathcal{I}_q = \{1, \ldots, l_{q,0}\}$ denotes the set of possible levels of the qth factor. We consider that each possible combination that can be formed from the different factors is a cell, and each of the cells are arranged to form a table.

The cells of the table are the elements $l_{[k]} = (l_k, \ldots, l_1)$ of the product, $\mathcal{I}$, of the level sets; that is $\mathcal{I} = \times_{q=1,\ldots,k}\mathcal{I}_q$. We say that we have a k *− dimensional* $l_{k,0} \times l_{k-1,0} \times \cdots \times l_{1,0}$ table, so that the total number of possible cells is

$$l_0 = \prod_{q=1}^{k} l_{q,0}. \tag{3.1}$$

We consider that it is possible to classify each individual to a single cell. For each individual, we are interested in making the same series of measurements $\mathcal{C} = \{X_1, \ldots, X_{v_0}\}$. Let $\mathcal{C}_{l_{[k]}i} = \{X_{l_{[k]}1i}, \ldots, X_{l_{[k]}v_0i}\}$ be the values of the measurements for the ith individual in the cell $l_{[k]}$. For simplicity, as before, we use the same notation to denote the $v_0 \times 1$ column vector of measurements for the ith individual in the $l_{[k]}$th cell, namely $\mathcal{C}_{l_{[k]}i} = [X_{l_{[k]}1i} \ldots X_{l_{[k]}v_0i}]^T$.

We choose to reference the cells in two (equivalent) ways. We may make explicit the settings of each factor in the cell by referencing the cell as $l_{[k]}$. This notation may make it easier to highlight specific factor settings. For example, we may explicitly show the level of the qth factor by writing

$$l_{[k]} = l_{[k;q+1]} l_q l_{[q-1;1]} = (l_k, \ldots, l_{q+1}, l_q, l_{q-1}, \ldots, l_1). \tag{3.2}$$

Alternatively, we may label the cell by a number. The following definition makes our choice of number.

Definition 16 *The bijective mapping $f : \mathcal{I} \to \{1, \ldots, l_0\}$ defined by*

$$l_{[k]} \mapsto \left\{ \sum_{q=2}^{k} (l_q - 1) \prod_{q'=1}^{q-1} l_{q',0} \right\} + l_1 \tag{3.3}$$

enables us to reference the $l_{[k]}$th cell by a number, $g = f(l_{[k]})$.

Adopting this function enables us, for example, to write X_{gvi} for $X_{l_{[k]}vi}$ and $\mathcal{C}_{gi}$ for $\mathcal{C}_{l_{[k]}i}$.

For an example of the mapping, consider $k = 3$ with $l_{3,0} = 5$, $l_{2,0} = 4$ and $l_{1,0} = 3$ so that we are considering a $5 \times 4 \times 3$ table. The number of possible cells is thus $l_0 = 60$ and

$$f(l_3, l_2, l_1) = l_{1,0} l_{2,0}(l_3 - 1) + l_{1,0}(l_2 - 1) + l_1 \tag{3.4}$$

$$= 12(l_3 - 1) + 3(l_2 - 1) + l_1, \tag{3.5}$$

so that, for example, the cell $(1, 1, 1)$ maps to the first group and the cell $(2, 4, 1)$ maps to the twenty-second group.

We suppose that the individuals in each cell are drawn from a potentially infinite class of individuals and that individuals in the same cell are judged second-order exchangeable. We also judge that individuals are co-exchangeable across cells. The types of models we shall derive, for example in Subsection 3.1.2, may be linked to the structure of multivariate analysis of variance (MANOVA) models such as those in Chapter VI of Press (1989). By the representation theorem of Goldstein (1986a), see Section 1.6, for each cell $l_{[k]}$ we may introduce the collection of random quantities $\mathcal{M}(\mathcal{C}_{l_{[k]}}) = \{\mathcal{M}(X_{l_{[k]}1}), \ldots, \mathcal{M}(X_{l_{[k]}v_0})\}$, the collection of underlying mean components for the $l_{[k]}$th cell, and $\mathcal{R}_i(\mathcal{C}_{l_{[k]}}) = \{\mathcal{R}_i(X_{l_{[k]}1}), \ldots, \mathcal{R}_i(X_{l_{[k]}v_0})\}$, the collection of residual components for the ith individual in the $l_{[k]}$th cell, and write:

$$X_{l_{[k]}vi} = \mathcal{M}(X_{l_{[k]}v}) + \mathcal{R}_i(X_{l_{[k]}v}). \tag{3.6}$$

As before we regard each $\mathcal{M}(\mathcal{C}_{l_{[k]}})$, $\mathcal{R}_i(\mathcal{C}_{l_{[k]}})$ as the analogous column vectors, so that for example, $\mathcal{M}(\mathcal{C}_{l_{[k]}}) = [\mathcal{M}(X_{l_{[k]}1}) \dots \mathcal{M}(X_{l_{[k]}v_0})]^T$. Suppose that we have standardised our random quantities to have zero mean and judge that the second-order specifications given by equations (2.67) - (2.69) are still appropriate. That is we specify that:

$$
\begin{aligned}
Cov(\mathcal{M}(X_{l_{[k]}}), \mathcal{M}(X_{l'_{[k]}})) &= \alpha_{l_{[k]}l'_{[k]}} C \ \forall l_{[k]}, l'_{[k]}; &(3.7)\\
Cov(\mathcal{M}(X_{l_{[k]}}), \mathcal{R}_j(X_{l'_{[k]}})) &= 0 \ \forall l_{[k]}, l'_{[k]}, j; &(3.8)\\
Cov(\mathcal{R}_i(X_{l_{[k]}}), \mathcal{R}_j(X_{l'_{[k]}})) &= \begin{cases} \beta_{l_{[k]}} E & \forall l_{[k]} = l'_{[k]}, i = j; \\ 0 & \text{otherwise,} \end{cases} &(3.9)
\end{aligned}
$$

where, as before, we denote by C, E the general $v_0 \times v_0$ nonnegative definite matrices with (v, w)th entries $(C)_{vw} = c_{vw}$, $(E)_{vw} = e_{vw}$ respectively. At present, we have made no distinctions with the work in Chapter 2. Thus, we know that Theorem 13 may be applied to these specifications if we wish to adjust the mean components having observed a sample. We now wish to expand the work in Chapter 2 by considering in more detail the relationships between the cells. In this chapter, we shall assume that the $\alpha_{l_{[k]}l'_{[k]}}$ and $\beta_{l_{[k]}}$ may be decomposed in the following way.

$$
\begin{aligned}
\alpha_{l_{[k]}l'_{[k]}} &= \prod_{q=1}^{k} \alpha^{(q)}_{l_q l'_q} \ \forall l_{[k]}, l'_{[k]} \in \mathcal{I}; &(3.10)\\
\beta_{l_{[k]}} &= \prod_{q=1}^{k} \beta^{(q)}_{l_q} \ \forall l_{[k]} \in \mathcal{I}. &(3.11)
\end{aligned}
$$

Let $A^{(q)}$ be the $l_{q,0} \times l_{q,0}$ matrix with (l, l')th entry $(A^{(q)})_{ll'} = \alpha^{(q)}_{ll'}$. We will have need to refer to the column representation of $A^{(q)}$ which is

$$A^{(q)} = [a^{(q)}_1 \dots a^{(q)}_{l_{q,0}}], \tag{3.12}$$

where $a^{(q)}_{l_q}$ is the $l_{q,0} \times 1$ column vector

$$a^{(q)}_{l_q} = [\alpha^{(q)}_{1l_q} \dots \alpha^{(q)}_{l_{q,0}l_q}]^T = [\alpha^{(q)}_{l_q 1} \dots \alpha^{(q)}_{l_q l_{q,0}}]^T, \tag{3.13}$$

since $(A^{(q)})^T = A^{(q)}$. Let $B^{(q)}$ be the $l_{q,0} \times l_{q,0}$ diagonal matrix with (l, l)th entry $(B^{(q)})_{ll} = \beta^{(q)}_l$. In Chapter 2, we collected the α_{gh} and β_g into the matrices A and B; A having (g, h)th entry $(A)_{gh} = \alpha_{gh}$ and the diagonal matrix B having (g, g)th entry $(B)_{gh} = \beta_g$. By making use of the mapping f given by equation (3.3) of Definition 16 we may express the equivalent $l_0 \times l_0$ matrices A and B for our current situation. Then by making use of the decompositions of $\alpha_{l_{[k]}l'_{[k]}}$ and $\beta_{l_{[k]}}$ given by equations (3.10) and (3.11) respectively, it is straightforward to see that we may write

$$
\begin{aligned}
A &= \otimes_{q=1}^{k} A^{(q)}; &(3.14)\\
B &= \otimes_{q=1}^{k} B^{(q)}. &(3.15)
\end{aligned}
$$

The notation $\otimes_{q=1}^{k} P^{(q)}$ for conformable matrices $P^{(1)}, \ldots, P^{(k)}$ is used as a shorthand for the direct product $P^{(k)} \otimes \cdots \otimes P^{(1)}$. Notice how the ordering of the $P^{(q)}$ differs from that of the direct sum $\oplus_{q=1}^{k} P^{(q)}$, see equation (2.114).

Thus, from equation (3.7) and using equation (3.14) we may rewrite equation (2.70) as

$$Var(\mathcal{M}(\mathcal{C})) \;=\; A \otimes C \;=\; \{\otimes_{q=1}^{k} A^{(q)}\} \otimes C. \tag{3.16}$$

3.1.1 The examiner revisited

For means of illustration, we return to our examiner example. Suppose that to this point, the examiner has been considering a single exam, when in fact there are four papers to be sat. Each paper is designed to the same format, with the same number of questions and each individual question designed in a comparable way For each paper sat, each of his two markers receive some scripts and he will ask for a sample of n individuals' marks of each paper to be returned to him for checking. Thus, each marker will mark $4n$ scripts and return the marks on these papers to the examiner. Each paper is to be marked anonymously so that it is not possible to detect the four papers of each individual (the examiner could find this out, but the total number of students is sufficiently large that he is unwilling to invest the time and effort in working this out. He assumes this will also apply to the markers when they mark the scripts as they are coded differently for each paper and not collected together in any particular order). Thus, only the individual question marks on a fixed paper are known to be from the same individual. From past experience, the examiner knows that students score more highly on some types of exam than others and he believes this is more a feature of the exam, so that he does not expect it to affect the two markers in different ways. The examiner lets X_{tgvi} denote the mark of the ith individual marked by the gth marker on the vth question of the tth paper. Thus, there are two classification criteria, namely the marker, with two levels, and the paper, with four levels. Thus, $l_0 = 8$. The cells are of the form $l_{[2]} = (t, g) = tg$.

The examiner asserts that a specification as given by equations (3.7) - (3.9) is appropriate, with equations (3.10), (3.11) also being valid. Thus, the examiner must specify the 4×4 matrices $A^{(2)}$ and $B^{(2)}$, and the 2×2 matrices $A^{(1)}$, $B^{(1)}$, C and E. To keep the similarity with the development of the previous chapter, the examiner judges that:

$$A^{(1)} = \begin{pmatrix} 1 & \gamma \\ \gamma & 1 \end{pmatrix};\ B^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \tag{3.17}$$

$$C = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix};\ E = \begin{pmatrix} 4 & 3.5 \\ 3.5 & 4 \end{pmatrix}. \tag{3.18}$$

which should be immediately compared with the specification for the single exam paper of Chapter 2, see equation (2.117). The examiner also judges that

$$A^{(2)} = \begin{pmatrix} 0.98 & 0.79 & 0.70 & 0.74 \\ 0.79 & 0.87 & 0.69 & 0.72 \\ 0.70 & 0.69 & 0.71 & 0.67 \\ 0.74 & 0.72 & 0.67 & 0.77 \end{pmatrix};\ B^{(2)} = \begin{pmatrix} 0.83 & 0 & 0 & 0 \\ 0 & 0.61 & 0 & 0 \\ 0 & 0 & 0.74 & 0 \\ 0 & 0 & 0 & 0.57 \end{pmatrix}. \tag{3.19}$$

3.1.2 A possible situation where our beliefs may be held

Recall that in Chapter 2, we considered the potential model of scaled exchangeability as one which fitted our belief specifications. A similar model may again provide an example application of our beliefs.

Suppose that we consider that a similar model is appropriate in this multifactor scenario and that we judge, in a similar vein to equation (2.71), that for all $l_{[k]} \in \mathcal{I}$

$$\mathcal{M}(\mathcal{C}_{l_{[k]}}) = \left\{ \prod_{q=1}^{k} \mathcal{A}_{l_q}^{(q)} \right\} \mathcal{M}(X), \tag{3.20}$$

where $\mathcal{A}_{l_q}^{(q)}$, $\mathcal{M}(X)$ are random quantities. Since we have adjusted our quantities to have prior mean zero, then $E(\mathcal{A}_{l_q}^{(q)}\mathcal{M}(X)) = 0$. For each factor q, collect the $\mathcal{A}_{l_q}^{(q)}$s together to form the $l_{q,0} \times 1$ vector $\mathcal{A}^{(q)} = [\mathcal{A}_1^{(q)} \ldots \mathcal{A}_{l_{q,0}}^{(q)}]^T$. $\mathcal{M}(X) = [\mathcal{M}(X_1) \ldots \mathcal{M}(X_{v_0})]^T$ is a $v_0 \times 1$ vector containing v_0 random quantities. Thus, in terms of equation (2.71), the random quantity $\mathcal{A}_{l_{[k]}}$ may be expressed as

$$\mathcal{A}_{l_{[k]}} = \left\{ \prod_{q=1}^{k} \mathcal{A}_{l_q}^{(q)} \right\}, \tag{3.21}$$

for all cells $l_{[k]} \in \mathcal{I}$. Once more we judge that there is no interaction between the quadratic products of the $\mathcal{A}_{l_{[k]}}$s and the $\mathcal{M}(X_v)$s, so that equation (2.72) remains valid. Now suppose that there is also no interaction between the quadratic products at each factor. That is for all $q \in \Delta$; for all $\tilde{\Delta} \subseteq \{\Delta \setminus q\}$; for all $l_{\tilde{\delta}}, l'_{\tilde{\delta}} \in \mathcal{I}_{\tilde{\delta}}$ we judge that

$$Cov(\mathcal{A}_{l_q}^{(q)}\mathcal{A}_{l'_q}^{(q)}, \textstyle\prod_{\tilde{\delta}\in\tilde{\Delta}}\mathcal{A}_{l_{\tilde{\delta}}}^{(\tilde{\delta})}\mathcal{A}_{l'_{\tilde{\delta}}}^{(\tilde{\delta})}) = 0. \tag{3.22}$$

Lemma 15 *If we judge that the model given by equation (3.20) is appropriate and we consider that equations (2.72) and (3.22) are valid specifications then for all* $l_{[k]}, l'_{[k]} \in \mathcal{I}$ *we have:*

$$\begin{aligned} Cov(\mathcal{M}(\mathcal{C}_{l_{[k]}}), \mathcal{M}(\mathcal{C}_{l'_{[k]}})) &= \left[\prod_{q=1}^{k}\left\{Cov(\mathcal{A}^{(q)}_{l_q}, \mathcal{A}^{(q)}_{l'_q}) + E(\mathcal{A}^{(q)}_{l_q})E(\mathcal{A}^{(q)}_{l'_q})\right\}\right] \\ &\quad \left\{Var(\mathcal{M}(X)) + E(\mathcal{M}(X))E(\mathcal{M}(X))^T\right\}. \quad (3.23) \end{aligned}$$

Thus, the specification for $[\mathcal{M}(\mathcal{C})]$ *is determined by the specification of the* $k+1$ *spaces,* $[\mathcal{A}^{(q)}]$ *for each* $q = 1, \dots, k$, *and* $[\mathcal{M}(X)]$.

Proof - Follows in the same way as the proof to Lemma 4. □

Setting for all $q \in \Delta$; for all $l_q, l'_q \in \mathcal{I}_q$; for all $v, w \in \{1, \dots, v_0\}$

$$\alpha^{(q)}_{l_q l'_q} = \left\{Cov(\mathcal{A}^{(q)}_{l_q}, \mathcal{A}^{(q)}_{l'_q}) + E(\mathcal{A}^{(q)}_{l_q})E(\mathcal{A}^{(q)}_{l'_q})\right\}; \quad (3.24)$$

$$c_{vw} = Cov(\mathcal{M}(X_v), \mathcal{M}(X_w)) + E(\mathcal{M}(X_v))E(\mathcal{M}(X_w)), \quad (3.25)$$

means that for this model we have

$$Var(\mathcal{M}(\mathcal{C})) = \{\otimes_{q=1}^{k} A^{(q)}\} \otimes C, \quad (3.26)$$

which mirrors the specification given by equation (3.16).

Consider that we gauge that the residual components may also be scaled in a similar way. We consider the following model to be valid

$$\mathcal{R}_i(X_{l_{[k]}}) = \mathcal{B}_{l_{[k]}}\mathcal{M}(X); \quad (3.27)$$

$$\mathcal{B}_{l_{[k]}} = \left\{\prod_{q=1}^{k} \mathcal{B}^{(q)}_{l_{[k]}}\right\}, \quad (3.28)$$

where for all $l_{[k]}, l'_{[k]}, \tilde{l}_{[k]}, \tilde{l}'_{[k]} \in \mathcal{I}$; for all v, w, i, j we assign that

$$Cov(\mathcal{B}_{l_{[k]}}\mathcal{B}_{l'_{[k]}}, (\mathcal{R}_{\tilde{l}_{[k]}i}(X_v))^2) = 0; \quad (3.29)$$

$$Cov(\mathcal{R}_{\tilde{l}_{[k]}i}(X_v), \mathcal{R}_{\tilde{l}'_{[k]}j}(X_w)) = \begin{cases} E & \text{if } \tilde{l}_{[k]} = \tilde{l}'_{[k]},\ i = j; \\ 0 & \text{otherwise,} \end{cases} \quad (3.30)$$

and for all $q \in \Delta$; for all $\tilde{\Delta} \subseteq \{\Delta \setminus q\}$; for all $l_{\tilde{\delta}}, l'_{\tilde{\delta}} \in \mathcal{I}_{\tilde{\delta}}$ we judge that

$$Cov(\mathcal{B}^{(q)}_{l_q}\mathcal{B}^{(q)}_{l'_q}, \textstyle\prod_{\tilde{\delta}\in\tilde{\Delta}} \mathcal{B}^{(\tilde{\delta})}_{l_{\tilde{\delta}}}\mathcal{B}^{(\tilde{\delta})}_{l'_{\tilde{\delta}}}) = 0. \quad (3.31)$$

We collect these statements and their consequences together to form the following lemma.

Lemma 16 *If we assert that the model given by equation (3.27) is appropriate and we consider that equations (3.29), (3.30) and (3.31) are valid specifications, then for all $l_{[k]}, l'_{[k]} \in \mathcal{I}$; for all i, j we have:*

$$Cov(\mathcal{R}_i(X_{l_{[k]}}), \mathcal{R}_j(X_{l'_{[k]}})) = \begin{cases} [\prod_{q=1}^{k}\{Var(\mathcal{B}^{(q)}_{l_q}) + (\mathcal{B}^{(q)}_{l_q})^2\}]E & \textit{if } l_{[k]} = l'_{[k]}, i = j; \\ 0 & \textit{otherwise.} \end{cases} \tag{3.32}$$

Thus, the specification for each $[\mathcal{R}_i(X_{l_{[k]}})]$ is determined by the specification of the spaces, $[\mathcal{B}^{(q)}]$ for each $q \in \Delta$, and each $[\mathcal{R}_{l_{[k]}i}(X)]$.

Setting $\beta^{(q)}_{l_q} = \{Var(\mathcal{B}^{(q)}_{l_q}) + E(\mathcal{B}^{(q)}_{l_q})^2\}$ then we have for this model that

$$Cov(\mathcal{R}_i(X_{l_{[k]}}), \mathcal{R}_j(X_{l'_{[k]}})) = \begin{cases} [\prod_{q=1}^{k}\beta^{(q)}_{l_q}]E & \text{if } l_{[k]} = l'_{[k]}, i = j; \\ 0 & \text{otherwise.} \end{cases} \tag{3.33}$$

which, by substituting equation (3.11) into equation (3.9), we observe conforms with our specified model of interest.

Notice the scope that this specification and interpretation allows us. Recall that in Chapter 2 we showed that if we judged the $\mathcal{A}_g$s to be exchangeable, then we could derive a one-way layout multivariate analysis of variance model. We can derive similar models in this multifactor scenario. For example, suppose that the number of factors is two and that we judge that $\mathcal{A}^{(2)}_{l_2}$ is second-order exchangeable over levels, so that for all $l_2, l'_2 \in \mathcal{I}_2$, $l_2 \neq l'_2$ we have

$$E(\mathcal{A}^{(2)}_{l_2}) = m^{(2)}; \tag{3.34}$$

$$Var(\mathcal{A}^{(2)}_{l_2}) + (m^{(2)})^2 = \alpha^{(2)}; \tag{3.35}$$

$$Cov(\mathcal{A}^{(2)}_{l_2}, \mathcal{A}^{(2)}_{l'_2}) + (m^{(2)})^2 = \gamma^{(2)}, \tag{3.36}$$

and that $\mathcal{A}^{(1)}_{l_1}$ is also second-order exchangeable over levels, so that for all $l_1, l'_1 \in \mathcal{I}_1$, $l_1 \neq l'_1$ we have

$$E(\mathcal{A}^{(1)}_{l_1}) = m^{(1)}; \tag{3.37}$$

$$Var(\mathcal{A}^{(1)}_{l_1}) + (m^{(1)})^2 = \alpha^{(1)}; \tag{3.38}$$

$$Cov(\mathcal{A}^{(1)}_{l_1}, \mathcal{A}^{(1)}_{l'_1}) + (m^{(1)})^2 = \gamma^{(1)}. \tag{3.39}$$

Hence, for each q, the matrix $A^{(q)}$ has the form

$$A^{(q)} = \alpha^{(q)} I_{l_{q,0}} + \gamma^{(q)}(J_{l_{q,0}} - I_{l_{q,0}}). \tag{3.40}$$

We have the following theorem

Theorem 20 *If $k = 2$ and we consider that both $\mathcal{A}^{(2)}_{l_2}$ and $\mathcal{A}^{(1)}_{l_1}$ are second-order exchangeable over levels with second-order specifications given by equations (3.35) - (3.39), then for all $l_{[2]} \in \mathcal{I}$ we may introduce the further collections of random quantities $\tilde{\mathcal{M}}(X)$, $\mathcal{P}(X) = \{\mathcal{P}(X_1), \dots, \mathcal{P}(X_{l_{2,0}})\}$, $\mathcal{Q}(X) = \{\mathcal{Q}(X_1), \dots, \mathcal{Q}(X_{l_{1,0}})\}$ and $\mathcal{R}(X) = \{\mathcal{R}(X_{11}), \dots, \mathcal{R}(X_{l_{2,0}l_{1,0}})\}$ and write*

$$\mathcal{M}(X_{l_{[2]}}) = \tilde{\mathcal{M}}(X) + \mathcal{P}(X_{l_2}) + \mathcal{Q}(X_{l_1}) + \mathcal{R}(X_{l_2 l_1}), \tag{3.41}$$

where

$$\tilde{\mathcal{M}}(X) = \frac{1}{l_{2,0}l_{1,0}} \sum_{l_2=1}^{l_{2,0}} \sum_{l_1=1}^{l_{1,0}} \mathcal{M}(X_{l_{[2]}}); \tag{3.42}$$

$$\mathcal{P}(X_{l_2}) = \frac{1}{l_{1,0}} \sum_{l_1=1}^{l_{1,0}} \mathcal{M}(X_{l_{[2]}}) - \frac{1}{l_{2,0}l_{1,0}} \sum_{l_2=1}^{l_{2,0}} \sum_{l_1=1}^{l_{1,0}} \mathcal{M}(X_{l_{[2]}}); \tag{3.43}$$

$$\mathcal{Q}(X_{l_1}) = \frac{1}{l_{2,0}} \sum_{l_2=1}^{l_{2,0}} \mathcal{M}(X_{l_{[2]}}) - \frac{1}{l_{2,0}l_{1,0}} \sum_{l_2=1}^{l_{2,0}} \sum_{l_1=1}^{l_{1,0}} \mathcal{M}(X_{l_{[2]}}); \tag{3.44}$$

$$\mathcal{R}(X_{l_2 l_1}) = \mathcal{M}(X_{l_{[2]}}) - \frac{1}{l_{2,0}} \sum_{l_2=1}^{l_{2,0}} \mathcal{M}(X_{l_{[2]}}) - \frac{1}{l_{1,0}} \sum_{l_1=1}^{l_{1,0}} \mathcal{M}(X_{l_{[2]}}) + \frac{1}{l_{2,0}l_{1,0}} \sum_{l_2=1}^{l_{2,0}} \sum_{l_1=1}^{l_{1,0}} \mathcal{M}(X_{l_{[2]}}). \tag{3.45}$$

The collections $\tilde{\mathcal{M}}(X)$, $\mathcal{P}(X)$, $\mathcal{Q}(X)$ and $\mathcal{R}(X)$ are mutually uncorrelated and

$$Var(\tilde{\mathcal{M}}(X)) = \left\{ \frac{(\alpha^{(2)} + (l_{2,0}-1)\gamma^{(2)})(\alpha^{(1)} + (l_{1,0}-1)\gamma^{(1)})}{l_{2,0}l_{1,0}} \right\} C; \tag{3.46}$$

$$Cov(\mathcal{P}(X_{l_2}), \mathcal{P}(X_{l'_2})) = \begin{cases} \left\{ \frac{(l_{2,0}-1)(\alpha^{(2)}-\gamma^{(2)})(\alpha^{(1)}+(l_{1,0}-1)\gamma^{(1)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_2 = l'_2; \\ -\left\{ \frac{(\alpha^{(2)}-\gamma^{(2)})(\alpha^{(1)}+(l_{1,0}-1)\gamma^{(1)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_2 \neq l'_2; \end{cases} \tag{3.47}$$

$$Cov(\mathcal{Q}(X_{l_1}), \mathcal{Q}(X_{l'_1})) = \begin{cases} \left\{ \frac{(l_{1,0}-1)(\alpha^{(1)}-\gamma^{(1)})(\alpha^{(2)}+(l_{2,0}-1)\gamma^{(2)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_1 = l'_1; \\ -\left\{ \frac{(\alpha^{(1)}-\gamma^{(1)})(\alpha^{(2)}+(l_{2,0}-1)\gamma^{(2)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_1 \neq l'_1; \end{cases} \tag{3.48}$$

$$Cov(\mathcal{R}(X_{l_{[2]}}), \mathcal{R}(X_{l'_{[2]}})) = \begin{cases} \left\{ \frac{(l_{2,0}-1)(l_{1,0}-1)(\alpha^{(2)}-\gamma^{(2)})(\alpha^{(1)}-\gamma^{(1)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_{[2]} = l'_{[2]}; \\ \left\{ -\frac{(l_{2,0}-1)(\alpha^{(2)}-\gamma^{(2)})(\alpha^{(1)}-\gamma^{(1)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_2 = l'_2,\ l_1 \neq l'_1; \\ \left\{ -\frac{(l_{1,0}-1)(\alpha^{(2)}-\gamma^{(2)})(\alpha^{(1)}-\gamma^{(1)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_2 \neq l'_2,\ l_1 = l'_1; \\ \left\{ \frac{(\alpha^{(2)}-\gamma^{(2)})(\alpha^{(1)}-\gamma^{(1)})}{l_{2,0}l_{1,0}} \right\} C & \text{if } l_2 \neq l'_2,\ l_1 \neq l'_1. \end{cases} \tag{3.49}$$

Proof - The second-order specification induced over the quantities given by equations (3.42) - (3.45) may be easily verified through writing then in matrix terms as

$$\begin{aligned}
\tilde{\mathcal{M}}(X) &= \frac{1}{l_{2,0}l_{1,0}}(1_{l_{2,0}} \otimes 1_{l_{1,0}} \otimes I_{v_0})^T \mathcal{M}(\mathcal{C}); & (3.50)\\
\mathcal{P}(X_{l_2}) &= \frac{1}{l_{1,0}}(\epsilon_{l_{2,0}l_2} \otimes 1_{l_{1,0}} \otimes I_{v_0})^T \mathcal{M}(\mathcal{C}) - \tilde{\mathcal{M}}(X); & (3.51)\\
\mathcal{Q}(X_{l_1}) &= \frac{1}{l_{2,0}}(1_{l_{2,0}} \otimes \epsilon_{l_{1,0}l_1} \otimes I_{v_0})^T \mathcal{M}(\mathcal{C}) - \tilde{\mathcal{M}}(X); & (3.52)\\
\mathcal{R}(X_{l_2l_1}) &= (\epsilon_{l_{2,0}l_2} \otimes \epsilon_{l_{1,0}l_1} \otimes I_{v_0})^T \mathcal{M}(\mathcal{C}) - \mathcal{Q}(X_{l_1}) - \mathcal{P}(X_{l_2}) - \tilde{\mathcal{M}}(X). & (3.53)
\end{aligned}$$

and then using the specification for $Var(\mathcal{M}(\mathcal{C}))$ given by equation (3.26) and the representation of $A^{(q)}$ given by equation (3.40). □

Notice then that our full representation for the model is

$$X_{l_{[2]}i} = \tilde{\mathcal{M}}(X) + \mathcal{P}(X_{l_2}) + \mathcal{Q}(X_{l_1}) + \mathcal{R}(X_{l_2l_1}) + \mathcal{R}_i(X_{l_{[2]}}), \qquad (3.54)$$

which may be compared to a two-way layout multivariate analysis of variance model. Notice that a judgement of $\alpha^{(1)} = \gamma^{(1)}$ corresponds to $Cov(\mathcal{Q}(X_{l_1}), \mathcal{Q}(X_{l_1'})) = 0$ and $Cov(\mathcal{R}(X_{l_2l_1}), \mathcal{R}(X_{l_2'l_1'})) = 0$ for all $l_{[2]}, l_{[2]}' \in \mathcal{I}$ and the model reduces to a model comparable to a one-way layout. Observe also that if we only judge $\mathcal{A}_{l_1}^{(1)}$ to be second-order exchangeable over groups, then we may construct the corresponding one-way layout model. If we now restore the number of factors to k, it can be shown that assessments of second-order exchangeability for p of the k collections $\mathcal{A}^{(q)}$ will allow the construction of a model that may be compared to a p-way layout multivariate analysis of variance model.

Observe that the assessment of the collection $\mathcal{A}^{(q)}$ as being second-order exchangeable implies that we have judged that $Cov(\mathcal{M}(X_{l_{[k]}}), \mathcal{M}(X_{l_{[k]}'}))$ is invariant for the qth factor settings. Thus a relationship may be made with work on invariant covariance models of Consonni & Dawid (1985) and Dawid (1988). Indeed, we may derive the model given by equation (3.54) by considering various invariance properties of $Cov(\mathcal{A}_{l_{[k]}}, \mathcal{A}_{l_{[k]}'})$ without using equation (3.21).

We emphasise once more, that this model merely serves as an example of when our beliefs may hold. There is nothing in our beliefs to force this model. It should, perhaps, also be emphasised here that, even if we judge this model to be applicable, the results that we shall develop do not rely upon the exchangeability of the $\mathcal{A}^{(q)}$ (or the covariance invariance). Indeed, if we adopt the scaled exchangeability perspective to the examiner example, then the examiner has judged $\mathcal{A}^{(1)}$ to be exchangeable (so that the marker scalings are second-order exchangeable), but by observing equation (3.19), we observe that the marker has not judged the paper scalings to be second-order exchangeable.

3.2 Adjusting the mean components

We want to observe a sample of $n_{l_{[k]}} > 0$ individuals from the $l_{[k]}$th cell for each $l_{[k]} \in \mathcal{I}$. Using the mapping $f(l_{[k]}) = g$ given in equation (3.3), in the gth cell we want to observe the n_g exchangeable collections $\mathcal{C}_{g1}, \ldots, \mathcal{C}_{gn_g}$. The complete sample is the gth cell is collected together as $\mathcal{C}_g(n_g) = \cup_{g=1}^{n_g} \mathcal{C}_{gi}$. The total sample is $\mathcal{C}(N) = \cup_{g=1}^{l_0} \mathcal{C}_g(n_g)$. N denotes the collection of sample sizes, which we also denote by the $l_0 \times l_0$ diagonal matrix, $N = diag(n_1, \ldots, n_{l_0})$. We want to use these observations to revise our beliefs over the mean collection corresponding to the cells, namely

$$\mathcal{M}(\mathcal{C}) \;=\; \cup_{l_{[k]} \in \mathcal{I}} \mathcal{M}(\mathcal{C}_{l_{[k]}}) \;=\; \cup_{g=1}^{l_0} \mathcal{M}(\mathcal{C}_g). \tag{3.55}$$

By appealing to Theorem 8, we have that the collection of sample means, $\mathcal{S}_N(\mathcal{C}) = \cup_{g=1}^{l_0} \mathcal{S}_{n_g}(\mathcal{C}_g)$, are Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting $[\mathcal{M}(\mathcal{C})]$. Once more, we shall make use of the vector representations of $\mathcal{M}(\mathcal{C})$, $\mathcal{C}(N)$ and $\mathcal{S}_N(\mathcal{C})$, that is $\mathcal{M}(\mathcal{C}) = [\mathcal{M}(\mathcal{C}_1)^T \ldots \mathcal{M}(\mathcal{C}_{l_0})]^T$; $\mathcal{C}(N) = [\mathcal{C}_1(n_1)^T \ldots \mathcal{C}_{l_0}(n_{l_0})^T]^T$; $\mathcal{S}_N(\mathcal{C}) = [\mathcal{S}_{n_1}(\mathcal{C}_1)^T \ldots \mathcal{S}_{n_{l_0}}(\mathcal{C}_{l_0})^T]^T$.

Currently we have placed no restriction upon the type of sample we could take. However, we shall proceed by assuming that the sample sizes are constrained by having to have the following form

$$n_{l_{[k]}} \;=\; \prod_{q=1}^{k} n_{l_q}^{(q)}, \tag{3.56}$$

so that the sample size for each cell is determined by the levels of the factors of that group. Hence, if $k = 3$ then $n_{pqr} = n_p n_q n_r$. Notice that equation (3.56) includes, but allows scope beyond, the balanced design.

An interpretation of this choice could be made by considering allocating individuals in the whole population to the cells, where the factors are judged to be independent for this allocation. Thus, we could regard our table of cells for this purpose as a contingency table. As is well known, see for example Everitt (1992), the expected number allocated to each cell has a maximum likelihood estimate of a form comparable to equation (3.56). It should be noted that whilst this choice of sample will not be optimal for all choices of design, for example if we wanted to specifically learn about a certain cell, the results that we develop will enable us to provide upper bounds for quantities such as variance reduction and these bounds may be found over a wider sample space than the balanced design. Notice that the examiner example naturally breaks down into this form. In the balanced case, we may take $n_t^{(2)} = 1$ for each of the four papers, and $n_g^{(1)} = n$ for each of the two markers. Notice that if the examiner asks for n_1 marks on each paper to be returned from the first marker and $n_2 \neq n_1$ marks on each paper to be returned from the second marker, then a similar breakdown occurs, by taking $n_t^{(2)} = 1$ and $n_g^{(1)} = n_g$.

Let $N^{(q)}$ be the $l_{q,0} \times l_{q,0}$ diagonal matrix with (l,l)th entry $(N^{(q)})_{ll} = n_l^{(q)}$. Then, by observing the mapping $f(l_{[k]}) = g$ given in equation (3.3), we may write

$$N = \otimes_{q=1}^{k} N^{(q)}. \tag{3.57}$$

Thus, in the examiner example, we have $N^{(2)} = I_4$ and $N^{(1)} = nI_2$. N is then equal to nI_8 as required.

By utilising equations (3.14), (3.15) and (3.57), we may write Lemma 6 as

Lemma 17 *The second-order specifications for the mean components, $\mathcal{M}(\mathcal{C})$, and the sample means, $\mathcal{S}_N(\mathcal{C})$, may be expressed as*

$$\begin{aligned} Var(\mathcal{M}(\mathcal{C})) &= \{\otimes_{p=1}^{k} A^{(p)}\} \otimes C; & (3.58)\\ Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) &= \{\otimes_{p=1}^{k} A^{(p)}\} \otimes C; & (3.59)\\ Var(\mathcal{S}_N(\mathcal{C})) &= \{\otimes_{p=1}^{k} A^{(p)}\} \otimes C + \{\otimes_{p=1}^{k} (N^{(p)})^{-1} B^{(p)}\} \otimes E. & (3.60) \end{aligned}$$

We will occasionally make use of the following notational devices. For all $q = 1, \dots, k$, we may write

$$\begin{aligned} C^{(q)} &= \{\otimes_{p=1}^{q} A^{(p)}\} \otimes C; & (3.61)\\ E^{(q)} &= \{\otimes_{p=1}^{q} (N^{(p)})^{-1} B^{(p)}\} \otimes E, & (3.62) \end{aligned}$$

so that, for example, we may write

$$Var(\mathcal{S}_N(\mathcal{C})) = C^{(k)} + E^{(k)}. \tag{3.63}$$

If $\tilde{\Delta} \subseteq \Delta$, which we may express as $\tilde{\Delta} = \{\tilde{\delta}_1, \dots, \tilde{\delta}_r\}$, where $\tilde{\delta}_1 < \cdots < \tilde{\delta}_r$, then we adopt the following shorthands

$$\begin{aligned} C^{(\tilde{\Delta})} &= \{\otimes_{p=1}^{r} A^{(\tilde{\delta}_p)}\} \otimes C; & (3.64)\\ E^{(\tilde{\Delta})} &= \{\otimes_{p=1}^{r} (N^{(\tilde{\delta}_p)})^{-1} B^{(\tilde{\delta}_p)}\} \otimes E. & (3.65) \end{aligned}$$

3.3 Underlying canonical variable and group problems

In Theorem 13, we saw that the solution of the underlying canonical variable and group problems completely determined the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$. We adjust the notation to the present notion of cells by interpretation of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$. We firstly use the underlying variable problem to find the canonical directions for the adjustment of the mean components, $\mathcal{M}(\mathcal{C}_{l_{[k]}})$, of the $l_{[k]}$th cell by a sample of $n_{l_{[k]}}$ individuals drawn from that cell. From equation (2.97) of Theorem 11, for each $s = 1, \ldots, v_0$, we have

$$Z_{[\mathcal{M}(\mathcal{C}_{l_{[k]}})/\mathcal{C}_{l_{[k]}}(n_{l_{[k]}})]s} = \sqrt{\frac{1}{\alpha_{l_{[k]}l_{[k]}}}} Y_s^T \mathcal{M}(\mathcal{C}_{l_{[k]}}). \tag{3.66}$$

For each s, we then collect all the corresponding directions across the cells to form the collection $\mathcal{Z}_{(N)s}$, so that

$$\mathcal{Z}_{(N)s} = \{Z_{[\mathcal{M}(\mathcal{C}_1)/\mathcal{C}_1(n_1)]s}, \ldots, Z_{[\mathcal{M}(\mathcal{C}_{l_0})/\mathcal{C}_{l_0}(n_{l_0})]s}\}. \tag{3.67}$$

We may equally express $\mathcal{Z}_{(N)s}$ as the $l_0 \times 1$ vector

$$\mathcal{Z}_{(N)s} = D(I_{l_0} \otimes Y_s^T)\mathcal{M}(\mathcal{C}), \tag{3.68}$$

which should be compared to equation (2.126), where we now have

$$D = \otimes_{p=1}^{k} D^{(p)}; \tag{3.69}$$

$$D^{(p)} = diag\left(\sqrt{\frac{1}{\alpha_{11}^{(p)}}}, \ldots, \sqrt{\frac{1}{\alpha_{l_{p,0}l_{p,0}}^{(p)}}}\right), \tag{3.70}$$

as the equivalent definition of D, to that given in equation (2.125). The canonical directions and resolutions for the adjustment of each $[\mathcal{Z}_{(N)s}]$ by $\mathcal{C}(N)$, for $s = 1, \ldots, v_0$, are found from the solution of the underlying canonical variable problem, as shown in Theorem 12. When collected together, Theorem 13 shows that these are the canonical directions and resolutions for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$. With our current specifications, we shall show that we may split the underlying canonical group problem down into a series of subproblems. We make the following definition.

Definition 17 *The underlying canonical qth factor directions are defined as the columns of the matrix $W^{(q)} = [W_1^{(q)} \ldots W_{l_{q,0}}^{(q)}]$ solving the generalised eigenvalue problem*

$$A^{(q)}W^{(q)} = (A^{(q)} + (N^{(q)})^{-1}B^{(q)})W^{(q)}\Psi^{(q)}, \tag{3.71}$$

where $\Psi^{(q)} = diag(\psi_1^{(q)}, \ldots, \psi_{l_{q,0}}^{(q)})$ is the matrix of eigenvalues. $W^{(q)}$ is chosen so that $(W^{(q)})^T A^{(q)} W^{(q)} = I_{l_{q,0}}$, $(W^{(q)})^T(A^{(q)} + (N^{(q)})^{-1}B^{(q)})W^{(q)}\Psi^{(q)} = I_{l_{q,0}}$. The ordered eigenvalues $1 > \psi_1^{(q)} \geq \ldots \geq \psi_{l_{q,o}}^{(q)} > 0$ are termed the underlying canonical qth factor resolutions.

To motivate this definition, write (q, l) to denote the cell

$$f_{[k;q+1]} l f_{[q-1;1]} = (f_k, \ldots, f_{q+1}, l, f_{q-1}, \ldots, f_1), \tag{3.72}$$

and, for each $s = 1, \ldots, v_0$, form the collection

$$\mathcal{Z}_{(N^{(q)})s} = \{Z_{[\mathcal{M}(\mathcal{C}_{(q,1)})/\mathcal{C}_{(q,1)}(n_{(q,1)})]s}, \ldots, Z_{[\mathcal{M}(\mathcal{C}_{(q,l_{q,0})})/\mathcal{C}_{(q,l_{q,0})}(n_{(q,l_{q,0})})]s}\}, \tag{3.73}$$

so that $\mathcal{Z}_{(N^{(q)})s}$ is the collection of the s canonical directions for the adjustment of each $[\mathcal{M}(\mathcal{C}_{(q,l)})]$ by $\mathcal{C}_{(q,l)}(n_{(q,l)})$ where $l = 1, \ldots, l_{q,0}$. Thus, we collect together the directions corresponding to the cells where only the qth factor varies, each remaining factor being held at a fixed level. In vector terms, we have that

$$\mathcal{Z}_{(N^{(q)})s} = M_{(q)} \mathcal{Z}_{(N)s} = M_{(q)} D(I_{l_0} \otimes Y_s^T) \mathcal{M}(\mathcal{C}), \tag{3.74}$$

where

$$M_{(q)} = \{\otimes_{p=q+1}^{k} \epsilon_{l_{p,0} f_p}^T\} \otimes I_{l_{q,0}} \otimes \{\otimes_{p=1}^{q-1} \epsilon_{l_{p,0} f_p}^T\}. \tag{3.75}$$

We want to consider using the sample corresponding to the cells (q, l) for each $l \in \mathcal{I}_q$ to learn about $\mathcal{Z}_{(N^{(q)})s}$. We represent the total sample by $\mathcal{C}(N^{(q)}) = \cup_{l=1}^{l_{q,0}} \mathcal{C}(n_{(q,l)})$, and by $\mathcal{S}_{N^{(q)}}(\mathcal{C}) = \cup_{l=1}^{l_{q,0}} \mathcal{S}_{n_{(q,l)}}(\mathcal{C}_{(q,l)})$ we represent the corresponding collection of sample means. In vector form, we have that

$$\mathcal{S}_{N^{(q)}}(\mathcal{C}) = (M_{(q)} \otimes I_{v_0}) \mathcal{S}_N(\mathcal{C}). \tag{3.76}$$

Lemma 18 *The second-order relationships between $\mathcal{Z}_{(N^{(q)})s}$ and $\mathcal{S}_{N^{(q)}}(\mathcal{C})$ may be expressed as follows*

$$\begin{aligned}
Var(\mathcal{Z}_{(N^{(q)})s}) &= D^{(q)} A^{(q)} D^{(q)}; && (3.77)\\
Cov(\mathcal{Z}_{(N^{(q)})s}, \mathcal{S}_{N^{(q)}}(\mathcal{C})) &= D^{(q)}[\epsilon_{v_0 s}^T \otimes \sqrt{a} A^{(q)}] \times \\
&\quad [(I_{l_{q,0}} \otimes Y_1) \ldots (I_{l_{q,0}} \otimes Y_{v_0})]^{-1}; && (3.78)\\
Var(\mathcal{S}_{N^{(q)}}(\mathcal{C})) &= [I_{l_{q,0}} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{l_{q,0}} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{l_{q,0}} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0} \{\phi_s a A^{(q)} + (1-\phi_s) b (N^{(q)})^{-1} B^{(q)}\}] \times \\
&\quad [(I_{l_{q,0}} \otimes Y_1) \ldots (I_{l_{q,0}} \otimes Y_{v_0})]^{-1}; && (3.79)\\
Cov(\mathcal{S}_{N^{(q)}}(\mathcal{C}), \mathcal{Z}_{(N^{(q)})s}) &= [I_{l_{q,0}} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{l_{q,0}} \otimes \epsilon_{v_0 1}^T \\ \vdots \\ I_{l_{q,0}} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\epsilon_{v_0 s} \otimes \{\phi_s \sqrt{a} A^{(q)}\}] D^{(q)}, && (3.80)
\end{aligned}$$

where

$$a = \prod_{p=1,p\neq q}^{k} \alpha_{f_pf_p}^{(p)}; \tag{3.81}$$

$$b = \prod_{p=1,p\neq q}^{k} \frac{\beta_{f_p}^{(p)}}{n_{f_p}^{(p)}}. \tag{3.82}$$

Proof - Equation (3.77) follows from equation (2.127) since

$$\begin{aligned}
Var(\mathcal{Z}_{(N^{(q)})s}) &= M_{(q)} Var(\mathcal{Z}_{(N)s}) M_{(q)}^T && (3.83)\\
&= M_{(q)} \{\otimes_{p=1}^k D^{(p)} A^{(p)} D^{(p)}\} M_{(q)}^T && (3.84)\\
&= \{\otimes_{p\neq q}^k \epsilon_{l_p,0f_p}^T D^{(p)} A^{(p)} D^{(p)} \epsilon_{l_p,0f_p}\} D^{(q)} A^{(q)} D^{(q)} && (3.85)\\
&= D^{(q)} A^{(q)} D^{(q)}. && (3.86)
\end{aligned}$$

From equations (3.74) and (3.76) we have

$$\begin{aligned}
Cov(\mathcal{Z}_{(N^{(q)})s}, \mathcal{S}_{N^{(q)}}(\mathcal{C})) &= M_{(q)} Cov(\mathcal{Z}_{(N)s}, \mathcal{S}_N(\mathcal{C}))(M_{(q)}^T \otimes I_{v_0}) && (3.87)\\
&= M_{(q)} D(I_{l_0} \otimes Y_s^T) Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C}))(M_{(q)}^T \otimes I_{v_0}) && (3.88)\\
&= M_{(q)} D(I_{l_0} \otimes Y_s^T)[\{\otimes_{p=1}^k A^{(p)}\} \otimes C](M_{(q)}^T \otimes I_{v_0}) && (3.89)\\
&= \{\otimes_{p\neq q}^k \epsilon_{l_p,0f_p}^T D^{(p)} A^{(p)} \epsilon_{l_p,0f_p}\} D^{(q)} [A^{(q)} \otimes Y_s^T C] && (3.90)\\
&= D^{(q)} [\sqrt{a} A^{(q)} \otimes Y_s^T C]. && (3.91)
\end{aligned}$$

Equation (3.78) follows by comparing equation (3.91) with equation (2.135).

$$\begin{aligned}
&Var(\mathcal{S}_{N^{(q)}}(\mathcal{C})) = (M_{(q)} \otimes I_{v_0}) Var(\mathcal{S}_N(\mathcal{C}))(M_{(q)}^T \otimes I_{v_0}) && (3.92)\\
&= (M_{(q)} \otimes I_{v_0})[(\{\otimes_{p=1}^k A^{(p)}\} \otimes C) + (\{\otimes_{p=1}^k (N^{(p)})^{-1} B^{(p)}\} \otimes E)] \times \\
&\qquad (M_{(q)}^T \otimes I_{v_0}) && (3.93)\\
&= [(\otimes_{p\neq q}^k \epsilon_{l_p,0f_p}^T A^{(p)} \epsilon_{l_p,0f_p}\} \otimes A^{(q)} \otimes C) + (\otimes_{p\neq q}^k \epsilon_{l_p,0f_p}^T (N^{(p)})^{-1} B^{(p)} \epsilon_{l_p,0f_p}\} \\
&\qquad \otimes (N^{(q)})^{-1} B^{(q)} \otimes E)] && (3.94)\\
&= \{(aA^{(q)} \otimes C) + (b(N^{(p)})^{-1} B^{(p)} \otimes E)\}. && (3.95)
\end{aligned}$$

Equation (3.79) follows by applying Lemma 7 to equation (3.95) with $S = aA^{(q)}$, $T = b(N^{(p)})^{-1}B^{(p)}$ and $X_t = I_{l_{q,0}}$ for all $t = 1, \dots, v_0$.

$$\begin{aligned}
Cov(\mathcal{S}_{N^{(q)}}(\mathcal{C}), \mathcal{Z}_{(N^{(q)})s}) &= (M_{(q)} \otimes I_{v_0}) Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{Z}_{(N)s}) M_{(q)}^T && (3.96)\\
&= (M_{(q)} \otimes I_{v_0}) Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}))(I_{l_0} \otimes Y_s) D^T M_{(q)}^T && (3.97)\\
&= [\{\otimes_{p\neq q}^k \epsilon_{l_p,0f_p}^T A^{(p)} D^{(p)} \epsilon_{l_p,0f_p}\} A^{(q)} \otimes CY_s] D^{(q)} && (3.98)\\
&= (\sqrt{a} A^{(q)} \otimes C)(I_{l_{q,0}} \otimes Y_s) D^{(q)}, && (3.99)
\end{aligned}$$

and equation (3.80) follows by comparing equation (3.99) with equation (2.141). □

Denote the resolution transform for the adjustment of $[\mathcal{Z}_{(N^{(q)})s}]$ by $\mathcal{C}(N^{(q)})$ as $T_{[\mathcal{Z}_{(N^{(q)})s}/\mathcal{C}(N^{(q)})]}$. We have the following theorem.

Theorem 21 *For the adjustment of $[\mathcal{Z}_{(N^{(q)})s}]$ by $\mathcal{C}(N^{(q)})$, the resolution transform matrix is calculated as*

$$T_{[\mathcal{Z}_{(N^{(q)})s}/\mathcal{C}(N^{(q)})]} = (D^{(q)})^{-1}\{\phi_s a A^{(q)} + (1-\phi_s)b(N^{(q)})^{-1}B^{(q)}\}\{\phi_s a A^{(q)}\}D^{(q)} \quad (3.100)$$

The canonical directions are given by

$$Z_{(N^{(q)})ds} = \{(D^{(q)})^{-1}W_d^{(q)}\}^T \mathcal{Z}_{(N^{(q)})s}, \quad (3.101)$$

for each $d = 1, \dots, l_{q,0}$ with corresponding canonical resolutions given by

$$\lambda_{(N^{(q)})ds} = \frac{a\psi_d^{(q)}\phi_s}{a\psi_d^{(q)}\phi_s + b(1-\psi_d^{(q)})(1-\phi_s)}. \quad (3.102)$$

$W_d^{(q)}$ is the dth underlying canonical qth factor direction as given in Definition 17; $\psi_d^{(q)}$ is the corresponding dth underlying canonical qth factor resolution.

Proof - Notice that for each s, $[\mathcal{Z}_{(N^{(q)})s}] \subset [\mathcal{M}(\mathcal{C})]$. It thus follows immediately, in an identical fashion to Theorem 8, that $\mathcal{S}_{N^{(q)}}(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N^{(q)})$ for adjusting $[\mathcal{Z}_{(N^{(q)})s}]$. Thus, $T_{[\mathcal{Z}_{(N^{(q)})s}/\mathcal{C}(N^{(q)})]} = T_{[\mathcal{Z}_{(N^{(q)})s}/\mathcal{S}_{N^{(q)}}(\mathcal{C})]}$. By inverting equation (3.77), post multiplying by equation (3.78) and then the inversion of (3.79) and finally by (3.80) and using equation (1.74), we have that

$$\begin{aligned} T_{[\mathcal{Z}_{(N^{(q)})s}/\mathcal{S}_{N^{(q)}}(\mathcal{C})]} &= \\ &(D^{(q)})^{-1}\{\phi_s a A^{(q)} + (1-\phi_s)b(N^{(q)})^{-1}B^{(q)}\}\{\phi_s a A^{(q)}\}D^{(q)}, \end{aligned} \quad (3.103)$$

from which equation (3.100) follows immediately. From the solution of the underlying group problem as given by equation (3.71), it is straightforward to see that

$$\begin{aligned} T_{[\mathcal{Z}_{(N^{(q)})s}/\mathcal{S}_{N^{(q)}}(\mathcal{C})]} &= (D^{(q)})^{-1}W^{(q)}\{\phi_s a\Psi^{(q)}\}\times \\ &\{\phi_s a\Psi^{(q)} + (1-\phi_s)b(I_{l_{q,0}} - \Psi^{(q)})\}^{-1}(W^{(q)})^{-1}D^{(q)}, \end{aligned} \quad (3.104)$$

so that

$$\begin{aligned} T_{[\mathcal{Z}_{(N^{(q)})s}/\mathcal{S}_{N^{(q)}}(\mathcal{C})]}((D^{(q)})^{-1}W^{(q)}) &= \\ &(D^{(q)})^{-1}W^{(q)}\{\phi_s a\Psi^{(q)}\}\{\phi_s a\Psi^{(q)} + (1-\phi_s)b(I_{l_{q,0}} - \Psi^{(q)})\}^{-1}. \end{aligned} \quad (3.105)$$

Hence, $\Lambda_{(N^{(q)})s} = \{\phi_s a\Psi^{(q)}\}\{\phi_s a\Psi^{(q)} + (1-\phi_s)b(I_{l_{q,0}} - \Psi^{(q)})\}^{-1}$ is the matrix whose diagonal elements are the canonical resolutions of the adjustment. To confirm that the $((D^{(q)})^{-1}W_d^{(q)})^T\mathcal{Z}_{(N^{(q)})_s}$ are the corresponding canonical resolutions, we verify that they are mutually uncorrelated with prior variance one. Notice that by making use of equation (3.77), we have that

$$\begin{aligned} Cov(((D^{(q)})^{-1}W_d^{(q)})^T\mathcal{Z}_{(N^{(q)})_s}, ((D^{(q)})^{-1}W_{d'}^{(q)})^T\mathcal{Z}_{(N^{(q)})_s}) &= \\ (W_d^{(q)})^T(D^{(q)})^{-1}Var(\mathcal{Z}_{(N^{(q)})_s})(D^{(q)})^{-1}(W_{d'}^{(q)}) & \quad (3.106) \\ = \ (W_d^{(q)})^T A^{(q)}(W_{d'}^{(q)}) & \quad (3.107) \\ = \ \delta_{dd'}, & \quad (3.108) \end{aligned}$$

where equation (3.107) follows from the choice of $W^{(q)}$ in Definition 17. □

Thus, we see how the underlying qth factor problem can be directly associated to a learning situation. Notice that the adjustment in each $[\mathcal{Z}_{(N^{(q)})_s}]$ provides the same qualitative information for each s, with straightforward modifications for the quantitative information given by the canonical resolutions. Notice that only the sample sizes given by $N^{(q)}$ will alter the qualitative information provided by the canonical directions, that is $W^{(d)}$ only depends upon $N^{(q)}$. Observe that there is a further property. The canonical directions do not depend upon either a or b, whilst the resolutions may be easily modified to handle changes to a or b. The terms a and b are the only terms in equation (3.100) that involve the factor settings for the cells which we hold fixed in the adjustment. Thus, if we consider the collection of cells,

$$f_{[k;q+1]}lf_{[q-1;1]} \quad = \quad (f_k, \ldots, f_{q+1}, l, f_{q-1}, \ldots, 1), \tag{3.109}$$

for $l = 1, \ldots, l_{q,0}$ for any choice of the levels $f_{[k;q+1]}$ and $f_{[q-1;1]}$ then the adjustment described above handles the adjustment of the canonical group directions corresponding to these cells; the only impact of changing $f_{[k;q+1]}$ and $f_{[q-1;1]}$ is to change a and b. There are $(l_0/l_{q,0})$ possible factor settings for the $f_{[k;q+1]}$ and $f_{[q-1;1]}$. Thus, the problem addressed only depends on what is happening to the qth factor settings and has the same interpretive consequences for any settings of the remaining $k-1$ factors. We now show that the solution of the underlying canonical qth factor problem for each $q = 1, \ldots, k$ provides the solution to the underlying canonical group problem.

Theorem 22 *If $A = \otimes_{q=1}^k A^{(q)}$ and $N^{-1}B = \otimes_{q=1}^k (N^{(q)})^{-1}B^{(q)}$ then the underlying canonical group directions are the columns of the matrix $W = [W_1 \ldots W_{l_0}]$ solving the generalised eigenvalue problem*

$$\{\otimes_{q=1}^k A^{(q)}\}W \quad = \quad [\{\otimes_{q=1}^k A^{(q)}\} + \{\otimes_{q=1}^k (N^{(q)})^{-1}B^{(q)}\}]W\Psi, \tag{3.110}$$

where $\Psi = diag(\psi_1, \ldots, \psi_{l_0})$ is the matrix of eigenvalues. Ψ may be written as

$$\Psi \quad = \quad P^T\{\otimes_{q=1}^k \Psi^{(q)}\}[\{\otimes_{q=1}^k \Psi^{(q)}\} + \{\otimes_{q=1}^k (I_{l_{q,0}} - \Psi^{(q)})\}]^{-1}P, \tag{3.111}$$

where P is a $l_0 \times l_0$ column permutation matrix chosen so that $1 > \psi_1 \geq \ldots \geq \psi_{l_0} > 0$ and $\Psi^{(q)}$ is the matrix of underlying canonical qth factor resolutions as defined in Definition 17. W may be expressed as

$$W = \{\otimes_{q=1}^{k} W^{(q)}\}P, \tag{3.112}$$

with $W^{(q)}$ the matrix of underlying canonical qth factor directions as given in Definition 17.

Proof - Equation (3.110) follows immediately from equation (2.123). P is a $l_0 \times l_0$ permutation matrix, so each column is of the form $\epsilon_{l_0 l}$ for some unique l. It thus follows immediately that $P^T P = I_{l_0}$. Thus, if (W, Ψ) solve equation (3.110) then so do $(WP, P^T \Psi P)$. We show that $W = \otimes_{q=1}^{k} W^{(q)}$ and $\Psi = \{\otimes_{q=1}^{k} \Psi^{(q)}\}[\{\otimes_{q=1}^{k} \Psi^{(q)}\} + \{\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})\}]^{-1}$ are solutions. Consider

$$\{\otimes_{q=1}^{k} A^{(q)}\}\{\otimes_{q=1}^{k} W^{(q)}\}\{\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})\} = \otimes_{q=1}^{k} A^{(q)} W^{(q)}(I_{l_{q,0}} - \Psi^{(q)}). \tag{3.113}$$

Notice from the solution of the underlying canonical qth factor problem as given by equation (3.71), we may write

$$A^{(q)} W^{(q)}(I_{l_{q,0}} - \Psi^{(q)}) = (N^{(q)})^{-1} B^{(q)} W^{(q)} \Psi^{(q)}. \tag{3.114}$$

Substituting equation (3.114) into equation (3.113) gives

$$\begin{aligned} \{\otimes_{q=1}^{k} A^{(q)}\}\{\otimes_{q=1}^{k} W^{(q)}\}\{\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})\} &= \otimes_{q=1}^{k}(N^{(q)})^{-1} B^{(q)} W^{(q)} \Psi^{(q)} \quad (3.115) \\ &= \{\otimes_{q=1}^{k}(N^{(q)})^{-1} B^{(q)}\}\{\otimes_{q=1}^{k} W^{(q)}\}\{\otimes_{q=1}^{k} \Psi^{(q)}\}. \quad (3.116) \end{aligned}$$

Adding $\{\otimes_{q=1}^{k} A^{(q)}\}\{\otimes_{q=1}^{k} W^{(q)}\}\{\otimes_{q=1}^{k} \Psi^{(q)}\}$ to both sides gives

$$\begin{aligned} &\{\otimes_{q=1}^{k} A^{(q)}\}\{\otimes_{q=1}^{k} W^{(q)}\}[\{\otimes_{q=1}^{k} \Psi^{(q)}\} + \{\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})\}] = \\ &\quad [\{\otimes_{q=1}^{k} A^{(q)}\} + \{\otimes_{q=1}^{k}(N^{(q)})^{-1} B^{(q)}\}]\{\otimes_{q=1}^{k} W^{(q)}\}\{\otimes_{q=1}^{k} \Psi^{(q)}\}, \end{aligned} \tag{3.117}$$

and the result follows. □

3.3.1 The underlying canonical variable and group problems for the examiner

As we stated, the only difference between the work in this chapter and that of Chapter 2 is the specification made in the underlying canonical group problem as defined by Definition 13. The underlying canonical variable problem is as defined by Definition 12. Thus, the examiner may make immediate use of the work of

Subsection 2.4.2. By Theorem 22, the solution of the underlying canonical group problem is obtained by the solution of the underlying canonical qth factor problem for each q. For the examiner, the underlying canonical 1st factor problem may be considered as the underlying canonical marker problem. Since $N^{(1)} = nI_2$, this problem is

$$A^{(1)}W^{(1)} = (A^{(1)} + nB^{(1)})W^{(1)}\Psi^{(1)}. \tag{3.118}$$

By comparing equation (3.17) to equation (2.117) and using the fact that $N^{(1)} = nI_2$, the marker may immediately observe that this problem is identical to the analysis he performed in Subsection 2.4.4, letting $W^{(1)}$ and $\Psi^{(1)}$ equate to the W and Ψ of equation (2.156). Thus, the examiner needs only to solve the underlying canonical paper problem. Since $N^{(2)} = I_4$, his problem is

$$A^{(2)}W^{(2)} = (A^{(2)} + B^{(2)})W^{(2)}\Psi^{(2)}, \tag{3.119}$$

where $A^{(2)}$ and $B^{(2)}$ are as defined by equation (3.19). Solving the problem yields

$$W^{(2)} = \begin{pmatrix} 0.2505 & 2.0529 & 0.6152 & 0.0970 \\ 0.3297 & -0.2167 & -2.5522 & -0.4804 \\ 0.2432 & -0.8584 & 0.5641 & 2.9189 \\ 0.3330 & -1.2055 & 1.4953 & -2.3654 \end{pmatrix}; \tag{3.120}$$

$$\Psi^{(2)} = \begin{pmatrix} 0.8161 & 0 & 0 & 0 \\ 0 & 0.1695 & 0 & 0 \\ 0 & 0 & 0.1471 & 0 \\ 0 & 0 & 0 & 0.0940 \end{pmatrix}. \tag{3.121}$$

The solution to the underlying canonical group problem is then found immediately using Theorem 22.

3.4 The adjustment of the mean components by the observed sample

By combining the reduction of the underlying group problem as given in Theorem 22 with the results of Theorem 13 we get the following corollary which shows how the canonical variable and qth factor analysis, for each $q = 1, \ldots, k$ completely determine the adjustment of the full collection $\mathcal{M}(\mathcal{C})$ by $\mathcal{C}(N)$.

Corollary 7 *The resolution transform matrix for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$ may be calculated as*

$$T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(N)]} = \left[\{\otimes_{p=1}^{k} A^{(p)}\} \otimes C + \{\otimes_{p=1}^{k} (N^{(p)})^{-1} B^{(p)}\} \otimes E\right]^{-1} \left[\{\otimes_{p=1}^{k} A^{(p)}\} \otimes C\right] \tag{3.122}$$

Letting $d_{[k]}$ be a shorthand for the cell $(d_k, \ldots, d_1)$, the collection $Z_{(N)} = \{Z_{(N)d_{[k]}s}\}$ for $d_q = 1, \ldots, l_{q,0}$, $q = 1, \ldots, k$ and $s = 1, \ldots, v_0$ is the set of canonical directions for the adjustment, where

$$Z_{(N)d_{[k]}s} = \left[\left\{\otimes_{q=1}^{k} W_{d_q}^{(q)}\right\} \otimes Y_s\right]^T \mathcal{M}(C). \tag{3.123}$$

The corresponding canonical resolutions are given by

$$\lambda_{(N)d_{[k]}s} = \frac{\psi_{d_{[k]}}\phi_s}{\psi_{d_{[k]}}\phi_s + (1-\psi)_{d_{[k]}}(1-\phi_s)}, \tag{3.124}$$

where

$$\psi_{d_{[k]}} = \prod_{q=1}^{k} \psi_{d_q}^{(q)}; \tag{3.125}$$

$$(1-\psi)_{d_{[k]}} = \prod_{q=1}^{k}(1-\psi_{d_q}^{(q)}). \tag{3.126}$$

The resolution ratio for the canonical direction $Z_{(N)d_{[k]}s}$ is given by

$$RR_{C(N)}(Z_{(N)d_{[k]}s}) = \frac{\lambda_{(N)d_{[k]}s}}{1-\lambda_{(N)d_{[k]}s}} = \left\{\prod_{q=1}^{k}\frac{\psi_{d_q}^{(q)}}{1-\psi_{d_q}^{(q)}}\right\}\frac{\phi_s}{1-\phi_s}. \tag{3.127}$$

Proof - From Theorem 22 we have that W, Ψ in Theorem 13 may, ignoring the ordering of the resolutions in order of magnitude, be expressed as

$$W = \otimes_{q=1}^{k} W^{(q)}; \tag{3.128}$$

$$\Psi = \{\otimes_{q=1}^{k}\Psi^{(q)}\}[\{\otimes_{q=1}^{k}\Psi^{(q)}\} + \{\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})\}]^{-1}, \tag{3.129}$$

so that equation (3.123) follows from equation (3.128). From equation (3.129) we have

$$(I_{l_0} - \Psi) = \{\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})\}[\{\otimes_{q=1}^{k}\Psi^{(q)}\} + \{\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})\}]^{-1}. \tag{3.130}$$

Substituting equations (3.129) and (3.130) into equation (2.170) gives

$$\begin{aligned}\Lambda_{(N)} = {} & \{(\otimes_{q=1}^{k}\Psi^{(q)}) \otimes \Phi\} \times \\ & [\{(\otimes_{q=1}^{k}\Psi^{(q)}) \otimes \Phi\} + \{(\otimes_{q=1}^{k}(I_{l_{q,0}} - \Psi^{(q)})) + (I_{v_0} - \Phi)\}]^{-1},\end{aligned} \tag{3.131}$$

from which equations (3.125) and (3.126) follow. □

Hence, we have the advantage that the $l_0 \times l_0$ underlying canonical group problem can be broken down into k underlying problems; those of the underlying canonical qth factor problem for each $q \in \Delta$. These problems are of size $l_{q,0} \times l_{q,0}$ and have an interpretable form, namely that of learning in the $\mathcal{Z}_{(N^{(q)})_s}$s, which provides added understanding for the types of information the design conveys and also the role of the factor settings in the design process. The elegant features of the balanced design remain in this multifactor design. We have the following corollaries which should be compared to Corollary 4.

Corollary 8 *If the sample sizes in the cells* $(f_{[k;q+1]}l_q f_{[q-1;1]})$ *for each* $l_q \in \mathcal{I}_q$ *are the same, that is we take* $N^{(q)} = n^{(q)} I_{l_{q,0}}$*, then the underlying canonical qth factor directions,* $W^{(q)} = [W_1^{(q)} \ldots W_{l_{q,0}}^{(q)}]$*, are the same for each* $n^{(q)}$ *and the corresponding canonical qth factor directions are given by*

$$\psi^{(q)}_{(n^{(q)})d} = \frac{n^{(q)}\psi^{(q)}_{(1)d}}{(n^{(q)}-1)\psi^{(q)}_{(1)d}+1}, \tag{3.132}$$

where $\Psi^{(q)}_{(1)} = diag(\psi^{(q)}_{(1)1}, \ldots, \psi^{(q)}_{(1)l_{q,0}})$ *solves* $A^{(q)}W^{(q)} = (A^{(q)}+B^{(q)})W^{(q)}\Psi^{(q)}_{(1)}$*. In this case we may write*

$$\lambda_{(N:N^{(q)}=n^{(q)}I_{l_{q,0}})d_{[k]}s} = \frac{n^{(q)}\lambda_{(N:N^{(q)}=I_{l_{q,0}})d_{[k]}s}}{(n^{(q)}-1)\lambda_{(N:N^{(q)}=I_{l_{q,0}})d_{[k]}s}+1}; \tag{3.133}$$

$$RR_{C(N:N^{(q)}=n^{(q)}I_{l_{q,0}})}(Z_{d_{[k]}s}) = n^{(q)}RR_{C(N:N^{(q)}=I_{l_{q,0}})}(Z_{d_{[k]}s}). \tag{3.134}$$

We may utilise the result of this corollary to develop the case of the complete balanced design.

Corollary 9 *If the sample size in each cell,* $l_{[k]} \in \Delta$*, is* n*, that is we have a balanced design, then the canonical directions* $Z_{(nI_{l_0})d_{[k]}s} = [\{\otimes_{q=1}^k W_{d_q}^{(q)}\} \otimes Y_s]^T \mathcal{M}(\mathcal{C})$ *are the same for all* n *and if, for each* $q \in \Delta$*,* $\Psi^{(q)}_{(1)} = diag(\psi^{(q)}_{(1)1}, \ldots, \psi^{(q)}_{(1)l_{q,0}})$ *solves* $A^{(q)}W^{(q)} = (A^{(q)}+B^{(q)})W^{(q)}\Psi^{(q)}_{(1)}$ *then*

$$\lambda_{(nI_{l_0})d_{[k]}s} = \frac{n\lambda_{(I_{l_0})d_{[k]}s}}{(n-1)\lambda_{(I_{l_0})d_{[k]}s}+1} \tag{3.135}$$

$$= \frac{n\left\{\prod_{q=1}^k \psi^{(q)}_{(1)d_q}\right\}\phi_s}{n\left\{\prod_{q=1}^k \psi^{(q)}_{(1)d_q}\right\}\phi_s + \left\{\prod_{q=1}^k (1-\psi^{(q)}_{(1)d_q})\right\}(1-\phi_s)}; \tag{3.136}$$

$$RR_{C(nI_{l_0})}(Z_{(nI_{l_0})d_{[k]}s}) = n\left\{\frac{\lambda_{(I_{l_0})d_{[k]}s}}{1-\lambda_{(I_{l_0})d_{[k]}s}}\right\} \tag{3.137}$$

$$= n\left\{\prod_{q=1}^k \frac{\psi^{(q)}_{(1)d_q}}{1-\psi^{(q)}_{(1)d_q}}\right\}\frac{\phi_s}{1-\phi_s}. \tag{3.138}$$

We may then use Corollary 9, in a similar way to Corollary 5 to simplify design problems for choosing sample sizes to achieve specific variance reductions.

Corollary 10 *In the balanced design, the sample size n in each cell required to achieve a proportionate variance reduction of κ for $Z_{(nI_{l_0})d_{[k]}s}$, is*

$$n \geq \left\{\frac{\kappa}{1-\kappa}\right\}\left\{\frac{1-\lambda_{(I_{l_0})d_{[k]}s}}{\lambda_{(I_{l_0})d_{[k]}s}}\right\} \tag{3.139}$$

$$= \left\{\frac{\kappa}{1-\kappa}\right\}\left\{\prod_{q=1}^{k}\frac{1-\psi^{(q)}_{(1)d_q}}{\psi^{(q)}_{(1)d_q}}\right\}\frac{1-\phi_s}{\phi_s}. \tag{3.140}$$

If the minimal canonical resolution for $n=1$ is

$$\lambda_{min} = \frac{\left\{\prod_{q=1}^{k}\psi^{(q)}_{min}\right\}\phi_{min}}{n\left\{\prod_{q=1}^{k}\psi^{(q)}_{min}\right\}\phi_{min}+\left\{\prod_{q=1}^{k}(1-\psi^{(q)}_{min})\right\}(1-\phi_{min})}, \tag{3.141}$$

then a sample size of

$$n_{min} = Int\left|\left\{\frac{\kappa}{1-\kappa}\right\}\left\{\prod_{q=1}^{k}\frac{1-\psi^{(q)}_{min}}{\psi^{(q)}_{min}}\right\}\frac{1-\phi_{min}}{\phi_{min}}\right|+1 \tag{3.142}$$

in each cell is the minimum sample which is sufficient to achieve a proportionate variance reduction of κ for every element of $[\mathcal{M}(\mathcal{C})]$.

$Int|x|$ denotes the integer part of x. It should be observed here that we can use this result, coupled with the costs of sampling, to find an upper bound on the cost of an experiment that seeks proportional variance reductions of κ. This bound is, of course, not just restricted to samples of the form given by equation (3.56) but for any sample where we sample $n_{l_{[k]}}$ individuals in the $l_{[k]}$th cell.

3.5 The full solution to the examiner's problem

By Corollary 7, the examiner knows his solution of the underlying canonical problems which is obtained in Subsection 3.3.1 is all he needs to find the canonical structure for the full adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_8)$. The co-ordinates of the canonical resolutions may be found by calculating $W^{(2)} \otimes W^{(1)} \otimes Y$. The examiner performs this calculation and the results are given in Table 3.1. For example,

Direction	$Z_{(N)111}$	$Z_{(N)112}$	$Z_{(N)121}$	$Z_{(N)122}$	$Z_{(N)211}$	$Z_{(N)212}$	$Z_{(N)221}$	$Z_{(N)222}$
Multiplier	τ_+	τ_+	τ_-	τ_-	τ_+	τ_+	τ_-	τ_-
Component								
$\mathcal{M}(X_{111})$	0.1446	0.1120	0.1446	0.1120	1.1852	0.9181	1.1852	0.9181
$\mathcal{M}(X_{112})$	-0.1446	0.1120	-0.1446	0.1120	-1.1852	0.9181	-1.1852	0.9181
$\mathcal{M}(X_{121})$	0.1446	0.1120	-0.1446	-0.1120	1.1852	0.9181	-1.1852	-0.9181
$\mathcal{M}(X_{122})$	-0.1446	0.1120	0.1446	-0.1120	-1.1852	0.9181	1.1852	-0.9181
$\mathcal{M}(X_{211})$	0.1904	0.1474	0.1904	0.1474	-0.1251	-0.0969	-0.1251	-0.0969
$\mathcal{M}(X_{212})$	-0.1904	0.1474	-0.1904	0.1474	0.1251	-0.0969	0.1251	-0.0969
$\mathcal{M}(X_{221})$	0.1904	0.1474	-0.1904	-0.1474	-0.1251	-0.0969	0.1251	0.0969
$\mathcal{M}(X_{222})$	-0.1904	0.1474	0.1904	-0.1474	0.1251	-0.0969	-0.1251	0.0969
$\mathcal{M}(X_{311})$	0.1404	0.1088	0.1404	0.1088	-0.4956	-0.3839	-0.4956	-0.3839
$\mathcal{M}(X_{312})$	-0.1404	0.1088	-0.1404	0.1088	0.4956	-0.3839	0.4956	-0.3839
$\mathcal{M}(X_{321})$	0.1404	0.1088	-0.1404	-0.1088	-0.4956	-0.3839	0.4956	0.3839
$\mathcal{M}(X_{322})$	-0.1404	0.1088	0.1404	-0.1088	0.4956	-0.3839	-0.4956	0.3839
$\mathcal{M}(X_{411})$	0.1923	0.1489	0.1923	0.1489	-0.6960	-0.5391	-0.6960	-0.5391
$\mathcal{M}(X_{412})$	-0.1923	0.1489	-0.1923	0.1489	0.6960	-0.5391	0.6960	-0.5391
$\mathcal{M}(X_{421})$	0.1923	0.1489	-0.1923	-0.1489	-0.6960	-0.5391	0.6960	0.5391
$\mathcal{M}(X_{422})$	-0.1923	0.1489	0.1923	-0.1489	0.6960	-0.5391	-0.6960	0.5391

Direction	$Z_{(N)111}$	$Z_{(N)112}$	$Z_{(N)121}$	$Z_{(N)122}$	$Z_{(N)211}$	$Z_{(N)212}$	$Z_{(N)221}$	$Z_{(N)222}$
Multiplier	τ_+	τ_+	τ_-	τ_-	τ_+	τ_+	τ_-	τ_-
Component								
$\mathcal{M}(X_{111})$	0.3552	0.2751	0.3552	0.2751	0.0560	0.0434	0.0560	0.0434
$\mathcal{M}(X_{112})$	-0.3552	0.2751	-0.3552	0.2751	-0.0560	0.0434	-0.0560	0.0434
$\mathcal{M}(X_{121})$	0.3552	0.2751	-0.3552	-0.2751	0.0560	0.0434	-0.0560	-0.0434
$\mathcal{M}(X_{122})$	-0.3552	0.2751	0.3552	-0.2751	-0.0560	0.0434	0.0560	-0.0434
$\mathcal{M}(X_{211})$	-1.4735	-1.1414	-1.4735	-1.1414	-0.2774	-0.2148	-0.2774	-0.2148
$\mathcal{M}(X_{212})$	1.4735	-1.1414	1.4735	-1.1414	0.2774	-0.2148	0.2774	-0.2148
$\mathcal{M}(X_{221})$	-1.4735	-1.1414	1.4735	1.1414	-0.2774	-0.2148	0.2774	0.2148
$\mathcal{M}(X_{222})$	1.4735	-1.1414	-1.4735	1.1414	0.2774	-0.2148	-0.2774	0.2148
$\mathcal{M}(X_{311})$	0.3257	0.2523	0.3257	0.2523	1.6852	1.3054	1.6852	1.3054
$\mathcal{M}(X_{312})$	-0.3257	0.2523	-0.3257	0.2523	-1.6852	1.3054	-1.6852	1.3054
$\mathcal{M}(X_{321})$	0.3257	0.2523	-0.3257	-0.2523	1.6852	1.3054	-1.6852	-1.3054
$\mathcal{M}(X_{322})$	-0.3257	0.2523	0.3257	-0.2523	-1.6852	1.3054	1.6852	-1.3054
$\mathcal{M}(X_{411})$	0.8633	0.6687	0.8633	0.6687	-1.3657	-1.0578	-1.3657	-1.0578
$\mathcal{M}(X_{412})$	-0.8633	0.6687	-0.8633	0.6687	1.3657	-1.0578	1.3657	-1.0578
$\mathcal{M}(X_{421})$	0.8633	0.6687	-0.8633	-0.6687	-1.3657	-1.0578	1.3657	1.0578
$\mathcal{M}(X_{422})$	-0.8633	0.6687	0.8633	-0.6687	1.3657	-1.0578	-1.3657	1.0578

Table 3.1: The canonical directions for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_8)$. The multipliers are $\tau_+ = (1+\gamma)^{-\frac{1}{2}}$ and $\tau_- = (1-\gamma)^{-\frac{1}{2}}$.

$\lambda_{(N)tgs}$	$g=1$, $s=1$	$g=1$, $s=2$	$g=2$, $s=1$	$g=2$, $s=2$
$t=1$	$\frac{0.4897(1+\gamma)n}{0.4897(1+\gamma)n+0.0736}$	$\frac{0.4897(1-\gamma)n}{0.4897(1-\gamma)n+0.0736}$	$\frac{0.1166(1+\gamma)n}{0.1166(1+\gamma)n+0.1576}$	$\frac{0.1166(1-\gamma)n}{0.1166(1-\gamma)n+0.1576}$
$t=2$	$\frac{0.1017(1+\gamma)n}{0.1017(1+\gamma)n+0.3322}$	$\frac{0.1017(1-\gamma)n}{0.1017(1-\gamma)n+0.3322}$	$\frac{0.0242(1+\gamma)n}{0.0242(1+\gamma)n+0.7119}$	$\frac{0.0242(1-\gamma)n}{0.0242(1-\gamma)n+0.7119}$
$t=3$	$\frac{0.0883(1+\gamma)n}{0.0883(1+\gamma)n+0.3412}$	$\frac{0.0883(1-\gamma)n}{0.0883(1-\gamma)n+0.3412}$	$\frac{0.0210(1+\gamma)n}{0.0210(1+\gamma)n+0.7311}$	$\frac{0.0210(1-\gamma)n}{0.0210(1-\gamma)n+0.7311}$
$t=4$	$\frac{0.0564(1+\gamma)n}{0.0564(1+\gamma)n+0.3624}$	$\frac{0.0564(1-\gamma)n}{0.0564(1-\gamma)n+0.3624}$	$\frac{0.0134(1+\gamma)n}{0.0134(1+\gamma)n+0.7766}$	$\frac{0.0134(1-\gamma)n}{0.0134(1-\gamma)n+0.7766}$

Table 3.2: The corresponding canonical resolutions to the canonical directions of Table 3.1 for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_8)$.

$$
\begin{aligned}
Z_{(N)111} &= (1+\gamma)^{-\frac{1}{2}}\{0.1446\mathcal{M}(X_{111}) - 0.1446\mathcal{M}(X_{112})+ \\
&\qquad \ldots + 0.1923\mathcal{M}(X_{421}) - 0.1923\mathcal{M}(X_{422})\} \quad (3.143)\\
&= 0.2505[\sqrt{1/2(1+\gamma)}\{\sqrt{(2/3)}\mathcal{M}(X_{111}) - \sqrt{(2/3)}\mathcal{M}(X_{112}) + \\
&= \ldots + .3330[\sqrt{1/2(1+\gamma)}\{\sqrt{(2/3)}\mathcal{M}(X_{421}) - \sqrt{(2/3)}\mathcal{M}(X_{422})\}] \quad (3.144)
\end{aligned}
$$

The corresponding canonical resolutions are the diagonal elements of the matrix $(\Psi^{(2)}\otimes\Psi^{(1)}\otimes\Phi)\ \{(\Psi^{(2)}\otimes\Psi^{(1)}\otimes\Phi)\ +((I_4-\Psi^{(2)})\otimes(I_2-\Psi^{(1)})\otimes(I_2-\Phi))\}^{-1}$. From his work in Subsection 3.3.1, the examiner may immediately calculate these. He displays them in Table 3.2. Thus, if $\gamma > 0$ then the largest resolution for all quantities in $[\mathcal{M}(\mathcal{C})]$ is for quantities proportional to $Z_{(N)111}$. This is approximately proportional to $\mathcal{M}(X_{..1}) - \mathcal{M}(X_{..2})$, the difference between the questions, averaged over papers and markers. Likewise, $Z_{(N)112}$ is approximately proportional to the overall mean, $\mathcal{M}(X_{...})$. $Z_{(N)122}$ is approximately proportional to $\mathcal{M}(X_{.1.}) - \mathcal{M}(X_{.2.})$ and $Z_{(N)121}$ is approximately proportional to $\mathcal{M}(X_{.11}) - \mathcal{M}(X_{.12}) - \mathcal{M}(X_{.21}) + \mathcal{M}(X_{.22})$.

For $\gamma > 0$ the smallest resolution for all quantities in $[\mathcal{M}(\mathcal{C})]$ is for quantities proportional to $Z_{(N)422}$. Notice that there are a number of small eigenvalues so that large samples will be required to achieve a variance reduction of κ for these quantities.

3.6 Marginal tables

The solution of the full k-dimensional problem in terms of the canonical structure enables us to investigate the learning of any quantity $\mathcal{X} \in [\mathcal{M}(\mathcal{C})]$. In many designs, we may only be interested in learning about the relationships between a few of the factors. That is, we may want to investigate marginal quantities. Analysis of the full space of course enables us to do this, but the results become clearer if we restrict our attention to learning in the subspace of interest.

We consider forming marginal spaces. Let $\tilde{\Delta} \subset \Delta$, then a $\tilde{\Delta}$-marginal space is obtained by only classifying the quantities of interest according to the criteria in $\tilde{\Delta}$. That is we only consider the classification variables in $\tilde{\Delta}$. For ease of notation, we shall assume that we are interested in the marginal when $\tilde{\Delta} = \{1, \ldots, q\}$. We shall conclude the section with the more general case of any $\tilde{\Delta} \subset \Delta$. The available cells are then the elements $l_{[q]} = (l_q, \ldots, l_1)$ of the product $\mathcal{I}_{[q;1]} = \times_{p=1,\ldots,q}\mathcal{I}_p$. For each cell $l_{[q]}$ we have the marginalised collection of random quantities $\mathcal{M}(\mathcal{C}_{l_{[q]}}) = \{\mathcal{M}(X_{l_{[q]}1}), \ldots, \mathcal{M}(X_{l_{[q]}v_0})\}$ that we are interested in learning about, where $\mathcal{M}(X_{l_{[q]}v})$ is defined as

$$\mathcal{M}(X_{l_{[q]}v}) = \sum_{p=q+1}^{k} \sum_{l_p=1}^{l_{p,0}} \mathcal{M}(X_{l_{[k]}v}), \tag{3.145}$$

so that the full mean collection is $\mathcal{M}(\mathcal{C}_{(q)}) = \cup_{l_{[q]}} \mathcal{M}(\mathcal{C}_{l_{[q]}})$, which in vector form is

$$\mathcal{M}(\mathcal{C}_{(q)}) = M\mathcal{M}(\mathcal{C}), \tag{3.146}$$

where

$$M = \left\{\otimes_{p=q+1}^{k} 1_{l_{p,0}}^{T}\right\} \otimes \left\{\otimes_{p=1}^{q} I_{l_{p,0}}\right\} \otimes I_{v_0}. \tag{3.147}$$

We interpret the mapping f, given in Definition 16, for the cells $l_{[k;q+1]}$ to be

$$l_{[k;q+1]} \mapsto \left\{\sum_{g=q+2}^{k} (l_g - 1) \prod_{g'=q+1}^{g-1} l_{g',0}\right\} + l_{q+1}. \tag{3.148}$$

The following lemma describes the relationships.

Lemma 19 *The relationships between the marginalised mean components and the sample means may be expressed as*

$$\begin{aligned}
Var(\mathcal{M}(\mathcal{C}_{(q)})) &= \left\{\otimes_{p=q+1}^{k} 1_{l_{p,0}}^{T} A^{(p)} 1_{l_{p,0}}\right\} \otimes C^{(q)}; &(3.149)\\
Cov(\mathcal{M}(\mathcal{C}_{(q)}), \mathcal{S}_N(\mathcal{C})) &= \left\{\otimes_{p=q+1}^{k} 1_{l_{p,0}}^{T} A^{(p)}\right\} \otimes C^{(q)}; &(3.150)\\
Var(\mathcal{S}_N(\mathcal{C})) &= \Big[\left\{\otimes_{p=q+1}^{k} (W^{(p)})^{-1}(A^{(p)} + (N^{(p)})^{-1}B^{(p)})^{-1}\right\} \\
&\quad \otimes I_{l_{[q;1],0}v_0}\Big]^{-1} \left[\oplus_{p'=1}^{l_{[k;q+1],0}} \left\{\psi_{p'}^{(k;q+1)} C^{(q)} + (1-\psi_{p'})^{(k;q+1)} E^{(q)}\right\}\right] \times \\
&\quad \left[\left\{\otimes_{p=q+1}^{k} (W^{(p)})^{-1}\right\} \otimes I_{l_{[q;1],0}v_0}\right]; &(3.151)\\
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}_{(q)})) &= \Big[\left\{\otimes_{p=q+1}^{k} (W^{(p)})^{-1}(A^{(p)} + (N^{(p)})^{-1}B^{(p)})^{-1}\right\} \\
&\quad \otimes I_{l_{[q;1],0}v_0}\Big]^{-1} \left\{\oplus_{p=1}^{l_{[k;q+1],0}} \psi_{p}^{(k;q+1)} C^{(q)}\right\} \left[\left\{\otimes_{p=q+1}^{k} (W^{(p)})^{-1} 1_{l_{p,0}}\right\} \otimes I_{l_{[q;1],0}v_0}\right] &(3.152)
\end{aligned}$$

where

$$\psi_{p'}^{(k;q+1)} = \prod_{p=q+1}^{k} \psi_{l_p}^{(p)} \text{ for } p' = f(l_{[k;q+1]}); \tag{3.153}$$

$$(1-\psi_{p'})^{(k;q+1)} = \prod_{p=q+1}^{k}(1-\psi_{l_p}^{(p)}) \text{ for } p' = f(l_{[k;q+1]}). \tag{3.154}$$

Proof - Equation (3.149) follows by noting that

$$Var(\mathcal{M}(\mathcal{C}_{(q)})) = Var(M\mathcal{M}(\mathcal{C})) = MVar(\mathcal{M}(\mathcal{C}))M^T, \tag{3.155}$$

and using the representation of $Var(\mathcal{M}(\mathcal{C}))$ given by equation (3.58).

Equation (3.150) follows by noting that

$$Cov(\mathcal{M}(\mathcal{C}_{(q)}), \mathcal{S}_N(\mathcal{C})) = Cov(M\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) = MCov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) \tag{3.156}$$

and using the representation of $Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C}))$ given by equation (3.59).

Equation (3.151) may be seen by taking the representation of $Var(\mathcal{S}_N(\mathcal{C}))$ given by equation (3.60) and premultiplying by $\left[\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}(A^{(p)}+(N^{(p)})^{-1}B^{(p)})^{-1}\}\right.$ $\left.\otimes I_{l_{[q;1],0}v_0}\right]$ and post multiplying by $\left[\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}\}\otimes I_{l_{[q;1],0}v_0}\right]$ and making use of the underlying canonical pth factor directions and resolutions, as given by Definition 17, for $p = q+1, \ldots, k$ and also the following two identities:

$$\oplus_{p=1}^{l_{[k;q+1],0}} \psi_p^{(k;q+1)} = \otimes_{p=q+1}^{k} \Psi^{(p)}; \tag{3.157}$$

$$\oplus_{p=1}^{l_{[k;q+1],0}} (1-\psi_p)^{(k;q+1)} = \otimes_{p=q+1}^{k} (I_{l_{p,0}} - \Psi^{(p)}). \tag{3.158}$$

Equation (3.152) follows in a similar fashion to equation (3.151) with a final post multiplication by M^T. □

Denote the resolution transform for the adjustment of the marginalised mean components by the full sample, $\mathcal{C}(N)$, by $T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{C}(N)]}$. We have the following theorem.

Theorem 23 $T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{C}(N)]}$ *may be computed as*

$$T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{C}(N)]} = \frac{1}{a_q}\sum_{p'=q+1}^{k}\sum_{l_{p'}=1}^{l_{p',0}}\left[\prod_{p=q+1}^{k}\left\{\sum_{g=1}^{l_{p,0}}(a_g^{(p)})^T W_{l_p}^{(p)}\right\}^2\right] \times$$

$$\left[\left(\prod_{p=q+1}^{k}\psi_{l_p}^{(p)}\right)C^{(q)} + \left(\prod_{p=q+1}^{k}(1-\psi_{l_p}^{(p)})\right)E^{(q)}\right]^{-1}\left(\prod_{p=q+1}^{k}\psi_{l_p}^{(p)}\right)C^{(q)} \tag{3.159}$$

where

$$a_q = \prod_{p=q+1}^{k} \left\{ 1_{l_{p,0}}^T A^{(p)} 1_{l_{p,0}} \right\}, \tag{3.160}$$

and $a_g^{(p)}$ is as given by equation (3.13). $W_{l_p}^{(p)}$ is the l_pth underlying canonical pth factor direction. $C^{(q)}$, $E^{(q)}$ are as given in equations (3.61) and (3.62). The canonical directions and resolutions of the adjustment of $[\mathcal{M}(\mathcal{C}_{(q)})]$ by $\mathcal{C}(N)$ may be deduced from those of the unmarginalised adjustment, namely that of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$. If $Z_{(N)d_{[k]}s}$, as given by equation (3.123), is a canonical direction of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$ with corresponding resolution $\lambda_{(N)d_{[k]}s}$, as given by equation (3.124), then $Z_{(N)d_{[q;1]}s}$ is a canonical direction of the adjustment of $[\mathcal{M}(\mathcal{C}_{(q)})]$ by $\mathcal{C}(N)$ with corresponding resolution $\lambda_{(N)d_{[q;1]}s}$, where

$$Z_{(N)d_{[q;1]}s} = \frac{1}{\sqrt{a_q}} \left[\left\{ \otimes_{p=1}^{q} W_{d_p}^{(p)} \right\} \otimes Y_s \right]^T \mathcal{M}(\mathcal{C}_{(q)}); \tag{3.161}$$

$$\lambda_{(N)d_{[q;1]}s} = \frac{1}{a_q} \sum_{p'=q+1}^{k} \sum_{l_{p'}=1}^{l_{p',0}} \left[\prod_{p=q+1}^{k} \left\{ \sum_{g=1}^{l_{p,0}} (a_g^{(p)})^T W_{l_p}^{(p)} \right\} \right]^2 \lambda_{(N)d_{[k]}s}. \tag{3.162}$$

This gives us the complete set of eigenvalues and eigenvectors.

Proof - From Theorem 8, $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for the adjustment of $[\mathcal{M}(\mathcal{C}_{(q)})]$. Hence, $T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{C}(N)]} = T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{S}_N(\mathcal{C})]}$, where $T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{S}_N(\mathcal{C})]}$ denotes the resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C}_{(q)})]$ by $\mathcal{S}_N(\mathcal{C})$. Defining

$$Var^{-1}Cov = \{Var(\mathcal{M}(\mathcal{C}_{(q)}))\}^{-1} Cov(\mathcal{M}(\mathcal{C}_{(q)}), \mathcal{S}_N(\mathcal{C})); \tag{3.163}$$

$$Cov^T = Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}_{(q)})), \tag{3.164}$$

then since all of the matrices are of full rank, $T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{S}_N(\mathcal{C})]}$ may be computed as

$$T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{S}_N(\mathcal{C})]} = Var^{-1}Cov\{Var(\mathcal{S}_N(\mathcal{C}))\}^{-1}Cov^T. \tag{3.165}$$

From equations (3.149) and (3.150) and making use of a as given by equation (3.160) we have

$$Var^{-1}Cov = \left[\frac{1}{a_q} \otimes (C^{(q)})^{-1} \right] \left[\left\{ \otimes_{p=q+1}^{k} 1_{l_{p,0}}^T A^{(p)} \right\} \otimes C^{(q)} \right] \tag{3.166}$$

$$= \frac{1}{a_q} \left[\left\{ \otimes_{p=q+1}^{k} 1_{l_{p,0}}^T A^{(p)} \right\} \otimes I_{l_{[q;1],0}v_0} \right] \tag{3.167}$$

$$= \frac{1}{a_q} \left[\left\{ \otimes_{p=q+1}^{k} 1_{l_{p,0}}^T A^{(p)} W^{(p)} \right\} \otimes I_{l_{[q;1],0}v_0} \right] \left[\left\{ \otimes_{p=q+1}^{k} (W^{(p)})^{-1} \right\} \otimes I_{l_{[q;1],0}v_0} \right]. \tag{3.168}$$

From the column representations of $A^{(p)}$, as given by equation (3.12), and $W^{(p)}$, as given in Definition 17, we find that

$$A^{(p)}W^{(p)} = \begin{pmatrix} (a_1^{(p)})^T W_1^{(p)} & \dots & (a_1^{(p)})^T W_{l_{p,0}}^{(p)} \\ \vdots & \ddots & \vdots \\ (a_{l_{p,0}}^{(p)})^T W_1^{(p)} & \dots & (a_{l_{p,0}}^{(p)})^T W_{l_{p,0}}^{(p)} \end{pmatrix}, \tag{3.169}$$

so that

$$1_{l_{p,0}}^T A^{(p)}W^{(p)} = \left[\sum_{g=1}^{l_{g,0}} (a_g^{(p)})^T W_1^{(p)} \dots \sum_{g=1}^{l_{g,0}} (a_g^{(p)})^T W_{l_{p,0}}^{(p)}\right]. \tag{3.170}$$

Letting

$$x_{p'} = \prod_{p=q+1}^{k} \left\{\sum_{g=1}^{l_{g,0}} (a_g^{(p)})^T W_{l_p}^{(p)}\right\} \text{ for } p' = f(l_{[k;q+1]}), \tag{3.171}$$

then substituting equation (3.170) into equation (3.168) and making use of equation (3.171) we find that

$$Var^{-1}Cov = \frac{1}{a_q}\left[x_1 I_{l_{[q;1],0}v_0} \dots x_{l_{[k;q+1],0}} I_{l_{[q;1],0}v_0}\right]\left[\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}\} \otimes I_{l_{[q;1],0}v_0}\right] \tag{3.172}$$

Now from our choice of canonical pth factor directions, we have that

$$(W^{(p)})^{-1} 1_{l_{p,0}} = \left\{1_{l_{p,0}}^T [(W^{(p)})^T]^{-1}\right\}^T \tag{3.173}$$

$$= \left\{1_{l_{p,0}}^T A^{(p)} W^{(p)}\right\}^T. \tag{3.174}$$

Making use of equation (3.174) in equation (3.152) we find that

$$Cov^T = [\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}(A^{(p)} + (N^{(p)})^{-1}B^{(p)})^{-1}\}$$

$$\otimes I_{l_{[q;1],0}v_0}]^{-1} \left\{\oplus_{p=1}^{l_{[k;q+1],0}} \psi_p^{(k;q+1)} C^{(q)}\right\} \left[x_1 I_{l_{[q;1],0}v_0} \dots x_{l_{[q;p+1],0}} I_{l_{[q;1],0}v_0}\right]^T. \tag{3.175}$$

Full rank specifications enable us to invert equation (3.149). Taking the inverse and premultiplying by equation (3.172) and post multiplying by equation (3.175) gives $T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{S}_N(\mathcal{C})]} = T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{C}(N)]}$ the representation of equation (3.159).

Following the proof to Corollary 7 it is straightforward to see that $(C^{(q)} + E^{(q)})^{-1}C^{(q)}$ has eigenvectors $\left\{\otimes_{p=1}^{q} W_{d_p}^{(p)}\right\} \otimes Y_s$ for $d_p = 1, \dots, l_{p,0}$ for each $p = 1, \dots, q$ and $s = 1, \dots, v_0$ where $W_{d_p}^{(p)}$ is the d_pth underlying canonical pth factor direction and Y_s the sth underlying canonical variable direction. The corresponding eigenvalues are given by

$$\lambda_{d_{[q]}s} = \frac{\left\{\prod_{p=1}^{q} \psi_{d_p}^{(p)}\right\} \phi_s}{\left\{\prod_{p=1}^{q} \psi_{d_p}^{(p)}\right\} \phi_s + \left\{\prod_{p=1}^{q}(1 - \psi_{d_p}^{(p)})\right\} (1 - \phi_s)}, \tag{3.176}$$

where $\psi^{(p)}_{d_p}$ is the d_pth underlying canonical pth factor resolution and ϕ_s is the sth underlying canonical variable resolution. The canonical structure follows immediately by noting that $(\xi C^{(q)}+\eta E^{(q)})^{-1}\xi C^q$ shares the same eigenvectors as $(C^{(q)}+E^{(q)})^{-1}C^q$ with eigenvalues given by

$$\lambda_{d_{[q]}s}(\xi,\eta) = \frac{\xi\lambda_{d_{[q]}s}}{\xi\lambda_{d_{[q]}s}+(1-\eta)(1-\lambda_{d_{[q]}s})}. \quad (3.177)$$

□

3.6.1 Marginalising over any $k-r$ factors

We now consider that we only want to classify the mean components in terms of r factors, for $r=0,\dots,k-1$. The case when $r=0$ being the marginalisation of the mean components to a single cell containing the v_0 variables, whilst if $r=k$ then we perform no marginalisation. Our classification criteria is then the set $\tilde{\Delta}\subset\Delta$, where $\tilde{\Delta}=\{\tilde{\delta}_1,\dots,\tilde{\delta}_r\}$. Without loss of generality we may take $\tilde{\delta}_1<\cdots<\tilde{\delta}_r$. The available cells are then the elements $l_{[\tilde{\Delta}]}=(l_{\tilde{\delta}_r},\dots,l_{\tilde{\delta}_1})$ of the product $\mathcal{I}_{\tilde{\Delta}}=\times_{\tilde{\delta}\in\tilde{\Delta}}\mathcal{I}_{\tilde{\delta}}$. In this case we have

$$l_{\tilde{\Delta}} = \prod_{p=1}^{r} l_{\tilde{\delta}_r,0} \quad (3.178)$$

possible cells. For each cell $l_{[\tilde{\Delta}]}$, we have the marginalised collection of mean components $\mathcal{M}(\mathcal{C}_{l_{[\tilde{\Delta}]}})=\{\mathcal{M}(X_{l_{[\tilde{\Delta}]}1}),\dots,\mathcal{M}(X_{l_{[\tilde{\Delta}]}v_0})\}$ where $\mathcal{M}(X_{l_{[\tilde{\Delta}]}v})$ is defined as

$$\mathcal{M}(X_{l_{[\tilde{\Delta}]}v}) = \sum_{d\notin\tilde{\Delta}}\sum_{l_\delta=1}^{l_{\delta,0}}\mathcal{M}(X_{l_{[k]}v}). \quad (3.179)$$

In vector form, the full collection of mean components is given by

$$\mathcal{M}(\mathcal{C}_{(\tilde{\Delta})}) = M_{\tilde{\Delta}}\mathcal{M}(\mathcal{C}), \quad (3.180)$$

where

$$M_{\tilde{\Delta}} = \left\{\otimes_{p=1}^{k}\left(I_{\tilde{\Delta}}(p)I_{l_{p,0}}+(1-I_{\tilde{\Delta}}(p))1^T_{l_{p,0}}\right)\right\}\otimes I_{v_0}, \quad (3.181)$$

where $I_{\tilde{\Delta}}(p)$ is the indicator function of $\tilde{\Delta}$, so that:

$$I_{\tilde{\Delta}}(p) = \begin{cases} 1 & \text{if } p\in\tilde{\Delta}; \\ 0 & \text{otherwise.}\end{cases} \quad (3.182)$$

Once more we are interested in using the full sample, $\mathcal{C}(N)$ to adjust $[\mathcal{M}(\mathcal{C}_{(\tilde{\Delta})})]$. We denote the resolution transform for this adjustment by $T_{[\mathcal{M}(\mathcal{C}_{(\tilde{\Delta})})/\mathcal{C}(N)]}$. The following corollary to Theorem 23 summarises the adjustment.

Corollary 11 *$T_{[\mathcal{M}(\mathcal{C}_{(\tilde{\Delta})})/\mathcal{C}(N)]}$ may be computed as*

$$T_{[\mathcal{M}(\mathcal{C}_{(\tilde{\Delta})})/\mathcal{C}(N)]} = \frac{1}{a_{\tilde{\Delta}}} \sum_{\delta' \notin \tilde{\Delta}} \sum_{l_{\delta'}=1}^{l_{\delta,0}} \left[\prod_{\delta \notin \tilde{\Delta}} \left\{ \sum_{g=1}^{l_{\delta,0}} (a_g^{(\delta)})^T W_{l_\delta}^{(\delta)} \right\}^2 \right] \times$$
$$\left[\left(\prod_{\delta \notin \tilde{\Delta}} \psi_{l_\delta}^{(\delta)} \right) C^{(\tilde{\Delta})} + \left(\prod_{\delta \notin \tilde{\Delta}} (1 - \psi_{l_\delta}^{(\delta)}) \right) E^{(\tilde{\Delta})} \right]^{-1} \left(\prod_{\delta \notin \tilde{\Delta}} \psi_{l_\delta}^{(\delta)} \right) C^{(\tilde{\Delta})}, \quad (3.183)$$

where

$$a_{\tilde{\Delta}} = \prod_{\delta \notin \tilde{\Delta}} \left\{ 1_{l_{\delta,0}}^T A^{(\delta)} 1_{l_{\delta,0}} \right\}, \quad (3.184)$$

and $a_g^{(\delta)}$ is as given by equation (3.13). $W_{l_\delta}^{(\delta)}$ is the l_δth underlying canonical δth factor direction. $C^{(\tilde{\Delta})}$, $E^{(\tilde{\Delta})}$ are as given in equations (3.64) and (3.65).

The canonical directions and resolutions of the adjustment of $[\mathcal{M}(\mathcal{C}_{(\tilde{\Delta})})]$ by $\mathcal{C}(N)$ may be deduced from those of the unmarginalised adjustment, namely that of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$. If $Z_{(N)d_{[k]}s}$, as given by equation (3.123), is a canonical direction of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$ with corresponding resolution $\lambda_{(N)d_{[k]}s}$, as given by equation (3.124), then $Z_{(N)d_{[\tilde{\Delta}]}s}$ is a canonical direction of the adjustment of $[\mathcal{M}(\mathcal{C}_{(\tilde{\Delta})})]$ by $\mathcal{C}(N)$ with corresponding resolution $\lambda_{(N)d_{[\tilde{\Delta}]}s}$, where

$$Z_{(N)d_{[\tilde{\Delta}]}s} = \frac{1}{\sqrt{a_{\tilde{\Delta}}}} \left[\left\{ \otimes_{p=1}^{r} W_{d_{\tilde{\delta}_p}}^{(\tilde{\delta}_p)} \right\} \otimes Y_s \right]^T \mathcal{M}(\mathcal{C}_{(\tilde{\Delta})}); \quad (3.185)$$

$$\lambda_{(N)d_{[\tilde{\Delta}]}s} = \frac{1}{a_{\tilde{\Delta}}} \sum_{\delta' \notin \tilde{\Delta}} \sum_{l_{\delta'}=1}^{l_{\delta',0}} \left[\prod_{\delta \notin \tilde{\Delta}} \left\{ \sum_{g=1}^{l_{\delta,0}} (a_g^{(\delta)})^T W_{l_\delta}^{(\delta)} \right\}^2 \right] \lambda_{(N)d_{[k]}s}. \quad (3.186)$$

This gives us the complete set of eigenvalues and eigenvectors.

Note then, that to find the $Z_{(N)d_{[\tilde{\Delta}]}s}$s and the $\lambda_{(N)d_{[\tilde{\Delta}]}s}$s, we need only solve our underlying canonical problems; no other calculations are necessary. Thus, solving the underlying problems enables us to have the solution to both the adjustment of the mean components and also for any marginalisation of the factors. Note that complete marginalisation of the factors leaves the adjustment providing the same qualitative information as for the single group adjustment, the structure being that of the underlying variable problem given in Definition 12. Of course, if we are only interested in the solution of the marginalisation problem, then we need only to solve the underlying canonical $\tilde{\delta}$th factor problems for each $\tilde{\delta} \in \tilde{\Delta}$ and the underlying canonical variable problem to find the $Z_{(N)d_{[\tilde{\Delta}]}s}$s, since these coupled with equation (3.183) yield the solution of the corresponding $\lambda_{(N)d_{[\tilde{\Delta}]}s}$s. Thus, we may choose to only solve the canonical factor directions relevant to the design. The downside may be a loss of understanding in how the sample sizes, $N^{(\delta)}$ for $\delta \in \Delta \setminus \tilde{\Delta}$ effect the information as we now explain.

Observe how the solution of this marginalised problem makes the choice of sample size and the role of the sample sizes clearer for selection in the marginalised space, with sample space of the form given by equation (3.56). The canonical resolutions, $Z_{(N)d_{[q;1]}s}$, only depend on the matrices $N^{(\tilde{\delta})}$ for $\tilde{\delta} \in \tilde{\Delta}$. If these are held fixed, then the qualitative features of the adjustment remain the same for all choices of the $N^{(\delta)}$s for $\delta \in \Delta \setminus \tilde{\Delta}$, the choices of the $N^{(\delta)}$s only effect the quantitative information. Thus, for example, we could use the $N^{(\tilde{\delta})}$s to fix the qualitative information and the $N^{(\delta)}$s to alter the quantitative nature of this information.

Notice that $\lambda_{(N)d_{[\tilde{\Delta}]}s}$ is the weighted sum of $\lambda_{(N)d_{[k]}s}$ over the marginalised levels. To see this, recall that

$$\sum_{g=1}^{l_{\delta,0}} (a_g^{(\delta)})^T W_{l_\delta}^{(\delta)} = \left(1_{l_{\delta,0}}^T A^{(\delta)} W^{(\delta)}\right)_{l_\delta}, \tag{3.187}$$

so that

$$\sum_{l_\delta=1}^{l_{\delta,0}} \left\{ \sum_{g=1}^{l_{\delta,0}} (a_g^{(\delta)})^T W_{l_\delta}^{(\delta)} \right\}^2 = 1_{l_{\delta,0}}^T A^{(\delta)} W^{(\delta)} (W^{(\delta)})^T A^{(\delta)} 1_{l_{\delta,0}}. \tag{3.188}$$

Now, from the choice of $W^{(q)}$ in the underlying canonical qth factor problem as given by Definition 17, namely that $(W^{(q)})^T A^{(q)} W^{(q)} = I_{l_{q,0}}$, we have that

$$\sum_{l_\delta=1}^{l_{\delta,0}} \left\{ \sum_{g=1}^{l_{\delta,0}} (a_g^{(\delta)})^T W_{l_\delta}^{(\delta)} \right\}^2 = 1_{l_{\delta,0}}^T A^{(\delta)} 1_{l_{\delta,0}}. \tag{3.189}$$

Hence,

$$\begin{aligned}
\sum_{\delta' \notin \tilde{\Delta}} \sum_{l_{\delta'}=1}^{l_{\delta',0}} \left[\prod_{\delta \notin \tilde{\Delta}} \left\{ \sum_{g=1}^{l_{\delta,0}} (a_g^{(\delta)})^T W_{l_\delta}^{(\delta)} \right\}^2 \right] &= \prod_{\delta \notin \tilde{\Delta}} \left[\sum_{l_\delta=1}^{l_{\delta,0}} \left\{ \sum_{g=1}^{l_{\delta,0}} (a_g^{(\delta)})^T W_{l_\delta}^{(\delta)} \right\}^2 \right] && (3.190)\\
&= \prod_{\delta \notin \tilde{\Delta}} \left\{ 1_{l_{\delta,0}}^T A^{(\delta)} 1_{l_{\delta,0}} \right\} && (3.191)\\
&= a_{[\tilde{\Delta}]}, && (3.192)
\end{aligned}$$

confirming that we do, indeed, have a weighted sum.

3.6.2 The examiner considers marginalisation

The examiner decides that he wants to restrict his attention to the overall performance of each marker on the specific questions. Thus, he decides to marginalise over the papers. He then reduces his eight cells to two, namely

$$\mathcal{M}(\mathcal{C}_g) = \sum_{t=1}^{4} \mathcal{M}(\mathcal{C}_{tg}), \tag{3.193}$$

for $g = 1, 2$. From Theorem 23, and its' accompanying corollary, the examiner knows he has little effort to employ in calculating the canonical resolutions and directions for the adjustment over the $\mathcal{M}(\mathcal{C}_g)$s. He calculates $A^{(2)}W^{(2)}$.

$$A^{(2)}W^{(2)} = \begin{pmatrix} 0.9226 & 0.3477 & 0.0881 & 0.0084 \\ 0.8923 & -0.0270 & -0.2686 & -0.0304 \\ 0.7986 & -0.1296 & 0.0720 & 0.2240 \\ 0.8421 & -0.1402 & 0.1470 & -0.1398 \end{pmatrix}, \tag{3.194}$$

whence

$$\sum_{g=1}^{4}(a_g^{(2)})^T W_1^{(2)} = 3.4556; \sum_{g=1}^{4}(a_g^{(2)})^T W_2^{(2)} = 0.0508; \tag{3.195}$$

$$\sum_{g=1}^{4}(a_g^{(2)})^T W_3^{(2)} = 0.0385; \sum_{g=1}^{4}(a_g^{(2)})^T W_4^{(2)} = 0.0622. \tag{3.196}$$

Calculating $a_g = 11.95$, then the canonical resolutions of his adjustment are

$$\begin{aligned} \lambda_{(nI_8)ds} &= \frac{1}{11.95}\{3.4556^2\lambda_{(nI_8)1ds} + 0.0508^2\lambda_{(nI_8)2ds} + \\ &\qquad 0.0385^2\lambda_{(nI_8)3ds} + 0.0622^2\lambda_{(nI_8)4ds}\} \qquad (3.197) \\ &\simeq \lambda_{(nI_8)1ds}. \qquad (3.198) \end{aligned}$$

This is of no surprise to the examiner, for if he calculates the corresponding canonical resolutions, he finds that

$$Z_{(nI_8)ds} = \frac{1}{\sqrt{11.95}}\{W_d^{(1)} \otimes Y_s\}^T \mathcal{M}(\mathcal{C}_g). \tag{3.199}$$

So that, for example, $Z_{(nI_8)11}$ is proportional to $\mathcal{M}(X_{\cdot\cdot 1}) - \mathcal{M}(X_{\cdot\cdot 2})$. This quantity was, of course, geometrically close to $Z_{(N)111}$. Similar observations follow to link $Z_{(nI_8)ds}$ with $Z_{(nI_8)1ds}$ for each of the choices of d and s. Equation (3.197) is the partition of the overall reduction in variance for the quantity $Z_{(nI_8)ds}$ in terms of the resolutions from each canonical direction in the full adjustment. For further details of this resolution partition see Section 7 of Goldstein & Wooff (1998).

3.7 Sliced tables

In addition to marginal quantities, we may also be interested in learning about quantities in cells corresponding to slices of the table. These are circumstances when we fix some of the classification criteria at predetermined levels and then allow the remaining criteria to vary.

Consider $\tilde{\Delta} \subset \Delta$, where $\tilde{\Delta} = \{\tilde{\delta}_1, \dots, \tilde{\delta}_r\}$ and suppose that, for each $p = 1, \dots, r$, we set the level of the δ_pth factor at $f_{\tilde{\delta}_p}$ and write $f_{[\tilde{\Delta}]} = (f_{\tilde{\delta}_r}, \dots, f_{\tilde{\delta}_1})$.

Definition 18 *The $f_{[\tilde{\Delta}]}$-slice of the table consists of all cells $l_{[k]} = (l_k, \dots, l_1)$ of the product $\mathcal{I}_{[\Delta\backslash\tilde{\Delta}]}$ of the level sets $\tilde{\mathcal{I}}_p$, that is $\mathcal{I}_{[\Delta\backslash\tilde{\Delta}]} = \times_{p=1,\dots,k}\tilde{\mathcal{I}}_p$ where*

$$\tilde{\mathcal{I}}_p = \begin{cases} f_p & \text{if } p \in \tilde{\Delta}; \\ \{1, \dots, l_{p,0}\} & \text{if } p \notin \tilde{\Delta}. \end{cases} \tag{3.200}$$

The $f_{[\tilde{\Delta}]}$-slice thus has

$$l_{\Delta\backslash\tilde{\Delta}} = \prod_{\delta\in\Delta\backslash\tilde{\Delta}} l_{\delta,0} \tag{3.201}$$

possible cells.

As in the case of marginalisation, initially we shall consider a convenient form of $\tilde{\Delta}$ in order to best illustrate the results, before concluding with the corresponding results for the general $l_{[\tilde{\Delta}]}$-slice.

Suppose that we take $\tilde{\Delta} = \{q+1, \dots, k\}$ and $f_{[\tilde{\Delta}]} = (f_k, \dots, f_{q+1})$. We regard this as the $f_{[k;q+1]}l_{[q]}$-slice and the available cells are

$$f_{[k;q+1]}l_{[q]} = (f_k, \dots, f_{q+1}, l_q, \dots, l_1) \tag{3.202}$$

for $l_p = 1, \dots, l_{p,0}$ for each $p = 1, \dots, q$. Notice that in the $f_{[k;q+1]}l_{[q]}$-slice we are allowing the same factors to vary as we did in the $[q;1]$-marginal, the difference here being that the remaining factors are held fixed at given levels as opposed to being summed over. In each cell $f_{[k;q+1]}l_{[q]}$ we are interested in learning about the mean component vector attached to that cell, namely $\mathcal{M}(\mathcal{C}_{f_{[k;q+1]}l_{[q]}})$. Notice the distinction here with the marginal case when the quantity of interest attached to the corresponding cell was $\mathcal{M}(\mathcal{C}_{l_{[q]}})$. The full collection of mean components is $\mathcal{M}(\mathcal{C}_{[k;q+1](q)}) = \cup_{l_{[q]}}\mathcal{M}(\mathcal{C}_{f_{[k;q+1]}l_{[q]}})$, which in vector form is

$$\mathcal{M}(\mathcal{C}_{[k;q+1](q)}) = S\mathcal{M}(\mathcal{C}), \tag{3.203}$$

where

$$S = \left\{\otimes_{p=q+1}^{k} e_{l_{p,0},f_p}^T\right\} \otimes \left\{\otimes_{p=1}^{q} I_{l_{p,0}}\right\} \otimes I_{v_0}. \tag{3.204}$$

As we explained above, our motivation is to make explicit the connection between learning in the slice and the analysis of the full table. In the $l_{[k;q+1]}l_{[q]}$-slice, we are focusing attention on learning about the mean components in the subspace $[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})]$ of $[\mathcal{M}(\mathcal{C})]$. To make the link explicit, we consider using the $\mathcal{C}(N)$ to learn about the $\mathcal{M}(\mathcal{C}_{[k;q+1](q)})$. By the Bayes linear sufficiency of $\mathcal{S}_N(\mathcal{C})$ for $\mathcal{C}(N)$ for the adjustment of $[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})]$, we know that we need only consider the second-order relationships between $[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})$ and $\mathcal{S}_N(\mathcal{C})$. The relationships for these collections, induced by the second-order specifications given in equations (3.7) - (3.9) and using the decompositions given by equations (3.10) and (3.11), may be written as in the following lemma.

Lemma 20 *The relationships between the mean components corresponding to the $l_{[k;q+1]}l_{[q]}$-slice and the sample means may be expressed as*

$$
\begin{aligned}
Var(\mathcal{M}(\mathcal{C}_{[k;q+1](q)})) &= \left\{\prod_{p=q+1}^{k} \alpha_{f_pf_p}^{(p)}\right\} \otimes C^{(q)}; & (3.205)\\
Cov(\mathcal{M}(\mathcal{C}_{[k;q+1](q)}), \mathcal{S}_N(\mathcal{C})) &= \left\{\otimes_{p=q+1}^{k} e_{l_{p,0}f_p}^{T} A^{(p)}\right\} \otimes C^{(q)}; & (3.206)\\
Var(\mathcal{S}_N(\mathcal{C})) &= \left[\left\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}(A^{(p)} + (N^{(p)})^{-1}B^{(p)})^{-1}\right\}\right.\\
&\quad \left.\otimes I_{l_{[q;1],0}v_0}\right]^{-1} \left[\oplus_{p'=1}^{l_{[k;q+1],0}} \left\{\psi_{p'}^{(k;q+1)} C^{(q)} + (1-\psi_{p'})^{(k;q+1)} E^{(q)}\right\}\right] \times\\
&\quad \left[\left\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}\right\} \otimes I_{l_{[q;1],0}v_0}\right]; & (3.207)\\
Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}_{[k;q+1](q)})) &= \left[\left\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}(A^{(p)} + (N^{(p)})^{-1}B^{(p)})^{-1}\right\}\right.\\
&\quad \left.\otimes I_{l_{[q;1],0}v_0}\right]^{-1} \left\{\oplus_{p=1}^{l_{[k;q+1],0}} \psi_{p}^{(k;q+1)} C^{(q)}\right\} \times\\
&\quad \left[\left\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1} e_{l_{p,0}f_p}\right\} \otimes I_{l_{[q;1],0}v_0}\right], & (3.208)
\end{aligned}
$$

where

$$
\begin{aligned}
\psi_{p'}^{(k;q+1)} &= \prod_{p=q+1}^{k} \psi_{l_p}^{(p)} \textit{ for } p' = f(l_{[k;q+1]}); & (3.209)\\
(1-\psi_{p'})^{(k;q+1)} &= \prod_{p=q+1}^{k} (1-\psi_{l_p}^{(p)}) \textit{ for } p' = f(l_{[k;q+1]}), & (3.210)
\end{aligned}
$$

and $C^{(q)} = \left\{\otimes_{p=1}^{q} A^{(p)}\right\} \otimes C$ *and* $E^{(q)} = \left\{\otimes_{p=1}^{q} (N^{(p)})^{-1} B^{(p)}\right\} \otimes E$.

Proof - Equation (3.205) follows by noting that

$$
Var(\mathcal{M}(\mathcal{C}_{[k;q+1](q)})) = Var(S\mathcal{M}(\mathcal{C})) = SVar(\mathcal{M}(\mathcal{C}))S^T, \quad (3.211)
$$

and using the representation of $Var(\mathcal{M}(\mathcal{C}))$ given by equation (3.58).

Equation (3.206) follows by noting that

$$
\begin{aligned}
Cov(\mathcal{M}(\mathcal{C}_{[k;q+1](q)}), \mathcal{S}_N(\mathcal{C})) &= Cov(S\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) & (3.212)\\
&= SCov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})), & (3.213)
\end{aligned}
$$

and using the representation of $Cov(\mathcal{M}(\mathcal{C}), \mathcal{S}_N(\mathcal{C}))$ given by equation (3.59).

Equation (3.207) is identical to Equation (3.151).

To obtain Equation (3.208) first note that

$$\begin{aligned} Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}_{[k;q+1](q)})) &= Cov(\mathcal{S}_N(\mathcal{C}), S\mathcal{M}(\mathcal{C})) && (3.214)\\ &= Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}))S^T. && (3.215)\end{aligned}$$

Taking the representation of $Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}))$ given by the transpose of equation (3.59) and premultiplying by $\left[\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}(A^{(p)} + (N^{(p)})^{-1}B^{(p)})^{-1}\} \otimes I_{l_{[q;1],0}v_0}\right]$ and post multiplying by $\left[\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}\} \otimes I_{l_{[q;1],0}v_0}\right]$ and making use of the underlying canonical pth factor directions and resolutions, as given by Definition 17, for $p = q+1, \ldots, k$ and also the following two identities:

$$\begin{aligned} \oplus_{p=1}^{l_{[k;q+1],0}}\psi_p^{(k;q+1)} &= \otimes_{p=q+1}^{k}\Psi^{(p)}; && (3.216)\\ \oplus_{p=1}^{l_{[k;q+1],0}}(1-\psi_p)^{(k;q+1)} &= \otimes_{p=q+1}^{k}(I_{l_{p,0}} - \Psi^{(p)}), && (3.217)\end{aligned}$$

it is easy to verify that we may write

$$\begin{aligned} Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C})) &= \left[\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}(A^{(p)} + (N^{(p)})^{-1}B^{(p)})^{-1}\}\right.\\ &\left.\otimes I_{l_{[q;1],0}v_0}\right]^{-1}\left\{\oplus_{p=1}^{l_{[k;q+1],0}}\psi_p^{(k;q+1)}C^{(q)}\right\}\left[\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}\} \otimes I_{l_{[q;1],0}v_0}\right]. \quad (3.218)\end{aligned}$$

Equation (3.208) then follows by post multiplying $Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}))$ by S^T. □

Denote the resolution transform for the adjustment of the mean components corresponding to the $f_{[k;q+1]}l_{[q]}$-slice by the full sample, $\mathcal{C}(N)$ as $T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{C}(N)]}$. We have the following theorem.

Theorem 24 *$T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{C}(N)]}$ may be computed as*

$$\begin{aligned} T_{[\mathcal{M}(\mathcal{C}_{(q)})/\mathcal{C}(N)]} &= \frac{1}{b_q}\sum_{p'=q+1}^{k}\sum_{l_{p'}=1}^{l_{p',0}}\left[\prod_{p=q+1}^{k}\left\{(a_{f_p}^{(p)})^T W_{l_p}^{(p)}\right\}^2\right] \times \\ &\left[\left(\prod_{p=q+1}^{k}\psi_{l_p}^{(p)}\right)C^{(q)} + \left(\prod_{p=q+1}^{k}(1-\psi_{l_p}^{(p)})\right)E^{(q)}\right]^{-1}\left(\prod_{p=q+1}^{k}\psi_{l_p}^{(p)}\right)C^{(q)}, \quad (3.219)\end{aligned}$$

where

$$b_q = \prod_{p=q+1}^{k}\alpha_{f_pf_p}^{(p)}, \quad (3.220)$$

and $a_{f_p}^{(p)}$ is as given by equation (3.13). $W_{l_p}^{(p)}$ is the l_pth underlying canonical pth factor direction. $C^{(q)}$, $E^{(q)}$ are as given in equations (3.61) and (3.62).

The canonical directions and resolutions of the adjustment of $[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})]$ *by* $\mathcal{C}(N)$ *may be deduced from those of the adjustment of* $[\mathcal{M}(\mathcal{C})]$ *by* $\mathcal{C}(N)$*. If* $Z_{(N)d_{[k]}s}$*, as given by equation (3.123), is a canonical direction of the adjustment of* $[\mathcal{M}(\mathcal{C})]$ *by* $\mathcal{C}(N)$ *with corresponding resolution* $\lambda_{(N)d_{[k]}s}$*, as given by equation (3.124), then* $Z_{f_{[k;q+1]}d_{[q;1]}s}$ *is a canonical direction of the adjustment of* $[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})]$ *by* $\mathcal{C}(N)$ *with corresponding resolution* $\lambda_{f_{[k;q+1]}d_{[q;1]}s}$*, where*

$$Z_{f_{[k;q+1]}d_{[q;1]}s} = \frac{1}{\sqrt{b_q}}\left[\left\{\otimes_{p=1}^{q} W_{d_p}^{(p)}\right\} \otimes Y_s\right]^T \mathcal{M}(\mathcal{C}_{[k;q+1](q)}); \tag{3.221}$$

$$\lambda_{f_{[k;q+1]}d_{[q;1]}s} = \frac{1}{b_q}\sum_{p'=q+1}^{k}\sum_{l_{p'}=1}^{l_{p',0}}\left[\prod_{p=q+1}^{k}\left\{(a_{f_p}^{(p)})^T W_{l_p}^{(p)}\right\}^2\right]\lambda_{(N)d_{[k]}s}. \tag{3.222}$$

This gives us the complete set of eigenvalues and eigenvectors.

Proof - From Theorem 8, $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for the adjustment of $[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})]$. Letting $T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{S}_N(\mathcal{C})]}$ denote the resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})]$ by $\mathcal{S}_N(\mathcal{C})$, the Bayes linear sufficiency gives

$$T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{C}(N)]} = T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{S}_N(\mathcal{C})]}. \tag{3.223}$$

Defining

$$Var^{-1}Cov = \{Var(\mathcal{M}(\mathcal{C}_{[k;q+1](q)}))\}^{-1}Cov(\mathcal{M}(\mathcal{C}_{[k;q+1](q)}), \mathcal{S}_N(\mathcal{C})); \tag{3.224}$$

$$Cov^T = Cov(\mathcal{S}_N(\mathcal{C}), \mathcal{M}(\mathcal{C}_{[k;q+1](q)})), \tag{3.225}$$

then since all of the matrices are of full rank, $T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{S}_N(\mathcal{C})]}$ may be computed as

$$T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{S}_N(\mathcal{C})]} = Var^{-1}Cov\{Var(\mathcal{S}_N(\mathcal{C}))\}^{-1}Cov^T. \tag{3.226}$$

From equations (3.205) and (3.206) and making use of b as given by equation (3.220) we have

$$Var^{-1}Cov = \left[\frac{1}{b_q}\otimes(C^{(q)})^{-1}\right]\left[\left\{\otimes_{p=q+1}^{k} e_{l_{p,0}f_p}^T A^{(p)}\right\}\otimes C^{(q)}\right] \tag{3.227}$$

$$= \frac{1}{b_q}\left[\left\{\otimes_{p=q+1}^{k} e_{l_{p,0}f_p}^T A^{(p)}\right\}\otimes I_{l_{[q;1],0}v_0}\right] \tag{3.228}$$

$$= \frac{1}{b_q}\left[\left\{\otimes_{p=q+1}^{k} e_{l_{p,0}f_p}^T A^{(p)}W^{(p)}\right\}\otimes I_{l_{[q;1],0}v_0}\right] \times \left[\left\{\otimes_{p=q+1}^{k}(W^{(p)})^{-1}\right\}\otimes I_{l_{[q;1],0}v_0}\right] \tag{3.229}$$

Using the representation of $A^{(p)}W^{(p)}$ given by equation (3.169), we have

$$e_{l_{p,0}f_p}^T A^{(p)}W^{(p)} = \left[(a_{f_p}^{(p)})^T W_1^{(p)} \ldots (a_{f_p}^{(p)})^T W_{l_{p,0}}^{(p)}\right]. \tag{3.230}$$

Letting

$$y_{p'} = \prod_{p=q+1}^{k} \left\{ (a_{f_p}^{(p)})^T W_{l_p}^{(p)} \right\} \text{ for } p' = f(l_{[k;q+1]}), \tag{3.231}$$

then substituting equation (3.230) into equation (3.229) and making use of equation (3.231) we find that

$$Var^{-1}Cov = \frac{1}{b_q}\left[y_1 I_{l_{[q;1],0}v_0} \cdots y_{l_{[k;q+1],0}} I_{l_{[q;1],0}v_0} \right]\left[\{ \otimes_{p=q+1}^{k} (W^{(p)})^{-1} \} \otimes I_{l_{[q;1],0}v_0} \right] \tag{3.232}$$

Now from our choice of canonical pth factor directions, we have that

$$\begin{aligned} (W^{(p)})^{-1} e_{l_{p,0}f_p} &= \left\{ e_{l_{p,0}f_p}^T [(W^{(p)})^T]^{-1} \right\}^T & (3.233)\\ &= \left\{ e_{l_{p,0}f_p}^T A^{(p)} W^{(p)} \right\}^T . & (3.234) \end{aligned}$$

By substituting equation (3.230) into (3.234) and making use of equation (3.231) we find that

$$\{ \otimes_{p=q+1}^{k} (W^{(p)})^{-1} e_{l_{p,0}f_p} \} = \left[y_1 \ldots y_{l_{[k;q+1],0}} \right]^T . \tag{3.235}$$

Substituting equation (3.235) into equation (3.208) we find that

$$\begin{aligned} Cov^T &= [\{ \otimes_{p=q+1}^{k} (W^{(p)})^{-1} (A^{(p)} + (N^{(p)})^{-1} B^{(p)})^{-1} \} \\ \otimes I_{l_{[q;1],0}v_0}]^{-1} & \left\{ \oplus_{p'=1}^{l_{[k;q+1],0}} \psi_{p'}^{(k;q+1)} C^{(q)} \right\} \left[y_1 I_{l_{[q;1],0}v_0} \cdots y_{l_{[q;p+1],0}} I_{l_{[q;1],0}v_0} \right]^T . \end{aligned} \tag{3.236}$$

Full rank specifications enable us to invert equation (3.205). Taking the inverse and premultiplying by equation (3.232) and post multiplying by equation (3.236) and simplifying gives $T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{S}_N(\mathcal{C})]} = T_{[\mathcal{M}(\mathcal{C}_{[k;q+1](q)})/\mathcal{C}(N)]}$ the representation of equation (3.219). The derivation of the eigenstructure follows exactly as for that in Theorem 23. □

3.7.1 Taking a general $f_{[\tilde{\Delta}]}$-slice

We now return to the case where we consider $\tilde{\Delta} = \{\tilde{\delta}_1, \ldots, \tilde{\delta}_r\}$, where we assume, without loss of generality, that $\tilde{\delta}_1 < \cdots < \tilde{\delta}_r$. We then have $\Delta \setminus \tilde{\Delta}$ $\{\delta_1, \ldots, \delta_{k-r}\}$, where $\delta_1 < \cdots < \delta_{k-r}$. In each cell $f_{[\tilde{\Delta}]} l_{[\Delta \setminus \tilde{\Delta}]}$ we are interested in learning about the mean component vector attached to that cell, namely $\mathcal{M}(\mathcal{C}_{f_{[\tilde{\Delta}]} l_{[\Delta\setminus\tilde{\Delta}]}})$. The full collection of mean components is $\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\setminus\tilde{\Delta})}) = \cup_{l_{[\Delta\setminus\tilde{\Delta}]} \in \mathcal{I}_{[\Delta\setminus\tilde{\Delta}]}} \mathcal{M}(\mathcal{C}_{f_{[\tilde{\Delta}]} l_{[\Delta\setminus\tilde{\Delta}]}})$, which in vector form is

$$\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\setminus\tilde{\Delta})}) = S_{\tilde{\Delta}} \mathcal{M}(\mathcal{C}), \tag{3.237}$$

where

$$S_{\tilde{\Delta}} = \left\{\otimes_{p=1}^{k}\left(I_{\{\tilde{\Delta}\}}(p)e^{T}_{l_{p,0}f_p} + (1 - I_{\{\tilde{\Delta}\}}(p))I_{l_{p,0}}\right)\right\} \otimes I_{v_0}, \tag{3.238}$$

where $I_{\{\tilde{\Delta}\}}(p)$ is the indicator function of $\tilde{\Delta}$, so that

$$I_{\{\tilde{\Delta}\}}(p) = \begin{cases} 1 & \text{if } p \in \tilde{\Delta}; \\ 0 & \text{otherwise.} \end{cases} \tag{3.239}$$

The following corollary to Theorem 24 gives the computational form for resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\backslash\tilde{\Delta})})]$ by the collection of sample means $\mathcal{C}(N)$, denoted by $T_{[\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\backslash\tilde{\Delta})})/\mathcal{C}(N)]}$, and also the canonical directions and corresponding resolutions of the adjustment.

Corollary 12 $T_{[\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\backslash\tilde{\Delta})})/\mathcal{C}(N)]}$ *may be computed as*

$$T_{[\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\backslash\tilde{\Delta})})/\mathcal{C}(N)]} = \frac{1}{b_{\tilde{\Delta}}}\sum_{\tilde{\delta}'\in\tilde{\Delta}}\sum_{l_{\tilde{\delta}'}=1}^{l_{\tilde{\delta}',0}}\left[\prod_{\tilde{\delta}\in\tilde{\Delta}}\left\{(a^{(\tilde{\delta})}_{f_{\tilde{\delta}}})^{T}W^{(\tilde{\delta})}_{l_{\tilde{\delta}}}\right\}^{2}\right] \times$$

$$\left[\left(\prod_{\tilde{\delta}\in\tilde{\Delta}}\psi^{(\tilde{\delta})}_{l_{\tilde{\delta}}}\right)C^{(\Delta\backslash\tilde{\Delta})} + \left(\prod_{\tilde{\delta}\in\tilde{\Delta}}(1-\psi^{(\tilde{\delta})}_{l_{\tilde{\delta}}})\right)E^{(\Delta\backslash\tilde{\Delta})}\right]^{-1}\left(\prod_{\tilde{\delta}\in\tilde{\Delta}}\psi^{(\tilde{\delta})}_{l_{\tilde{\delta}}}\right)C^{(\Delta\backslash\tilde{\Delta})}, \tag{3.240}$$

where

$$b_{\tilde{\Delta}} = \prod_{\tilde{\delta}\in\tilde{\Delta}}\alpha^{(\tilde{\delta})}_{f_{\tilde{\delta}}f_{\tilde{\delta}}}, \tag{3.241}$$

and $a^{(\tilde{\delta})}_{f_{\tilde{\delta}}}$ *is as given by equation (3.13).* $W^{(\tilde{\delta})}_{l_{\tilde{\delta}}}$ *is the* $l_{\tilde{\delta}}$*th underlying canonical* $\tilde{\delta}$*th factor direction and*

$$C^{(\Delta\backslash\tilde{\Delta})} = \left\{\otimes_{p=1}^{k-r}A^{(\delta_p)}\right\} \otimes C; \tag{3.242}$$

$$E^{(\Delta\backslash\tilde{\Delta})} = \left\{\otimes_{p=1}^{k-r}(N^{(\delta_p)})^{-1}B^{(\delta_p)}\right\} \otimes E. \tag{3.243}$$

The canonical directions and resolutions of the adjustment of $[\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\backslash\tilde{\Delta})})]$ *by* $\mathcal{C}(N)$ *may be deduced from those of the adjustment of* $[\mathcal{M}(\mathcal{C})]$ *by* $\mathcal{C}(N)$*. If* $Z_{(N)d_{[k]}s}$*, as given by equation (3.123), is a canonical direction of the adjustment of* $[\mathcal{M}(\mathcal{C})]$ *by* $\mathcal{C}(N)$ *with corresponding resolution* $\lambda_{(N)d_{[k]}s}$*, as given by equation (3.124), then* $Z_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\backslash\tilde{\Delta}]}s}$ *is a canonical direction of the adjustment of* $[\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\backslash\tilde{\Delta})})]$ *by* $\mathcal{C}(N)$ *with corresponding resolution* $\lambda_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\backslash\tilde{\Delta}]}s}$*, where*

$$Z_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\backslash\tilde{\Delta}]}s} = \frac{1}{\sqrt{b_{\tilde{\Delta}}}}\left[\left\{\otimes_{p=1}^{k-r}W^{(\delta_p)}_{d_{\delta_p}}\right\} \otimes Y_s\right]^{T}\mathcal{M}(\mathcal{C}_{[\tilde{\Delta}](\Delta\backslash\tilde{\Delta})}); \tag{3.244}$$

$$\lambda_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\backslash\tilde{\Delta}]}s} = \frac{1}{b_{\tilde{\Delta}}}\sum_{\tilde{\delta}'\in\tilde{\Delta}}\sum_{l_{\tilde{\delta}'}=1}^{l_{\tilde{\delta}',0}}\left[\prod_{\tilde{\delta}\in\tilde{\Delta}}\left\{(a^{(\tilde{\delta})}_{f_{\tilde{\delta}}})^{T}W^{(\tilde{\delta})}_{l_{\tilde{\delta}}}\right\}^{2}\right]\lambda_{(N)d_{[k]}s}. \tag{3.245}$$

This gives us the complete set of eigenvalues and eigenvectors.

Observe that in the same way as the marginalisation problem, we need only to solve our underlying problems to find the $Z_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\setminus\tilde{\Delta}]}s}$s and the $\lambda_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\setminus\tilde{\Delta}]}s}$s. Thus, the solution of the underlying problems enable us to solve not only the adjustment of the mean components, but also all marginalisation and slicing problems related to the mean components. If we are only interested in the $f_{[\tilde{\Delta}]}$-slice adjustment, then we only need to solve the underlying canonical δth factor problems for each $\delta \in \Delta \setminus \tilde{\Delta}$ and the underlying canonical variable problem to yield the canonical structure for the adjustment.

Notice how the canonical resolutions, $Z_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\setminus\tilde{\Delta}]}s}$, only depend on the matrices $N^{(\delta)}$ for $\delta \in \Delta \setminus \tilde{\Delta}$. If these are held fixed, then the qualitative features of the adjustment remain the same for all choices of the $N^{(\tilde{\delta})}$s, for $\tilde{\delta} \in \tilde{\Delta}$. Hence, the $N^{(\tilde{\delta})}$s only effect the quantitative information. Notice further that the choice of fixed levels, $f_{\tilde{\delta}}$ for each $\tilde{\delta} \in \tilde{\Delta}$ also only effects the quantitative information of the design, that is the qualitative information of the adjustment of the $f_{[\tilde{\Delta}]}$-slice only depends upon $\tilde{\Delta}$. This has a great interpretative advantage for assessing the effect of the individual levels on the design, for if $\tilde{\delta} \in \tilde{\Delta}$ then the effect of changing $f_{\tilde{\delta}}$ in the $f_{[\tilde{\Delta}]}$-slice is restricted completely to the quantitative aspects of the adjustment.

It should be noted that the collection $\mathcal{Z}_{(N^{(q)})s}$ as given by equation (3.73) consists of the sth canonical variable directions for the adjustment of the cell mean components by a sample in that cell, for each of the cells in the $f_{[\tilde{\Delta}]}$-slice, where $\tilde{\Delta} = \Delta \setminus \{q\}$. Thus, the underlying canonical qth factor problem may be related to learning in the $f_{[\Delta\setminus\{q\}]}$-slice. Notice that by taking $k = 1$ we may mirror the interpretation of the underlying canonical groups problem as given by Definition 13.

In a similar way to marginalisation, $\lambda_{(N)f_{[\tilde{\Delta}]}d_{[\Delta\setminus\tilde{\Delta}]}s}$ is the weighted sum of $\lambda_{(N)d_{[k]}s}$ over the fixed levels. To see this, recall that

$$\begin{aligned} \sum_{l_{\tilde{\delta}}=1}^{l_{\tilde{\delta},0}} \left\{(a_{f_{\tilde{\delta}}}^{(\tilde{\delta})})^T W_{l_{\tilde{\delta}}}^{(\tilde{\delta})}\right\}^2 &= \epsilon_{l_{\tilde{\delta},0}f_{\tilde{\delta}}}^T A^{(\tilde{\delta})} W^{(\tilde{\delta})} (W^{(\tilde{\delta})})^T A^{(\tilde{\delta})} \epsilon_{l_{\tilde{\delta},0}f_{\tilde{\delta}}} && (3.246) \\ &= \epsilon_{l_{\tilde{\delta},0}f_{\tilde{\delta}}}^T A^{(\tilde{\delta})} \epsilon_{l_{\tilde{\delta},0}f_{\tilde{\delta}}} && (3.247) \\ &= \alpha_{f_{\tilde{\delta}}f_{\tilde{\delta}}}^{(\tilde{\delta})}. && (3.248) \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{\tilde{\delta}'\in\tilde{\Delta}} \sum_{l_{\tilde{\delta}'}=1}^{l_{\tilde{\delta}',0}} \left[\prod_{\tilde{\delta}\in\tilde{\Delta}} \left\{(a_{f_{\tilde{\delta}}}^{(\tilde{\delta})})^T W_{l_{\tilde{\delta}}}^{(\tilde{\delta})}\right\}^2\right] &= \prod_{\tilde{\delta}\in\tilde{\Delta}} \left[\sum_{l_{\tilde{\delta}}=1}^{l_{\tilde{\delta},0}} \left\{(a_{f_{\tilde{\delta}}}^{(\tilde{\delta})})^T W_{l_{\tilde{\delta}}}^{(\tilde{\delta})}\right\}^2\right] && (3.249) \\ &= \prod_{\tilde{\delta}\in\tilde{\Delta}} \alpha_{f_{\tilde{\delta}}f_{\tilde{\delta}}}^{(\tilde{\delta})} && (3.250) \\ &= b_{\tilde{\Delta}}, && (3.251) \end{aligned}$$

confirming that we do, indeed, have a weighted sum.

3.7.2 The examiner considers slicing

The examiner decides that he wants to restrict his attention to the performance of each marker on the questions on the 1st paper. Thus, he wishes to consider the $f_{[paper=1]}$ slice. The $f_{[paper=1]}$ slice has two possible cells, $(1,1) = 11$ and $(1,2) = 12$. We restrict our attention to learning about$[\mathcal{M}(\mathcal{C}_1)]$, where $[\mathcal{M}(\mathcal{C}_1)] = [\mathcal{M}(\mathcal{C}_{11})^T\mathcal{M}(\mathcal{C}_{12})^T]^T$. From Theorem 24, and its' accompanying corollary, the examiner knows he has to make little additional effort to calculate the canonical resolutions and directions for the adjustment over $[\mathcal{M}(\mathcal{C}_1)]$. Using the matrix $A^{(2)}W^{(2)}$ (see equation (3.194)) we have

$$(a_1^{(2)})^T W_1^{(1)} = 0.9226;\ (a_1^{(2)})^T W_2^{(1)} = 0.3477; \tag{3.252}$$
$$(a_1^{(2)})^T W_3^{(1)} = 0.0881;\ (a_1^{(2)})^T W_4^{(1)} = 0.0084, \tag{3.253}$$

and $b_{[paper=1]} = 0.98$. The canonical resolutions of his adjustment are

$$\lambda_{1ds} = \frac{1}{0.98}\{0.9226^2\lambda_{(nI_8)1ds} + 0.3477^2\lambda_{(nI_8)2ds} + 0.0881^2\lambda_{(nI_8)3ds} + 0.0084^2\lambda_{(nI_8)4ds}\}. \tag{3.254}$$

The corresponding canonical directions of his adjustment are

$$Z_{1ds} = \frac{1}{\sqrt{0.98}}\{W_d^{(1)} \otimes Y_s\}^T\mathcal{M}(\mathcal{C}_1). \tag{3.255}$$

Notice that if the examiner wished to consider, for example, the $f_{[paper=2]}$ slice, the only change would be the weights. The canonical resolutions are still the weighted sum and the canonical directions will have the same co-ordinate representation.

Chapter 4

Modelling using finite second-order exchangeability

SUMMARY

To this point, our work has progressed under the assumption that the individuals in each second-order exchangeable sequence come from a potentially infinite population. This is an idealisation and we seek to remove it from our models. We consider progressing as if each second-order exchangeable sequence was only finite in length. Notice that this also allows us to consider sequences where there is no concept of an infinite length. In Section 4.3, we review the finite analogue to the representation theorem of Section 1.5 and in Section 4.5 develop the analogous theory for the adjustment of finite second-order exchangeable sequences as that reviewed in Subsection 1.8.2. In Subsection 4.5.1, we show that the coherency conditions between the adjustment of exchangeable sequences resulting from different sample sizes found in the infinite case are also present in the finite scenario. In Section 4.6, we consider extendible and non-extendible sequences and show that the adjustment is qualitatively the same for all lengths of extendible second-order exchangeable sequences. In Section 4.7, we extend this to infinite sequences and show how we may regard the difference between the adjustment of finite and infinite second-order exchangeable sequences, with the same specifications between individuals, as being quantitative and not qualitative. From Section 4.10 onwards, we develop the finite analogue

to Sections 2.1 - 2.4. We show that, once more, the adjustment of the population structure may be solved as a series of subproblems: the underlying canonical variable problem of Section 4.12 is identical to that of Subsection 2.4.1; the underlying canonical finite group problems of Section 4.13 can be seen, in the limit, to be identical to the underlying canonical group problem of Subsection 2.4.3. Section 4.14 is the finite analogue of Subsection 2.4.5. The simple examiner example is again used to illustrate the theory.

4.1 Introduction

Recall Definition 1 and the judgement of n events as being exchangeable. We stated that the infinite sequence of events A_1, A_2, ... are (infinitely) exchangeable if every finite subsequence is exchangeable. Definition 1 effectively defines the sequence A_1, ..., A_n to be a finitely exchangeable sequence of events and Definition 2 extends this definition to random quantities. Infinite exchangeability of sequences of random quantities is the case where every finite subsequence of random quantities is judged exchangeable in the sense of Definition 2.

So far we have proceeded with sequences that we judged to be second-order exchangeable and potentially infinite in length. This allowed us to make use of the representation theorem for second-order infinitely exchangeable sequences, see Theorem 2, and the representation theorem for sequences of co-exchangeable infinite second-order exchangeable sequences, see Theorem 7. These representation theorems introduce random quantities corresponding to underlying mean components and the adjustment of these mean components following sampling from the sequence have particularly attractive and illuminating features.

In practice however, the assumption of a sequence being potentially infinite is a modelling simplification, for it is usually possible to give an upper bound to the length of the sequence under consideration. For example, we have motivated our work through an examiner considering the marks on scripts for exams sat by individuals. In reality, the sequence of individuals who could sit the exams is not potentially infinite but finite; there will only be so many individuals registered for the exam. In consideration though, it is not always easy or straightforward to specify the upper bound for the sequence. In the examiner example, the number of potential candidates may not be available to the examiner and he may be faced with a cost in finding this information out. Moreover, any upper bounds he gives may be rather arbitrary. As such, it is often easier to proceed with the simplification that the sequence was of potentially infinite length.

As with any modelling assumption, we would like to investigate the consequences of the the infinite approximation and how it effects both our modelling and learning. In the infinite exchangeable case, we proceed with the assumption that observables $X_1, \ldots, X_m$ are assumed to be part of an infinite exchangeable sequence. In the case of full exchangeability, we may use the representation theorem developed by de Finetti (1937) and reproduced as Theorem 1 in this thesis. However, Diaconis (1977; p272) shows, via means of an example of an urn containing two balls, one marked 0 and the other 1 and drawing a sample of size two, without replacement, from the urn, that one serious problem with this representation is that it does require an infinite sequence and is false if the sequence is finite. Diaconis (1977) and Diaconis & Freedman (1980b) go on to develop analogous representations when the sequence is finite and show that the difference in probabilities, for the finite sequence of length m and the infinite approximation, assigned to an event goes like $(1/m)$, suggesting that the infinite approximation causes no important difference.

It seems tempting, as Bernardo & Smith (1994; p171) say, "to wonder whether *every* finite sequence of exchangeable random quantities could be embedded in or extended to an infinitely exchangeable sequence of similarly defined random quantities". They produce an example which shows that this is not true, and that indeed a finitely exchangeable sequence cannot always be embedded into a larger finitely exchangeable sequence. This example is based on pure mathematics, of course there will be situations where there is no concept of being able to extend the sequence; the random quantities themselves may have no notion of belonging to a potentially infinite sequence. For example, excluding the Preface and Contents, I might be interested in the sequence of random quantities $X_1, \ldots, X_{586}$ where X_i is the number of occurrences of the word 'aesthetic' on page i of Bernardo & Smith (1994). I may judge this sequence to be exchangeable, but there is no notion of it being extendible; there are only 586 pages. Even if there was a sequence of infinite exchangeable random quantities for which I judged every subsequence of length 586 to have the same mathematical beliefs as $X_1, \ldots, X_{586}$, it would seem bizarre to use this as a basis for introducing the representation theorem for infinite random quantities over the sequence $X_1, \ldots, X_{586}$. Thus, the study of finite exchangeable sequences has a broader scope than the study of infinite exchangeable sequences for it will include exchangeable sequences that have no analogue in the infinite setting.

Hence, we see that the study of finite extendible sequences is important for two distinct reasons. If an exchangeable sequence is conceptually and mathematically extendible to an exchangeable sequence of longer length, indeed of possible infinite length, then we may desire to compare the consequences of the sequence length used in any modelling assumptions when the real length of the sequence is unknown. For example, before I physically turned to the last page and read its number, I didn't know the length of the sequence X_i, yet I would still have been willing to judge it exchangeably. I could, though, have put an upper bound on the length; I certainly knew it was finite. Notice that any exchangeable sequence we could conceive as being an extension of a subsequence of length two. We proceed as follows. We investigate the adjustment of finite exchangeable sequences of length m and show how these may be linked to the adjustment of sequences of length 2. This then allows us to study both infinitely extendible sequences and those that are not extendible, or only finitely so. For sequences of length m that are infinitely extendible, we draw the comparison with the adjustment of the infinite sequence into which the finite sequence is embedded. We then draw comparisons between the grouped modelling in Chapter 2 with the case when the individuals in each group are only finitely exchangeable. Once more, we proceed with second-order beliefs since the finite judgement does not overcome our doubts about being able to specify full probability models.

4.2 Finite second-order exchangeability

We are interested in a making a series of measurements $\mathcal{C} = \{X_1, X_2, \dots\}$, finite or infinite, on a finite collection of individuals. We shall consider that there are at most m individuals. The collection for each individual is generated from $\mathcal{C}$ by letting $\mathcal{C}_i = \{X_{1i}, X_{2i}, \dots\}$ be the measurements for the ith individual. We specify directly the prior means, variance and covariance for each pair of quantities and judge that the collection of measurements $\mathcal{C}$ is second-order exchangeable over the the full collection $\mathcal{C}^*$, see Definition 3 in Subsection 1.5.3 for full details; this also gives us our second-order specification. Again, to conform with the geometric representation of $[\mathcal{C}^*]$ as an inner product space, we standardise our quantities by subtracting the mean. Since we are considering that there are only m possible individuals that we could observe, $\mathcal{C}^*$ is the union of a finite number of individuals' collections, and so we say that the sequence of individuals is finitely second-order exchangeable.

Observe that whatever the number of specifications of prior means, variances and covariances involved in the specification of the full collection $\mathcal{C}^*$, all that is actually required is the consideration of two cases. The symmetries that we judge between the individuals means that the specification of the complete collection follows. Thus, in terms of specification, there is nothing different in how we may regard a finite or an infinite second-order exchangeable sequence. The second-order judgement makes less symmetry requirements upon our beliefs than the full finite exchangeability requirement of symmetry judgements related to the sequence length; see, for example, Definition 2.

However, the length of the sequence under consideration does effect the coherency statements we make between the individuals. We have to ensure that we do not specify any quantities in the full collection $\mathcal{C}^*$ to have negative variation. The following statement clarifies the restriction of the choice of the d_{vw} and c_{vw}, where $d_{vw} = Cov(X_{vi}, X_{wi})$ and $c_{vw} = Cov(X_{vi}, X_{wj})$ as given in Definition 3.

Statement 1 *Suppose that $\mathcal{C}^*$ is the union of m individuals' collections. Letting $\{v_1, v_2, \dots\}$ be a general finite subset of integers, $\eta = [\eta_1 \eta_2 \dots]^T$ be a general real valued vector, then d_{vw} and c_{vw} must satisfy the following:*

$$\sum_u \sum_{u'} \eta_u \eta_{u'} d_{v_u v_{u'}} \geq 0; \tag{4.1}$$

$$\sum_u \sum_{u'} \eta_u \eta_{u'} (d_{v_u v_{u'}} - c_{v_u v_{u'}}) \geq 0; \tag{4.2}$$

$$\sum_u \sum_{u'} \eta_u \eta_{u'} (d_{v_u v_{u'}} + (m-1) c_{v_u v_{u'}}) \geq 0. \tag{4.3}$$

Proof - If the three conditions above hold, then $[\mathcal{C}^*]$ is an inner-product space. The first follows by considering the variance of $Y_i = \sum_u \eta_u X_{v_u i}$, and the second by considering the variance of $Y_i - Y_j = \sum_u \eta_u (X_{v_u i} - X_{v_u j})$. The third condition follows from considering the variance of a general quantity $Y = \sum_{i=1}^m \sum_u a_i \eta_u X_{v_u i}$. There are two cases to consider:

Firstly, when $\sum_u \sum_{u'} \eta_u \eta_{u'} c_{v_u v_{u'}} \geq 0$. In this case, we find:

$$
\begin{aligned}
Var(Y) &= \sum_u \sum_{u'} \eta_u \eta_{u'} \left\{ \sum_{i=1}^{m} a_i^2 Cov(X_{v_u i}, X_{v_{u'} i}) + \right. \\
&\qquad \left. \sum_{i=1}^{m} \sum_{j \neq i}^{m} a_i a_j Cov(X_{v_u i}, X_{v_{u'} j}) \right\} && (4.4) \\
&= \sum_u \sum_{u'} \eta_u \eta_{u'} \left\{ d_{v_u v_{u'}} \sum_{i=1}^{m} a_i^2 + c_{v_u v_{u'}} \sum_{i=1}^{m} \sum_{j \neq i}^{m} a_i a_j \right\} && (4.5) \\
&= \left\{ \sum_u \sum_{u'} \eta_u \eta_{u'} (d_{v_u v_{u'}} - c_{v_u v_{u'}}) \right\} \left\{ \sum_{i=1}^{m} a_i^2 \right\} + \\
&\qquad \left\{ \sum_u \sum_{u'} \eta_u \eta_{u'} c_{v_u v_{u'}} \right\} \left\{ \sum_{i=1}^{m} a_i \right\}^2 && (4.6) \\
&\geq 0, && (4.7)
\end{aligned}
$$

where (4.7) follows by condition (4.2) and the assumption.

Secondly, when $\sum_u \sum_{u'} \eta_u \eta_{u'} c_{v_u v_{u'}} < 0$. In this case, we find:

$$
\begin{aligned}
Var(Y) &= \sum_u \sum_{u'} \eta_u \eta_{u'} \left\{ d_{v_u v_{u'}} \sum_{i=1}^{m} a_i^2 + c_{v_u v_{u'}} \sum_{i=1}^{m} \sum_{j \neq i}^{m} a_i a_j \right\} && (4.8) \\
&= \sum_u \sum_{u'} \eta_u \eta_{u'} \left\{ [d_{v_u v_{u'}} + (m-1) c_{v_u v_{u'}}] \sum_{i=1}^{m} a_i^2 - \right. \\
&\qquad \left. c_{v_u v_{u'}} \sum_{i=1}^{m} \sum_{j \neq i}^{m} a_i (a_i - a_j) \right\} && (4.9) \\
&= \left\{ \sum_u \sum_{u'} \eta_u \eta_{u'} [d_{v_u v_{u'}} + (m-1) c_{v_u v_{u'}}] \right\} \left\{ \sum_{i=1}^{m} a_i^2 \right\} - \\
&\qquad \left\{ \sum_u \sum_{u'} \eta_u \eta_{u'} c_{v_u v_{u'}} \right\} \left\{ \sum_{i=1}^{m} \sum_{j > i}^{m} (a_i - a_j)^2 \right\} && (4.10) \\
&\geq 0, && (4.11)
\end{aligned}
$$

where (4.11) follows by condition (4.3) and the assumption. □

4.2.1 The examiner's considerations

To highlight the parallels between the finite and infinite cases, we assume that the examiner in confronted with the same problem as in Chapter 1. For each individual sampled he will make the same set of measurements $\mathcal{C} = \{X_1, X_2\}$, where X_v is the mark on the vth question. The only difference between the scenario in Chapter 1 and the current one is that the examiner wishes to make his model more realistic in that he accepts that there are only a finite number, m, of possible individuals to sample from. Everything else is unchanged.

To aid with the comparison, we assume, indeed there is no discernible reason why he should change his opinion, that the examiner makes the same specifications over the individuals as he did previously. That is he makes the following specification:

$$E(X_{vi}) = 0 \ \forall v, i; \tag{4.12}$$

$$Cov(X_{vi}, X_{wi}) = \begin{cases} 5 & \text{if } v = w, \ \forall i; \\ 3.75 & \text{if } v \neq w, \ \forall i; \end{cases} \tag{4.13}$$

$$Cov(X_{vi}, X_{wj}) = \begin{cases} 1 & \text{if } v = w, \ \forall i \neq j; \\ 0.25 & \text{if } v \neq w, \ \forall i \neq j. \end{cases} \tag{4.14}$$

4.3 The representation theorem for finite second-order exchangeable beliefs

The representation theorem for infinite second-order exchangeable sequences is only valid if the sequence is infinite. However, all is not lost, for in an analogous way to the representation theorem for infinite second-order exchangeable sequences, reproduced as Theorem 2 in this thesis, Goldstein (1986a) derives a representation theorem for finitely second-order exchangeable measurements as follows.

Theorem 25 *If C is second-order exchangeable over the (finite) individuals, then we may introduce the further collections of random quantities $\tilde{\mathcal{M}}(C) = \{\tilde{\mathcal{M}}(X_1), \tilde{\mathcal{M}}(X_2), \ldots, \}$, and $\tilde{\mathcal{R}}_i(C) = \{\tilde{\mathcal{R}}_i(X_1), \tilde{\mathcal{R}}_2(X), \ldots\}$, and write*

$$X_{vi} = \tilde{\mathcal{M}}(X_v) + \tilde{\mathcal{R}}_i(X_v), \tag{4.15}$$

where

$$\tilde{\mathcal{M}}(X_v) = \frac{1}{m}\sum_{i=1}^{m} X_{vi}. \tag{4.16}$$

The collections $\tilde{\mathcal{M}}(C)$ and $\tilde{\mathcal{R}}_i(C)$ satisfy the following relationships

$$E(\tilde{\mathcal{M}}(X_v)) = m_v \ \forall v; \tag{4.17}$$

$$E(\tilde{\mathcal{R}}_i(X_v)) = 0 \ \forall v, i; \tag{4.18}$$

$$Cov(\tilde{\mathcal{M}}(X_v), \tilde{\mathcal{M}}(X_w)) = c_{vw} + \frac{1}{m}(d_{vw} - c_{vw}) \ \forall v, w; \tag{4.19}$$

$$Cov(\tilde{\mathcal{M}}(X_v), \tilde{\mathcal{R}}_j(X_w)) = 0 \ \forall v, w, j; \tag{4.20}$$

$$Cov(\tilde{\mathcal{R}}_i(X_v), \tilde{\mathcal{R}}_j(X_w)) = \begin{cases} \frac{m-1}{m}(d_{vw} - c_{vw}) & \text{if } i = j \ \forall v, w; \\ -\frac{1}{m}(d_{vw} - c_{vw}) & \text{otherwise.} \end{cases} \tag{4.21}$$

Notice that as we are dealing with only a finite number of individuals, it is conceivable that we sample each possible individual. The quantities in $\tilde{\mathcal{M}}(\mathcal{C})$ and in the $\tilde{\mathcal{R}}_i(\mathcal{C})$ are observable; this contrasts with the unobservable quantities in the infinitely exchangeable representation theorem. In this case, $\tilde{\mathcal{M}}(\mathcal{C})$ is the population mean collection for the individuals and the collections $\tilde{\mathcal{R}}_i(\mathcal{C})$ are the individual discrepancies from the overall means. The representation theorem may be extended to encompass linear combinations in an identical fashion to the infinite case. We let $\langle\tilde{\mathcal{M}}(\mathcal{C})\rangle$ and $\langle\tilde{\mathcal{R}}_i(\mathcal{C})\rangle$ be, respectively, the collection of finite linear combinations of the elements of $\tilde{\mathcal{M}}(\mathcal{C})$and $\tilde{\mathcal{R}}_i(\mathcal{C})$. Then for $\mathcal{X} \in \langle\mathcal{C}\rangle$, for each individual i, we have

$$\mathcal{X}_i = \tilde{\mathcal{M}}(\mathcal{X}) + \tilde{\mathcal{R}}_i(\mathcal{X}), \tag{4.22}$$

where $\tilde{\mathcal{M}}(\mathcal{X}) \in \langle\tilde{\mathcal{M}}(\mathcal{C})\rangle$ and $\tilde{\mathcal{R}}_i(\mathcal{X}) \in \langle\tilde{\mathcal{R}}_i(\mathcal{C})\rangle$.

4.3.1 The examiner applies the representation theorem

The examiner may then apply this representation theorem to his beliefs. He may introduce the collection of random quantities $\tilde{\mathcal{M}}(\mathcal{C}) = \{\tilde{\mathcal{M}}(X_1), \tilde{\mathcal{M}}(X_2)\}$, the collection of mean scores for the questions, and $\tilde{\mathcal{R}}_i(\mathcal{C}) = \{\tilde{\mathcal{R}}_i(X_1), \tilde{\mathcal{R}}_i(X_2)\}$, the individual residuals from the means and write:

$$X_{vi} = \tilde{\mathcal{M}}(X_v) + \tilde{\mathcal{R}}_i(X_v). \tag{4.23}$$

The newly introduced random quantities have the following induced beliefs:

$$E(\tilde{\mathcal{M}}(X_v)) = 0 \; \forall v; \tag{4.24}$$

$$E(\tilde{\mathcal{R}}_i(X_v)) = 0 \; \forall v, i; \tag{4.25}$$

$$Cov(\tilde{\mathcal{M}}(X_v), \tilde{\mathcal{M}}(X_w)) = \begin{cases} 1 + \frac{4}{m} & \text{if } v = w; \\ \frac{1}{4} + \frac{7}{2m} & \text{if } v \neq w; \end{cases} \tag{4.26}$$

$$Cov(\tilde{\mathcal{M}}(X_v), \tilde{\mathcal{R}}_j(X_w)) = 0 \; \forall v, w, j; \tag{4.27}$$

$$Cov(\tilde{\mathcal{R}}_i(X_v), \tilde{\mathcal{R}}_j(X_w)) = \begin{cases} \frac{4(m-1)}{m} & \text{if } i = j \text{ and } v = w; \\ \frac{7(m-1)}{2m} & \text{if } i = j \text{ and } v \neq w; \\ -\frac{4}{m} & \text{if } i \neq j \text{ and } v = w; \\ -\frac{7}{2m} & \text{if } i \neq j \text{ and } v \neq w. \end{cases} \tag{4.28}$$

4.4 The examiner's learning

Suppose, once more, that the examiner receives the scores on a single exam paper, and wishes to scrutinize the effect of those scores on his prior beliefs about the

Component	Resolution (sample size 1)
$\tilde{\mathcal{M}}(X_+)$	$\tilde{\lambda}_+ = \frac{m+6}{7m} = \frac{(m-1)\lambda_+ + 1}{m}$
$\tilde{\mathcal{M}}(X_-)$	$\tilde{\lambda}_- = \frac{3m+2}{5m} = \frac{(m-1)\lambda_- + 1}{m}$
$l_1\tilde{\mathcal{M}}(X_1) + l_2\tilde{\mathcal{M}}(X_2)$	$\frac{(5m+30)(l_1+l_2)^2\tilde{\lambda}_+ + (3m+2)(l_1-l_2)^2\tilde{\lambda}_-}{(8m+32)(l_1^2+l_2^2)+(4m+56)l_1 l_2}$

Table 4.1: Resolutions for the quantities in $[\tilde{\mathcal{M}}(\mathcal{C})]$, having seen the scores on a single exam paper

remaining individuals. He remains interested in learning about various linear combinations of the $\tilde{\mathcal{M}}(\mathcal{C})$. In the example of Chapter 1, we showed that the canonical resolutions in the finite case were proportional to $\mathcal{M}(X_+)$ and $\mathcal{M}(X_-)$. The theory developed in Goldstein & Wooff (1998) shows that the canonical resolutions for all samples sizes are the same. In this finite case, we shall focus upon the analogous quantities, namely

$$\tilde{\mathcal{M}}(X_+) = \tilde{\mathcal{M}}(X_1) + \tilde{\mathcal{M}}(X_2), \tag{4.29}$$

representing the total scored on the paper and

$$\tilde{\mathcal{M}}(X_+) = \tilde{\mathcal{M}}(X_1) - \tilde{\mathcal{M}}(X_2), \tag{4.30}$$

representing the difference in marks between the first and the second question. The quantities $\tilde{\mathcal{M}}(X_+)$ and $\tilde{\mathcal{M}}(X_-)$ are mutually uncorrelated and thus form an orthogonal basis for $[\tilde{\mathcal{M}}(\mathcal{C})]$, so that any element $l_1\tilde{\mathcal{M}}(X_1) + l_2\tilde{\mathcal{M}}(X_2) \in [\tilde{\mathcal{M}}(\mathcal{C})]$ may be expressed as a linear combination of the two stated quantities. The examiner draws up Table 4.1 to show the effect of receiving the scores on a single paper for his beliefs over the population mean collection.

The resolutions in Table 4.1 should be compared with those in Table 1.1 in Subsection 1.7.3. Compare the resolution for $\tilde{\mathcal{M}}(X_+)$ with $\mathcal{M}(X_+)$. We observe that in the finite case, the resolution is greater, and by expressing $\tilde{\lambda}_+$ in terms of λ_+, we see that the increment is $((1 - \lambda_+)/m)$. Now compare the resolution for $\tilde{\mathcal{M}}(X_-)$ with that of $\mathcal{M}(X_-)$. The identical comments apply; $\tilde{\lambda}_-$ is larger than λ_-, the difference being $((1 - \lambda_-)/m)$. It is no surprise that in the finite case, the resolutions are larger - when we take a sample we are actually seeing a definable amount of the population mean (in this case a $1/m$th of it) whereas in the infinite case, there is no such comparison of actually 'seeing' part of the underlying mean. The interest is that the difference in the finite and infinite resolutions takes the same form, hence creating a regular link between finite and infinite modelling. Observe that the resolution of $l_1\tilde{\mathcal{M}}(X_1) + l_2\tilde{\mathcal{M}}(X_2)$ may be found as a weighted average of

Component	Resolution (sample size n)
$\tilde{\mathcal{M}}(X_+)$	$\tilde{\lambda}_{(n)+} = \frac{n(m-1)\tilde{\lambda}_+}{(n-1)m\tilde{\lambda}_+ + (m-n)}$
$\tilde{\mathcal{M}}(X_-)$	$\tilde{\lambda}_{(n)-} = \frac{n(m-1)\tilde{\lambda}_-}{(n-1)m\tilde{\lambda}_- + (m-n)}$
$l_1\tilde{\mathcal{M}}(X_1) + l_2\tilde{\mathcal{M}}(X_2)$	$\frac{(5m+30)(l_1+l_2)^2\tilde{\lambda}_{(n)+} + (3m+2)(l_1-l_2)^2\tilde{\lambda}_{(n)-}}{(8m+32)(l_1^2+l_2^2)+(4m+56)l_1l_2}$

Table 4.2: Resolutions for the quantities in $[\tilde{\mathcal{M}}(\mathcal{C})]$ having seen the scores on the first n exam papers

the resolutions of $\tilde{\mathcal{M}}(X_+)$ and $\tilde{\mathcal{M}}(X_-)$. The canonical directions for this adjustment are thus proportional to $\tilde{\mathcal{M}}(X_+)$ and $\tilde{\mathcal{M}}(X_-)$. Hence, the canonical directions for the finite adjustment share, up to a scale factor to ensure a prior variance of one, the same co-ordinate representation as the canonical directions in the infinite case. It should be emphasised that these observations do not depend upon the length, m, of the finite sequence. The only place m has an effect is on the weighting between the two canonical resolutions for the general quantity, $l_1\tilde{\mathcal{M}}(X_1)+l_2\tilde{\mathcal{M}}(X_2)$. Indeed, note that as $m \to \infty$, then the weights are the same as those for $l_1\mathcal{M}(X_1)+l_2\mathcal{M}(X_2)$ (see Table 1.1).

Having observed these links between the finite and infinite assumptions for a single paper, we may wonder whether there are similarly strong coherence relations between adjustments in the finite case for samples of n papers as there was in the infinite case. Thus, the examiner would like to investigate the effect of the adjustment if he saw the scores on the first n scripts, where $n \leq m$. The examiner thus performs a Bayes linear adjustment of his beliefs having seen the scores of the first n individuals and once more calculates the resolutions in order to help him assess the effect of the adjustment. To investigate the relationships between sample sizes, he focuses attention upon the two quantities, $\tilde{\mathcal{M}}(X_+)$ and $\tilde{\mathcal{M}}(X_-)$. Table 4.2 summarises his calculations. There are a number of points to make, firstly between the finite sequence for changing the sample size, and secondly between different sequence lengths. It is immediate from the table that the canonical directions are proportional to $\tilde{\mathcal{M}}(X_+)$ and $\tilde{\mathcal{M}}(X_-)$. The elegant features of the adjustment of the underlying population structure induced by a second-order infinitely exchangeable sample remain true when we adopt the more realistic modelling assumption of finite second-order exchangeability. Irrespective of sample size, we may form an orthogonal grid which summarises the adjustment. We expect to learn most about quantities that are highly correlated with $\tilde{\mathcal{M}}(X_-)$ and least about those highly correlated with $\tilde{\mathcal{M}}(X_+)$. Thus, the effect of changing the sample size is quantitative, by modifying

Quantity	Sample size required for variance reduction of		
	50 percent	90 percent	95 percent
$\mathcal{M}(X_+)$	6	54	114
$\tilde{\mathcal{M}}(X_+)$, $m = 1000$	6	51	102
$\tilde{\mathcal{M}}(X_+)$, $m = 250$	6	43	76
$\tilde{\mathcal{M}}(X_+)$, $m = 50$	5	24	33
$\mathcal{M}(X_-)$	1	6	13
$\tilde{\mathcal{M}}(X_-)$, $m = 1000$	1	6	13
$\tilde{\mathcal{M}}(X_-)$, $m = 250$	1	6	12
$\tilde{\mathcal{M}}(X_-)$, $m = 50$	1	6	10

Table 4.3: Table showing the resolutions required to achieve the specified variance reductions in the given quantities

the canonical resolutions, and not qualitative, as the canonical directions are the same for all choices of sample size. For any choice of m, this grid has the same form; the canonical resolutions have, up to a scale factor to ensure a prior variance of one the same co-ordinate representation. This holds not just for finite m, but in the limit when we take m to infinity and we have an infinitely exchangeable sequence. Thus, in this case, for any length of sequence (provided the specification is coherent) for which the judgements between each pair of individuals is the same then the underlying features of the adjustment of the second-order exchangeable sequence induced by a second-order exchangeable sample remains the same not just for all possible sample sizes, n, but also for all possible sequence lengths m. In particular, it is straightforward for the examiner to assess the impact of an infinite assumption upon each quantity as the qualitative information is easily comparable.

One possible comparison may be to investigate the difference in sample sizes required to achieve a specific variance reduction for quantities of interest. The examiner constructs Table 4.3 to summarise these sample sizes. Hence, if we have $m = 250$, then the cost of the infinite assumption in achieving a proportionate variance reduction of 95 percent for $\tilde{\mathcal{M}}(X_-)$ for the analogous quantity $\mathcal{M}(X_-)$ is a single individual. As $\tilde{\mathcal{M}}(X_-)$ is proportional to the canonical direction with largest resolution, $\tilde{\mathcal{M}}(X_-)$ achieves the maximum variance reduction for quantities in $[\tilde{\mathcal{M}}(\mathcal{C})]$. As $\tilde{\mathcal{M}}(X_+)$ is proportional to the canonical direction with largest resolution, $\tilde{\mathcal{M}}(X_+)$ achieves the minimum variance reduction for quantities in $[\tilde{\mathcal{M}}(\mathcal{C})]$. The analogous comment is true for $\mathcal{M}(X_+)$ in $[\mathcal{M}(\mathcal{C})]$. Thus, if $m = 250$, the greatest additional number of individuals we would have to sample to ensure that all quantities in the space of interest received a variance reduction of 95 percent is $114 - 76 = 38$, an increase of 50 percent in the sample size required. Notice that if

$m = 1000$, this falls to $114 - 102 = 12$ further individuals.

4.5 Learning about the mean components

Suppose that we are about to observe a sample of $n \leq m$ exchangeable collections, which we label for convenience, $\mathcal{C}_1, \ldots, \mathcal{C}_n$ and we let $\mathcal{C}(n) = \cup_{i=1}^{n} \mathcal{C}_i$. We then want to use this data to revise our beliefs over the mean collection, $\tilde{\mathcal{M}}(\mathcal{C})$. In a similar style to Goldstein & Wooff (1998), we use the following notational shorthands for our adjusted quantities:

$$E_n(X) = E_{\mathcal{C}(n)}(X); \quad Var_n(X) = Var_{\mathcal{C}(n)}(X); \quad E_{\tilde{\mathcal{M}}}(X) = E_{\tilde{\mathcal{M}}(\mathcal{C})}(X). \tag{4.31}$$

For each $\mathcal{X} \in [\mathcal{C}]$, denote the average of the first n values by

$$\mathcal{S}_n(\mathcal{X}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i, \tag{4.32}$$

and the collection of average values by $\mathcal{S}_n(\mathcal{C}) = \{\mathcal{S}_n(X_1), \mathcal{S}_n(X_2), \ldots\}$. For each $X_{vi} \in \mathcal{C}(N)$ let

$$\mathcal{T}_i(X_v) = X_{vi} - \mathcal{S}_n(X_v), \tag{4.33}$$

so that $\mathcal{T}_i(\mathcal{C}(n)) = \{\mathcal{T}_i(X_1), \mathcal{T}_i(X_2), \ldots\}$ is the collection of residuals of the ith individual from the collection of sample means. The complete collection of sample mean residuals is then $\mathcal{T}(\mathcal{C}(n)) = \cup_{i=1}^{n} \mathcal{T}_i(\mathcal{C}(n))$. We have the following lemma.

Lemma 21 *The second-order relationships between the $\tilde{\mathcal{M}}(\mathcal{C})$s, the $\mathcal{S}_n(\mathcal{C})$s and the $\mathcal{T}_i(\mathcal{C}(n))$s may be expressed as*

$$Cov(\tilde{\mathcal{M}}(X_v), \mathcal{S}_n(X_w)) = c_{vw} + \frac{1}{m}(d_{vw} - c_{vw}) \ \forall v, w; \tag{4.34}$$

$$Cov(\mathcal{S}_n(X_v), \mathcal{S}_n(X_w)) = c_{vw} + \frac{1}{n}(d_{vw} - c_{vw}) \ \forall v, w; \tag{4.35}$$

$$Cov(\mathcal{S}_n(X_v), \mathcal{T}_j(X_w)) = 0 \ \forall v, w, j; \tag{4.36}$$

$$Cov(\tilde{\mathcal{M}}(X_v), \mathcal{T}_j(X_w)) = 0 \ \forall v, w, j; \tag{4.37}$$

$$Cov(\mathcal{T}_i(X_v), \mathcal{T}_j(X_w)) = \begin{cases} \frac{n-1}{n}(d_{vw} - c_{vw}) & if\ i = j;\ \forall v, w; \\ -\frac{1}{n}(d_{vw} - c_{vw}) & if\ i \neq j;\ \forall v, w. \end{cases} \tag{4.38}$$

Proof - The results follow from the definitions of the $\mathcal{S}_n(X_v)$s, the $\mathcal{T}_i(X_v)$s and the specifications given by equations (4.19) - (4.21). □

We now show that the collection of sample means is Bayes linear sufficient for the full sample for adjusting the collection of mean components.

Theorem 26 *If $\mathcal{C}$ is (finitely) exchangeable over individuals, then $\mathcal{S}_n(\mathcal{C}) = \{\mathcal{S}_n(X_1), \mathcal{S}_n(X_2), \dots\}$ is Bayes linear sufficient for $\mathcal{C}(n)$ for adjusting the collection $\mathcal{M}(\mathcal{C})$.*

Proof - From equation (4.36) we have that $[\mathcal{S}_n(\mathcal{C})] \perp [\mathcal{T}_i(\mathcal{C}(n))]$ for each i and from equation (4.37) we have that $[\tilde{\mathcal{M}}(\mathcal{C})] \perp [\mathcal{T}_i(\mathcal{C}(n))]$ for each i. Letting $\mathcal{T}(\mathcal{C}(n)) = \cup_i \mathcal{T}_i(\mathcal{C}(n))$ we have that

$$[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{C}(n)] = [\tilde{\mathcal{M}}(\mathcal{C})/(\mathcal{S}_n(\mathcal{C}) + \mathcal{T}(\mathcal{C}(n)))] \tag{4.39}$$

$$= [[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{T}(\mathcal{C}(n))]/[\mathcal{S}_n(\mathcal{C})/\mathcal{T}(\mathcal{C}(n))]] \tag{4.40}$$

$$= [\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_n(\mathcal{C})], \tag{4.41}$$

and hence the result. □

Thus, if we want to adjust the population mean collection corresponding to a finite sequence of second-order exchangeable systems by observing n individuals, then we only need the sample means to perform the adjustment.

4.5.1 Resolution transforms for (finitely) exchangeable systems

In Chapter 1, we explained how Goldstein & Wooff (1998) used the resolution transform to investigate the coherence conditions imposed on second-order infinitely exchangeable adjustments. We now show that there are similar coherence relationships obtained between adjustments based upon second-order finitely exchangeable samples. Denote the resolution transform for the mean collection $\mathcal{M}(\mathcal{C})$, based on n observations, $\mathcal{C}(n) = \bigcup_{r=1}^{n} \mathcal{C}_r$, by $\tilde{T}_n = E_{\tilde{\mathcal{M}}}\{E_n(\cdot)\}$. We have the following theorem.

Theorem 27 *The eigenvectors of $\tilde{T}_n$ are the same for each n. If $\tilde{Y}$ is an eigenvector of $\tilde{T}_1$ with corresponding eigenvalue $\tilde{\lambda}$, then the corresponding eigenvalue $\tilde{\lambda}_{(n)}$ for $\tilde{T}_n$ is*

$$\tilde{\lambda}_{(n)} = \frac{n(m-1)\tilde{\lambda}}{(n-1)m\tilde{\lambda} + (m-n)}. \tag{4.42}$$

Proof - By linear sufficiency, we have for any $\mathcal{X} \in [\mathcal{C}]$,

$$E_n(\tilde{\mathcal{M}}(\mathcal{X})) = E_{\mathcal{S}(n)}(\tilde{\mathcal{M}}(\mathcal{X})). \tag{4.43}$$

The space $[\mathcal{S}(n)]$ is the collection of all elements of the form

$$\mathcal{S}_n(\mathcal{X}) = \tilde{\mathcal{M}}(\mathcal{X}) + \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(\mathcal{X}), \tag{4.44}$$

for $\mathcal{X} \in [\mathcal{C}]$. Therefore, for any $W \in [\tilde{\mathcal{M}}(\mathcal{C})]$, the projection

$$E_{S(n)}(W) = \tilde{\mathcal{M}}(\mathcal{X}_W) + \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(\mathcal{X}_W), \tag{4.45}$$

for some $\mathcal{X}_W \in [\mathcal{C}]$. Therefore, as $\tilde{\mathcal{R}}_i(\mathcal{X}_W) \perp [\tilde{\mathcal{M}}(\mathcal{C})]\ \forall i, \mathcal{X}_W \in [\mathcal{C}]$, we have for any $W \in [\tilde{\mathcal{M}}(\mathcal{C})]$, for each n,

$$\begin{aligned} E_{\tilde{\mathcal{M}}}\{E_{S(n)}(W)\} &= E_{\tilde{\mathcal{M}}}\{\tilde{\mathcal{M}}(\mathcal{X}_W) + \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(\mathcal{X}_W)\} && (4.46)\\ &= \tilde{\mathcal{M}}(\mathcal{X}_W). && (4.47)\end{aligned}$$

Suppose that $\tilde{Y}$ is an eigenvector of $\tilde{T}_1$, with eigenvalue $1 > \tilde{\lambda} > 0$. Then from equation (4.44) we have

$$E_1(\tilde{Y}) = \tilde{\mathcal{M}}(U_{\tilde{Y}}) + \tilde{\mathcal{R}}_1(U_{\tilde{Y}}), \tag{4.48}$$

for some $U_{\tilde{Y}} \in [\mathcal{C}]$. We have

$$\begin{aligned} \tilde{\lambda}\tilde{Y} &= \tilde{T}_1(\tilde{Y}) && (4.49)\\ &= E_{\tilde{\mathcal{M}}}\{E_1(\tilde{Y})\} && (4.50)\\ &= \tilde{\mathcal{M}}(U_{\tilde{Y}}), && (4.51)\end{aligned}$$

with equation (4.51) following via sufficiency and equation (4.47). Therefore, $\tilde{Y}$ is an eigenvector of $\tilde{T}_1$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$,

$$Cov(\tilde{Y} - \tilde{\lambda}\tilde{Y} - \tilde{\mathcal{R}}_1(U_{\tilde{Y}}), \tilde{\mathcal{M}}(\mathcal{X}) + \tilde{\mathcal{R}}_1(\mathcal{X})) = 0, \tag{4.52}$$

or equivalently, if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1 - \tilde{\lambda})Cov(\tilde{Y}, \tilde{\mathcal{M}}(\mathcal{X})) = Cov(\tilde{\mathcal{R}}_1(U_{\tilde{Y}}), \tilde{\mathcal{R}}_1(\mathcal{X})). \tag{4.53}$$

Similarly, $\tilde{Y}$ is an eigenvector of $\tilde{T}_n$, with eigenvalue μ, if and only if, for some $Z_{\tilde{Y}} \in [\mathcal{C}]$,

$$E_n(\tilde{Y}) = \mu\tilde{Y} + \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(Z_{\tilde{Y}}), \tag{4.54}$$

or equivalently, if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1 - \mu)Cov(\tilde{Y}, \tilde{\mathcal{M}}(\mathcal{X})) = Cov(\frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(Z_{\tilde{Y}}), \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(\mathcal{X})). \tag{4.55}$$

Then as:

$$Cov(\frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(U), \frac{1}{n}\sum_{i=1}^{n} \tilde{\mathcal{R}}_i(V)) = \frac{(m-n)}{n(m-1)}Cov(\tilde{\mathcal{R}}_1(U), \tilde{\mathcal{R}}_1(V)) \tag{4.56}$$

for all $U, V \in [\mathcal{C}]$, we have that $U_{\tilde{Y}}$ satisfies (4.53), with eigenvalue $\tilde{\lambda}$, if and only if

$$Z_{\tilde{Y}} = \frac{n(m-1)}{(n-1)m\tilde{\lambda} + (m-n)} U_{\tilde{Y}} \tag{4.57}$$

satisfies (4.55) with eigenvalue $\mu = \tilde{\lambda}_{(n)}$ □

Goldstein & Wooff (1998) derive the corresponding result for infinitely exchangeable sequences by deriving the equivalent form for (4.56). The points they raise remain valid here. For the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C})]$ by $\mathcal{C}(n)$, the canonical directions $\tilde{Z}_1 = Var(\tilde{Y}_1)^{-\frac{1}{2}}\tilde{Y}_1$, $\tilde{Z}_2 = Var(\tilde{Y}_2)^{-\frac{1}{2}}\tilde{Y}_2$, ... are the same for each n, and are termed the canonical directions induced by (finite) exchangeability. Thus, the qualitative features of the adjustment are the same for all sample sizes and the effect of changing the sample size is to refashion the canonical resolutions in the manner given in equation (4.42). Compare the resolutions with those given in Table 4.2 for the examiner's problem when he receives the scores of the first n papers. Notice that from equation (1.63) we have that, for any sample size n, the adjusted variance for any $Y \in [\tilde{\mathcal{M}}(\mathcal{C})]$ is given by

$$Var_n(Y) = \sum_i \frac{(m-n)(1-\tilde{\lambda}_i)}{(n-1)m\tilde{\lambda}_i + (m-n)} Cov(Y, \tilde{Z}_i)^2, \tag{4.58}$$

and the resolution may be similarly expressed. Thus, in an identical manner to Goldstein & Wooff (1998), we see how we may exploit equation (4.42) to simplify design problems for which we are required to choose the sample size to achieve a specified variance reduction in elements of interest in the $[\tilde{\mathcal{M}}(\mathcal{C})]$. We have the following corollary.

Corollary 13 *Suppose that $\tilde{Y}$ is an eigenvector of $\tilde{T}_1$ with eigenvalue $\tilde{\lambda} > 0$. Then the sample size n required to achieve a proportionate variance reduction of κ for $\tilde{Y}$, $Var_n(\tilde{Y}) \leq (1-\kappa)Var(\tilde{Y})$, is*

$$n \geq \frac{\kappa(1-\tilde{\lambda})}{\tilde{\lambda}(1-\kappa) + (1/m)(1-\tilde{\lambda})}. \tag{4.59}$$

If the minimal eigenvalue of $\tilde{T}_1$ is $\tilde{\lambda}_{min}$, then to achieve a proportionate variance reduction of κ for every element of $[\tilde{\mathcal{M}}(\mathcal{C})]$ requires a sample size, rounded up, of

$$\frac{\kappa(1-\tilde{\lambda}_{min})}{\tilde{\lambda}_{min}(1-\kappa) + (1/m)(1-\tilde{\lambda}_{min})}. \tag{4.60}$$

We observed this corollary in action in Table 4.3

4.6 Extendible and non-extendible exchangeable sequences

Recall from the introduction that the use of finite exchangeability is desirable within the subjective framework as it acknowledges the necessarily finite nature of our actual exchangeability judgements. It is also more general than infinite exchangeability as it covers two possible scenarios. Firstly, the collection could form part of a larger (possibly infinite) sequence of exchangeable collections, or the sequence cannot be embedded in any longer sequence of exchangeable collections. There may be a number of reasons for this. Firstly, the situation may just not be extendible, there are no more cases to cover. Secondly, we may be able to embed our sequence into a longer sequence, but the maximal length may be restricted by the specifications having to remain consistent. We make the following definition

Definition 19 *Suppose that we have a sequence of m second-order exchangeable collections, $\mathcal{C}_1, \ldots, \mathcal{C}_m$. We say the sequence is M extendible if there is a sequence of $m + M$ second-order exchangeable collections, $\mathcal{C}_1, \ldots, \mathcal{C}_m, \mathcal{C}_{m+1}, \ldots, \mathcal{C}_{m+M}$.*

Thus, our consideration for each case and our considerations between each pair of cases is the same in the two sequences. From this definition, we can view infinite exchangeability as corresponding to the assumption of M-extendibility for all $M > m$. A second-order exchangeable sequence is, theoretically, extendible provided we have not introduced an invalid specification over quantities in $\mathcal{C}^*$. From the coherency conditions given in Statement 1, we can immediately impose upper bounds on the length of exchangeable sequence our judgements between two cases could conceivably allow.

Statement 2 *If $\sum_u \sum_{u'} \eta_u \eta_{u'} c_{v_u v_{u'}} \geq 0$, where $\{v_1, v_2, \ldots\}$ is a general finite subset of integers, and $\eta = [\eta_1 \eta_2 \ldots]^T$ is a general real valued vector, then the collection $\mathcal{C}$ is M − extendible for all M. If $Cov(X_{vi}, X_{vj}) = c_{vv} < 0$ for some v, $i \neq j$ then the collection $\mathcal{C}$ is an exchangeable sequence of length at most $1 + Int[|\rho_{min}|^{-1}]$ where*

$$\rho_{min} = \min_{v, i \neq j} Corr(X_{vi}, X_{vj}). \tag{4.61}$$

Proof - From the three conditions in Statement 1, we see that equation (4.3) is the only condition that depends upon the length of the sequence. Thus, we have M extendibility provided equation (4.3) holds. From the proof of Statement 1, we see this holds for all values of M if $\sum_u \sum_{u'} \eta_u \eta_{u'} c_{v_u v_{u'}} \geq 0$. Noting that $Corr(X_{vi}, X_{vj}) = (c_{vv}/d_{vv})$, then if $c_{vv} < 0$, by considering $Var(\mathcal{X}) =$

$Var(\sum_{i=1}^{m+M} X_{vi}) \geq 0$, where $m + M$ is the total length of the second-order exchangeable sequence, we must have

$$d_{vv} + (m + M - 1)c_{vv} \geq 0 \;\Leftrightarrow\; (m + M) \leq 1 - \frac{d_{vv}}{c_{vv}}. \tag{4.62}$$

The largest value of $m + M$ is then $1 + Int[|Corr(X_{vi}, X_{vj})|^{-1}]$ and by considering the cases where $c_{vv} < 0$ the result follows. $\square$

We are interested in the relationships between the adjustments made in extendible exchangeable sequences. From a design perspective, this is useful because we may have costs involved in actually determining the length of the sequence and would like to be able to assess the impact of changing the sequence length on the design, both qualitatively and quantitatively. Suppose that $\mathcal{C}^*$ is the union of m collections and $\mathcal{C}$ is second-order exchangeable over $\mathcal{C}^*$. Then we may make the length of the sequence explicit when we use the representation theorem by denoting by $\tilde{\mathcal{M}}^{[m]}(\mathcal{C}) = \{\tilde{\mathcal{M}}^{[m]}(X_1), \tilde{\mathcal{M}}^{[m]}(X_2), \dots\}$, the population mean collection, and by $\tilde{\mathcal{R}}_i^{[m]}(\mathcal{C}) = \{\tilde{\mathcal{R}}_i^{[m]}(X_1), \tilde{\mathcal{R}}_i^{[m]}(X_2), \dots\}$, the residual collection for the ith individual, so that

$$X_{vi} = \tilde{\mathcal{M}}^{[m]}(X_v) + \tilde{\mathcal{R}}_i^{[m]}(X_v). \tag{4.63}$$

Recall that any length second-order exchangeable sequence is uniquely specified by the consideration of just two cases. We shall show that this is essentially all we need consider for the adjustment of any second-order exchangeable sequence of length m. Denote by $\tilde{T}_n^{[m]} = E_{\tilde{\mathcal{M}}^{[m]}}\{E_n(\cdot)\}$ the resolution transform for the population mean collection $\tilde{\mathcal{M}}^{[m]}(\mathcal{C})$, based on n observations; the following theorem reveals the link between the structures.

Theorem 28 *Suppose that $\tilde{Y}^{[2]} = \sum_u \xi_u \tilde{\mathcal{M}}^{[2]}(X_{i_u})$, where $\{i_1, i_2, \dots\}$ is a general finite subset of integers, is an eigenvector of $\tilde{T}_1^{[2]}$, with eigenvalue $\tilde{\lambda}^{[2]}$. Then $\tilde{Y}^{[m]} = \sum_u \xi_u \tilde{\mathcal{M}}^{[m]}(X_{i_u})$ is an eigenvector of $\tilde{T}_1^{[m]}$, with eigenvalue*

$$\tilde{\lambda}^{[m]} = \frac{2(m-1)\tilde{\lambda}^{[2]} + (2-m)}{m}. \tag{4.64}$$

Proof - From the proof to Theorem 27, we have that $\tilde{Y}^{[2]}$ is an eigenvector of $\tilde{T}_1^{[2]}$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1 - \tilde{\lambda}^{[2]})Cov(\tilde{Y}^{[2]}, \tilde{\mathcal{M}}^{[2]}(\mathcal{X})) = Cov(\tilde{\mathcal{R}}_1^{[2]}(U_{\tilde{Y}^{[2]}}), \tilde{\mathcal{R}}_1^{[2]}(\mathcal{X})), \tag{4.65}$$

where

$$\begin{aligned} E_1(\tilde{Y}^{[2]}) &= \tilde{\mathcal{M}}^{[2]}(U_{\tilde{Y}^{[2]}}) + \tilde{\mathcal{R}}_1^{[2]}(U_{\tilde{Y}^{[2]}}) && (4.66)\\ &= \tilde{\lambda}^{[2]}\tilde{Y}^{[2]} + \tilde{\mathcal{R}}_1^{[2]}(U_{\tilde{Y}^{[2]}}). && (4.67)\end{aligned}$$

Similarly, $\tilde{Y}^{[m]}$ is an eigenvector of $\tilde{T}_1^{[m]}$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$(1-\tilde{\lambda}^{[m]})Cov(\tilde{Y}^{[m]}, \tilde{\mathcal{M}}^{[m]}(\mathcal{X})) = Cov(\tilde{\mathcal{R}}_1^{[m]}(U_{\tilde{Y}^{[m]}}), \tilde{\mathcal{R}}_1^{[m]}(\mathcal{X})), \tag{4.68}$$

where

$$\begin{aligned} E_1(\tilde{Y}^{[m]}) &= \tilde{\mathcal{M}}^{[m]}(U_{\tilde{Y}^{[m]}}) + \tilde{\mathcal{R}}_1^{[m]}(U_{\tilde{Y}^{[m]}}) && (4.69)\\ &= \tilde{\lambda}^{[m]}\tilde{Y}^{[m]} + \tilde{\mathcal{R}}_1^{[m]}(U_{\tilde{Y}^{[m]}}). && (4.70)\end{aligned}$$

Now, from equation (4.21), we have

$$\begin{aligned} Cov(\tilde{\mathcal{R}}_i^{[m]}(X_v), \tilde{\mathcal{R}}_i^{[m]}(X_w)) &= \frac{m-1}{m}(d_{vw} - c_{vw}) && (4.71)\\ &= \frac{2(m-1)}{m} Cov(\tilde{\mathcal{R}}_i^{[2]}(X_v), \tilde{\mathcal{R}}_i^{[2]}(X_w)) && (4.72)\end{aligned}$$

so that

$$Cov(\tilde{\mathcal{R}}_1^{[m]}(U_{\tilde{Y}^{[m]}}), \tilde{\mathcal{R}}_1^{[m]}(\mathcal{X})) = \frac{2(m-1)}{m} Cov(\tilde{\mathcal{R}}_1^{[2]}(U_{\tilde{Y}^{[2]}}), \tilde{\mathcal{R}}_1^{[2]}(\mathcal{X})). \tag{4.73}$$

Similarly, for equation (4.19), we have

$$\begin{aligned} Cov(\tilde{\mathcal{M}}^{[m]}(X_v), \tilde{\mathcal{M}}^{[m]}(X_w)) &= c_{vw} + \frac{1}{m}(d_{vw} - c_{vw}) && (4.74)\\ = Cov(\tilde{\mathcal{M}}^{[2]}(X_v), \tilde{\mathcal{M}}^{[2]}(X_w)) &+ \frac{2-m}{2m}(d_{vw} - c_{vw}) && (4.75)\\ = Cov(\tilde{\mathcal{M}}^{[2]}(X_v), \tilde{\mathcal{M}}^{[2]}(X_w)) &+ \frac{2-m}{m} Cov(\tilde{\mathcal{R}}_i^{[2]}(X_v), \tilde{\mathcal{R}}_i^{[2]}(X_w)). && (4.76)\end{aligned}$$

From equation (4.70) we have that

$$Cov(\tilde{Y}^{[m]}, \tilde{\mathcal{M}}^{[m]}(\mathcal{X})) = \frac{1}{\tilde{\lambda}^{[m]}} Cov(\tilde{\mathcal{M}}^{[m]}(U_{\tilde{Y}^{[m]}}), \tilde{\mathcal{M}}^{[m]}(\mathcal{X})), \tag{4.77}$$

so that, by equation (4.76), we have

$$\begin{aligned} &Cov(\tilde{Y}^{[m]}, \tilde{\mathcal{M}}^{[m]}(\mathcal{X})) = \\ &\quad Cov(\tilde{Y}^{[2]}, \tilde{\mathcal{M}}^{[2]}(\mathcal{X})) + \frac{2-m}{m\tilde{\lambda}^{[m]}} Cov(\tilde{\mathcal{R}}_1^{[2]}(U_{\tilde{Y}^{[2]}}), \tilde{\mathcal{R}}_1^{[2]}(\mathcal{X})). && (4.78)\end{aligned}$$

By substituting equations (4.73) and (4.78) into equation (4.68), we have that $\tilde{Y}^{[m]}$ is an eigenvector of $\tilde{T}_1^{[m]}$ if and only if, for all $\mathcal{X} \in [\mathcal{C}]$

$$\begin{aligned} &(1-\tilde{\lambda}^{[m]})Cov(\tilde{Y}^{[2]}, \tilde{\mathcal{M}}^{[2]}(\mathcal{X})) = \\ &\quad \left(\frac{2(m-1)}{m} - \frac{(2-m)(1-\tilde{\lambda}^{[m]})}{m\tilde{\lambda}^{[m]}}\right) Cov(\tilde{\mathcal{R}}_1^{[2]}(U_{\tilde{Y}^{[2]}}), \tilde{\mathcal{R}}_1^{[2]}(\mathcal{X})) && (4.79)\\ &= \frac{m\tilde{\lambda}^{[m]} - (2-m)}{m\tilde{\lambda}^{[m]}} Cov(\tilde{\mathcal{R}}_1^{[2]}(U_{\tilde{Y}^{[2]}}), \tilde{\mathcal{R}}_1^{[2]}(\mathcal{X})). && (4.80)\end{aligned}$$

We thus have that $U_{\tilde{Y}^{[2]}}$ satisfies equation (4.65), with eigenvalue $\tilde{\lambda}^{[2]}$, if and only if $U_{\tilde{Y}^{[m]}} = \alpha U_{\tilde{Y}^{[2]}}$ satisfies equation (4.80), with eigenvalue $\tilde{\lambda}^{[m]}$ where

$$\frac{m\tilde{\lambda}^{[m]} - (2-m)}{m\tilde{\lambda}^{[m]}(1-\tilde{\lambda}^{[m]})}\alpha \quad = \quad \frac{1}{1-\tilde{\lambda}^{[2]}}. \tag{4.81}$$

Using equations, (4.70) and (4.67) we have that

$$\alpha \quad = \quad \frac{\tilde{\lambda}^{[m]}}{\tilde{\lambda}^{[2]}}. \tag{4.82}$$

Substituting equation (4.82) into equation (4.81) and rearranging gives

$$\tilde{\lambda}^{[m]} \quad = \quad \frac{2(m-1)\tilde{\lambda}^{[2]} + (2-m)}{m}, \tag{4.83}$$

and hence the result. □

The eigenvectors of $\tilde{T}_1^{[2]}$ and $\tilde{T}_1^{[m]}$ thus share the same co-ordinate representation, with easily modified eigenvalues. Recall that we observed this feature in the examiner example. We noted that the quantities $\tilde{\mathcal{M}}(X_+)$ and $\tilde{\mathcal{M}}(X_-)$ for the adjustment were proportional to the canonical directions and that this feature did not depend upon m.

We may combine the results of Theorem 27 and Theorem 28 together to yield the following corollary.

Corollary 14 *If $\tilde{Y}^{[2]}$ is an eigenvector of $\tilde{T}_1^{[2]}$, with eigenvalue $\tilde{\lambda}^{[2]}$, then $\tilde{Y}^{[m]}$ is, for each $n \leq m$, an eigenvector of $\tilde{T}_n^{[m]}$ with eigenvalue*

$$\tilde{\lambda}^{[m]}_{(n)} \quad = \quad \frac{2n(m-1)^2\tilde{\lambda}^{[2]} - n(m-1)(m-2)}{2(n-1)m(m-1)\tilde{\lambda}^{[2]} - (n-2)m(m-1)}. \tag{4.84}$$

Hence, the canonical directions for the adjustment of $[\tilde{\mathcal{M}}^{[m]}(\mathcal{C})]$ have, up to a scale factor to ensure a prior variance of one, the same co-ordinate representation for all m, with simply modified canonical resolutions. Thus, the qualitative information provided by the adjustment remains the same for all possible sequence lengths and all possible sample sizes and the quantitative information is easy to compare across this, via equation (4.84). Thus, not only is it straightforward to compare the effect of changing the sample size for learning about $[\tilde{\mathcal{M}}^{[m]}(\mathcal{C})]$, but we now see that it is straightforward to compare the differences between learning about the corresponding quantities in differing $[\tilde{\mathcal{M}}^{[m]}(\mathcal{C})]$s. For example, if m_1 and m_2 are feasible sequence lengths we have that

$$\frac{m_1}{m_1-1}(1-\tilde{\lambda}^{[m_1]}) \quad = \quad \frac{m_2}{m_2-1}(1-\tilde{\lambda}^{[m_2]}). \tag{4.85}$$

Since the canonical resolutions share the same co-ordinate representation for all feasible choices of m, then by using equations (4.76) and (4.72) it is straightforward to compare $R_n(\tilde{\mathcal{M}}^{[m_1]}(\mathcal{X}))$ with $R_n(\tilde{\mathcal{M}}^{[m_2]}(\mathcal{X}))$ for any $\mathcal{X} \in \langle \mathcal{C} \rangle$

Note also the computational advantage. In order to compare all sample sizes for all sequence lengths, we need only to consider the transform for a sample of size one and a sequence of length two.

It should be emphasised that the work in this section remains valid for the two types of second-order exchangeable sequences of length two. The first when we have at most $m-2$ extendibility (for whatever reason) and the second when we have infinite extendibility. We proceed by assuming we have infinite extendibility and showing the links between infinite second-order exchangeable sequences and finite second-order exchangeable sequences.

4.7 Linking the adjustments of finite and infinite second-order exchangeable sequences

Suppose that we have a sequence of two second-order exchangeable collections $\mathcal{C}_1$ and $\mathcal{C}_2$ and that the sequence is M extendible for all M, so that there is an infinite sequence of collections $\mathcal{C}_1, \mathcal{C}_2, \ldots$ which is second-order exchangeable. It is straightforward to see that for this to be theoretically possible, equation (4.3) of Statement 1 must reduce to $\sum_u \sum_{u'} \eta_u \eta_{u'} c_{v_u v_{u'}} \geq 0$, as was suggested by Statement 2. We have the following theorem.

Theorem 29 *Suppose that $\tilde{Y}^{[2]} = \sum_u \xi_u \tilde{\mathcal{M}}^{[2]}(X_{i_u})$, where $\{i_1, i_2, \ldots\}$ is a general finite subset of integers, is an eigenvector of $\tilde{T}_1^{[2]}$, with eigenvalue $\tilde{\lambda}^{[2]}$. Then $Y = \sum_u \xi_u \mathcal{M}(X_{i_u})$ is an eigenvector of T_1, with eigenvalue*

$$\lambda = 2\tilde{\lambda}^{[2]} - 1. \tag{4.86}$$

Proof - By noting that,

$$\begin{aligned} Cov(\mathcal{M}(X_v), \mathcal{M}(X_w)) &= Cov(\tilde{\mathcal{M}}^{[2]}(X_v), \tilde{\mathcal{M}}^{[2]}(X_w)) - \\ &\qquad Cov(\tilde{\mathcal{R}}_i^{[2]}(X_v), \tilde{\mathcal{R}}_i^{[2]}(X_w)); \qquad (4.87) \\ Cov(\mathcal{R}_i(X_v), \mathcal{R}_i(X_w)) &= 2Cov(\tilde{\mathcal{R}}_i^{[2]}(X_v), \tilde{\mathcal{R}}_i^{[2]}(X_w)), \qquad (4.88) \end{aligned}$$

the result follows in a similar way to Theorem 28. □

By combining the results of Theorem 3 of Goldstein & Wooff (1998) (reproduced as Theorem 5 in this thesis) and Theorem 29 we have the following corollary.

Corollary 15 *If $\tilde{Y}^{[2]}$ is an eigenvector of $\tilde{T}_1^{[2]}$, with eigenvalue $\tilde{\lambda}^{[2]}$, then Y is, for each n, an eigenvector of T_n with eigenvalue*

$$\lambda_{(n)} = \frac{2n\tilde{\lambda}^{[2]} - n}{2(n-1)\tilde{\lambda}^{[2]} - (n-2)}. \tag{4.89}$$

Notice that by comparing equation (4.89) with equation (4.84) we see that

$$\lambda_{(n)} = \lim_{m\to\infty} \tilde{\lambda}_{(n)}^{[m]}. \tag{4.90}$$

Indeed, by comparing equation (4.87) with equation (4.76) and equation (4.88) with equation (4.72), we can easily see that for any $\mathcal{X} \in \langle \mathcal{C} \rangle$ we have

$$R_n(\mathcal{M}(\mathcal{X})) = \lim_{m\to\infty} R_n(\tilde{\mathcal{M}}^{[m]}(\mathcal{X})). \tag{4.91}$$

Thus, we may think of the adjustment of the mean space corresponding to an infinite exchangeable sequence by a sample of n individuals from that sequence as being the limit of the adjustment of a finite sequence by a sample of n individuals from the finite sequence.

Notice that we have expressed all our canonical resolutions in terms of $\tilde{\lambda}^{[2]}$. This allows us to include the theory for both infinitely extendible and finitely extendible sequences together since a second-order exchangeable sequence of length m which is not extendible may be conceived as a second-order exchangeable sequence of length 2 being $m-2$ extendible. In the case of infinite extendibility, we may choose to express the resolutions in terms of the resolutions for the infinite sequence. Using equations (4.84) and (4.89), we may write, for any infinitely extendible second-order exchangeable sequence of length m,

$$\tilde{\lambda}_{(n)}^{[m]} = \frac{\lambda_{(n)}}{\lambda_{(m)}}. \tag{4.92}$$

In particular, we have that

$$\tilde{\lambda}^{[m]} = \frac{(m-1)\lambda + 1}{m}. \tag{4.93}$$

Notice that equation (4.93) shows that the relationships displayed in Table 4.1 between the resolutions for $\tilde{\mathcal{M}}(X_+)$ and $\mathcal{M}(X_+)$ and between the resolutions for $\tilde{\mathcal{M}}(X_-)$ and $\mathcal{M}(X_-)$ are not quirks obtained by the simplicity of the examiner example but a consequence of the second-order exchangeability judgements. We may rearrange equation (4.92) into the form

$$\tilde{\lambda}_{(n)}^{[m]} = \lambda_{(n)} + \frac{n}{m}(1 - \lambda_{(n)}). \tag{4.94}$$

Thus, we could consider the qualitative difference between infinite and finite modelling to be that for the finite case, we need a finite model correction term for the canonical resolutions, this correction term is $\frac{n}{m}(1-\lambda_{(n)})$. If $\tilde{Y}_s^{[m]}$ is an eigenvector of $\tilde{T}_n^{[m]}$ with eigenvalue $\tilde{\lambda}_{(n)s}^{[m]}$ and Y_s, $\lambda_{(n)s}$ the corresponding eigenvector and eigenvalue for T_n, then from equation (4.94) we have that

$$Var_n(\tilde{Y}_s^{[m]}) \;=\; (1-(n/m))Var_n(Y). \tag{4.95}$$

The multiplier $(1-(n/m))$ is the same for each eigenvector of $\tilde{T}_n^{[m]}$. We could use the fraction (n/m) as a 'rule of thumb' for assessing the validity of the infinite approximation to the finite judgement; the smaller the value of (n/m), the greater the validity of the approximation. The benefit of this guide is that it does not depend upon any particular direction in $[\tilde{\mathcal{M}}^{[m]}(\mathcal{C})]$.

Of course, if we are specifically interested in given directions in $[\tilde{\mathcal{M}}^{[m]}(\mathcal{C})]$, which may not correspond to the canonical resolutions, then we may like a less ad hoc guide to the validity of the approximation. However, as the eigenvectors of $\tilde{T}_n^{[m]}$ and T_n share the same co-ordinate representation then it is straightforward to assess the difference between the infinite approximation and the finite reality. Suppose that $Z_s = Y_s = \sum_u \xi_u \mathcal{M}(X_{i_u})$ is a canonical direction of T_n, with canonical resolution $\lambda_{(n)s}$. Then, by Theorem 29 and Theorem 28, $\tilde{Z}_s^{[m]} = a_s\tilde{Y}_s^{[m]}$ is a canonical direction of $\tilde{T}_n^{[m]}$ with canonical resolution $\tilde{\lambda}_{(n)s}^{[m]}$. a_s is chosen to ensure that $\tilde{Z}_s^{[m]}$ has prior variance 1, so that $a_s^{-2} = Var(\tilde{Y}_s^{[m]})$. Now $Z_s \in [\mathcal{M}(\mathcal{C})]$ so suppose $Z_s = \mathcal{M}(\mathcal{Z}_s)$ for some $\mathcal{Z}_s \in \langle\mathcal{C}\rangle$. Hence, $\tilde{Z}_s^{[m]} = \tilde{\mathcal{M}}^{[m]}(a_s\mathcal{Z}_s)$. Notice from equations (4.19), (1.22) and (1.24) we may write for any $\mathcal{X} \in \langle\mathcal{C}\rangle$,

$$\begin{aligned} Cov(\tilde{\mathcal{M}}^{[m]}(\mathcal{Z}_s), \tilde{\mathcal{M}}^{[m]}(\mathcal{X})) &= \\ Cov(\mathcal{M}(\mathcal{Z}_s), \mathcal{M}(\mathcal{X})) &+ \frac{1}{m}Cov(\mathcal{R}_i(\mathcal{Z}_s), \mathcal{R}_i(\mathcal{X})). \end{aligned} \tag{4.96}$$

Hence,

$$a_s \;=\; \{1+(1/m)Var(\mathcal{R}_i(\mathcal{Z}_s))\}^{-\frac{1}{2}}. \tag{4.97}$$

Then, by using equation (1.62), we may write for any $\mathcal{X} \in \langle\mathcal{C}\rangle$,

$$\begin{aligned} Var_n(\tilde{\mathcal{M}}^{[m]}(\mathcal{X})) \;=\; \left(1-\frac{n}{m}\right)\sum_s (1-\lambda_{(n)s})a_s^2\{Cov(\mathcal{M}(\mathcal{Z}_s), \mathcal{M}(\mathcal{X})) + \\ (1/m)Cov(\mathcal{R}_i(\mathcal{Z}_s), \mathcal{R}_i(\mathcal{X}))\}^2. \end{aligned} \tag{4.98}$$

Equation (4.98) thus expresses the adjusted variance for any $\tilde{\mathcal{M}}^{[m]}(\mathcal{X}) \in [\tilde{\mathcal{M}}^{[m]}(\mathcal{C})]$, having observed a sample of size n, for any infinitely extendible exchangeable sequence in terms of relationships and adjustments of quantities purely in the infinite sequence. By noting that,

$$Var_n(\mathcal{M}(\mathcal{X})) = \sum_s (1 - \lambda_{(n)s})\{Cov(\mathcal{M}(\mathcal{Z}_s), \mathcal{M}(\mathcal{X}))\}^2, \tag{4.99}$$

it is straightforward to compare the effect of the infinite approximation for any sample size n and any sequence length m. Notice the rather simple dependence upon $(1/m)$. By letting $m \to \infty$, it is easy to confirm equation (4.91). Thus, not only do we know that qualitatively $\tilde{T}_n^{[m]}$ and T_n provide the same information, but also that the quantitative differences are straightforward to calculate via equations (4.98) and (4.99), providing an easy way to assess the differences between the more realistic modelling framework of finite second-order exchangeability, where the difficulty may lie in determining m, and the convenient use of infinite second-order exchangeability.

4.8 Prediction of future individuals

Suppose that we wish to consider the effect of observing n individuals for predicting the values for a further r individuals who are second-order exchangeable with those in the sample. Then, this adjustment is driven completely by the relationships between the individuals and as such, provided that there are the respective n and r individuals available, the adjustment does not depend upon the total number of individuals who could have been sampled. Thus, if there are only $m \geq n+r$ second-order exchangeable individuals, or an infinite number of second-order exchangeable individuals the adjustment will be the same, and is given by that in Section 8 of Goldstein & Wooff (1998). Since the canonical directions of the predictive adjustment share, up to a scale factor, the same co-ordinate representation as for the adjustment of the underlying mean components, then the same qualitative insights may be made if the sequence in fact only contains a finite number of individuals.

4.9 Implementing the theory; linking with the underlying canonical variable problem

Consider that for each individual we shall make a total of v_0 measurements. Thus, we may write C_i as the $v_0 \times 1$ vector, $C_i = [X_{1i} \dots X_{v_0 i}]^T$. As per usual, we collect the $\{c_{vw}\}$ into the $v_0 \times v_0$ matrix C with (v, w)th entry $(C)_{vw} = c_{vw}$ and the $\{e_{vw} = d_{vw} - c_{vw}\}$ into the $v_0 \times v_0$ matrix E with (v, w)th entry $(E)_{vw} = c_{vw}$.

We may then write

$$Var(\tilde{\mathcal{M}}^{[m]}(\mathcal{C})) = C + \frac{1}{m}E; \tag{4.100}$$

$$Cov(\tilde{\mathcal{R}}_i^{[m]}(\mathcal{C}), \tilde{\mathcal{R}}_j^{[m]}(\mathcal{C})) = \begin{cases} \frac{m-1}{m}E & \text{if } i = j; \\ -\frac{1}{m}E & \text{otherwise.} \end{cases} \tag{4.101}$$

We shall assume that the variance matrices under consideration are strictly positive definite and hence invertible. In the case where they are not, we would obtain the corresponding results by considering the adjustment over the linear span of the columns of the matrices that we need to invert. We have the following corollary which summarises our work in the previous two sections.

Corollary 16 *The resolution transform matrix, $\tilde{T}_n^{[m]}$, is calculated as*

$$\tilde{T}_n^{[m]} = \{C + (1/n)E\}^{-1}\{C + (1/m)E\}. \tag{4.102}$$

For each $s = 1, \dots, v_0$, the canonical directions for the adjustment of $[\tilde{\mathcal{M}}^{[m]}(\mathcal{C})]$ by $\mathcal{C}(n)$ are given by

$$\tilde{Z}_s^{[m]} = \lambda_{(m)s}^{\frac{1}{2}} Y_s^T \tilde{\mathcal{M}}^{[m]}(\mathcal{C}), \tag{4.103}$$

with the corresponding canonical resolutions given by $\tilde{\lambda}_{(n)s}^{[m]}$. Y_s is the sth underlying canonical variable direction as given by Definition 12.

Proof - We need only to verify that the canonical resolutions have a prior variance of one. We have that

$$Cov(\tilde{Z}_s^{[m]}, \tilde{Z}_t^{[m]}) = \lambda_{(m)s}^{\frac{1}{2}} \lambda_{(m)t}^{\frac{1}{2}} Y_s^T Var(\tilde{\mathcal{M}}^{[m]}(\mathcal{C})) Y_t \tag{4.104}$$

$$= \lambda_{(m)s}^{\frac{1}{2}} \lambda_{(m)t}^{\frac{1}{2}} Y_s^T (C + (1/m)E) Y_t \tag{4.105}$$

$$= \lambda_{(m)s}^{\frac{1}{2}} \lambda_{(m)t}^{\frac{1}{2}} \{[(m-1)/m](Y^TCY)_{st} + [1/(m\lambda_t)](Y^T(C+E)Y\Lambda)_{st}\} \tag{4.106}$$

$$= \frac{(m-1)\lambda_t + 1}{m\lambda_t} \times \delta_{st} \lambda_{(m)s}^{\frac{1}{2}} \lambda_{(m)t}^{\frac{1}{2}} \tag{4.107}$$

$$= \delta_{st}. \tag{4.108}$$

The result thus follows. □

4.10 Co-exchangeable finitely exchangeable systems

We introduced co-exchangeability in order to achieve a more realistic belief model. The problem at hand is one where we are interested in making a series of measurements $\mathcal{C} = \{X_1, \dots, X_{v_0}\}$ on a collection of individuals, where each individual can be classified as coming from one of g_0 groups. For each individual, we wish to measure the same set of v_0 variables. Let $\mathcal{C}_{gi} = \{X_{g1i}, \ldots , X_{gv_0i}\}$ be the values of the measurements for the ith individual in the gth group. In Chapter 2 we proceeded as if there were, in each group, a potentially infinite amount of individuals that could be observed. However, as we argued at the start of this Chapter, this is not really the case and it is usually possible to give an upper bound to the number of individuals that are available in each group. We shall proceed in this vein, assuming that in the gth group, there are m_g possible individuals. We collect the measurements for all the individuals in the gth group together as the collection $\mathcal{C}_g^* = \cup_{i=1}^{m_g}\mathcal{C}_{gi}$, and the total collection of measurements, namely those for all the individuals in all the groups, are collected together in the collection $\mathcal{C}^* = \cup_{g=1}^{g_0}\mathcal{C}_g^*$. In an analogous way to Chapter 2, we consider that each $\mathcal{C}_g^*$ is second-order exchangeable over the individuals and $\mathcal{C}_1^*$, $\mathcal{C}_2^*$, ..., $\mathcal{C}_{g_0}^*$ are pairwise co-exchangeable across the individuals. That is, our specifications take the form of equations (2.3) - (2.6). Thus, we change not the relationships between the individuals, but just the judgement as to how many possible individuals there are. In an analogous manner to Theorem 7, the representation theorem for finite exchangeable sequences may then be adapted to incorporate finite co-exchangeability as the following theorem. The representation follows immediately from Goldstein (1986a).

Theorem 30 *Suppose that $\mathcal{C}_1^*$, $\mathcal{C}_2^*$, $\mathcal{C}_3^*$, ..., are a sequence of co-exchangeable finite exchangeable systems. Then for each system g we introduce the further collections of random quantities $\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g) = \{\tilde{\mathcal{M}}^{[m_g]}(X_{g1}), \tilde{\mathcal{M}}^{[m_g]}(X_{g2}), \ldots\}$, and $\tilde{\mathcal{R}}_i^{[m_g]}(\mathcal{C}_g) = \{\tilde{\mathcal{R}}_i^{[m_g]}(X_{g1}), \tilde{\mathcal{R}}_i^{[m_g]}(X_{g2}), \ldots\}$, and $\forall g, v, i$ write*

$$X_{gvi} = \tilde{\mathcal{M}}^{[m_g]}(X_{gv}) + \tilde{\mathcal{R}}_i^{[m_g]}(X_{gv}), \tag{4.109}$$

where

$$\tilde{\mathcal{M}}^{[m_g]}(X_{gv}) = \frac{1}{m_g}\sum_{i=1}^{m_g} X_{gvi}. \tag{4.110}$$

The collections $\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)$ and $\tilde{\mathcal{R}}_i^{[m_g]}(\mathcal{C}_g)$ satisfy, for all v, w, the following relation-

ships

$$\begin{aligned}
E(\tilde{\mathcal{M}}^{[m_g]}(X_{gv})) &= m_{gv}\ \forall g; & (4.111)\\
E(\tilde{\mathcal{R}}_i^{[m_g]}(X_{gv})) &= 0\ \forall g, i; & (4.112)\\
Cov(\tilde{\mathcal{M}}^{[m_g]}(X_{gv}), \tilde{\mathcal{M}}^{[m_h]}(X_{hw})) &= \begin{cases} c_{ggvw} + \frac{1}{m_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h; \\ c_{ghvw} & \text{if } g \neq h; \end{cases} & (4.113)\\
Cov(\tilde{\mathcal{M}}^{[m_g]}(X_{gv}), \tilde{\mathcal{R}}_j^{[m_h]}(X_{hw})) &= 0\ \forall g, h, j; & (4.114)\\
Cov(\tilde{\mathcal{R}}_i^{[m_g]}(X_{gv}), \tilde{\mathcal{R}}_j^{[m_h]}(X_{hw})) &= \begin{cases} \frac{m_g - 1}{m_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h,\ i = j; \\ -\frac{1}{m_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h,\ i \neq j \\ 0 & \text{otherwise.} \end{cases} & (4.115)
\end{aligned}$$

Notice that the residual vectors are uncorrelated across systems and that it is only the within system, different individuals residual vectors that have a correlation induced by the finite population judgement. The residual vectors are uncorrelated with the mean component vectors. Thus, if $\tilde{\mathcal{M}}(\mathcal{C}) = \cup_g \tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)$ represents the complete collection of mean components and $\tilde{\mathcal{R}}(\mathcal{C}) = \cup_g \cup_i \tilde{\mathcal{R}}_i^{[m_g]}(\mathcal{C}_g)$ the complete collection of residual components, then from equation (4.114) we have that

$$[\tilde{\mathcal{M}}(\mathcal{C})] \perp [\tilde{\mathcal{R}}(\mathcal{C})]. \tag{4.116}$$

4.10.1 Adjusting the mean components

Suppose that we make observations in g_0 systems, observing $0 < n_g \leq m_g$ individuals in the gth system, for $g = 1, \dots, g_0$. Thus, the sample in the gth system is the collection $\mathcal{C}_g(n_g)$ and the total sample is the collection $\mathcal{C}(N)$. $\mathcal{S}_{n_g}(\mathcal{C}_g)$ denotes the collection of sample means in the gth system, with the complete collection of sample means being $\mathcal{S}_N(\mathcal{C})$. $\mathcal{T}_i(\mathcal{C}_g(n_g))$ denotes the collection of sample mean residuals of the ith individual in the gth system with the complete collection of sample mean residuals given by $\mathcal{T}(\mathcal{C}(N))$.

Lemma 22 *The second-order relationships between the $\tilde{\mathcal{M}}^{[m_g]}(X_{gv})$s, the $\mathcal{S}_{n_g}(X_{gv})$s* *and the $\mathcal{T}_i(X_{gv})$s may be expressed, for all v, w, as*

$$Cov(\tilde{\mathcal{M}}^{[m_g]}(X_{gv}), \mathcal{S}_{n_h}(X_{hw})) = \begin{cases} c_{ggvw} + \frac{1}{m_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h; \\ c_{ghvw} & \text{otherwise}; \end{cases} \tag{4.117}$$

$$Cov(\mathcal{S}_{n_g}(X_{gv}), \mathcal{S}_{n_h}(X_{hw})) = \begin{cases} c_{ggvw} + \frac{1}{n_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h; \\ c_{ghvw} & \text{otherwise}; \end{cases} \tag{4.118}$$

$$Cov(\mathcal{S}_{n_g}(X_{gv}), \mathcal{T}_j(X_{hw})) = 0 \ \forall g, h, j; \tag{4.119}$$

$$Cov(\tilde{\mathcal{M}}^{[m_g]}(X_{gv}), \mathcal{T}_j(X_{hw})) = 0 \ \forall g, h, j; \tag{4.120}$$

$$Cov(\mathcal{T}_i(X_{gv}), \mathcal{T}_j(X_{hw})) = \begin{cases} \frac{n_g - 1}{n_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h,\ i = j; \\ -\frac{1}{n_g}(d_{gvw} - c_{ggvw}) & \text{if } g = h,\ i \neq j; \\ 0 & \text{otherwise}. \end{cases} \tag{4.121}$$

Proof - Since $\mathcal{S}_{n_g}(X_{gv})$ and $\mathcal{T}_i(X_{gv})$ only depend upon n_g individuals, then there is no difference as to whether they are sampled from a finite or infinite collection and equations (4.118), (4.119) and (4.121) are identical to the respective equations (2.18), (2.19) and (2.21). Equations (2.17) and (4.120) follow from equations (4.113) - (4.115). □

The resolution transform for the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C})]$ by $\mathcal{S}_N(\mathcal{C})$ is denoted by $T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$ and the resolution transform for the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C})]$ by $\mathcal{C}(N)$ is denoted by $T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{C}(N)]}$. We now show that the collection of sample means is Bayes linear sufficient for the full sample for adjusting the population mean collections.

Theorem 31 *If $\mathcal{C}_1^*, \dots, \mathcal{C}_{g_0}^*$ are a chain of co-exchangeable finite exchangeable systems then $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting the collection $\tilde{\mathcal{M}}(\mathcal{C})$. Equivalently, for each $\tilde{\mathcal{M}}(\mathcal{X}) \in [\tilde{\mathcal{M}}(\mathcal{C})]$ we have*

$$E_{\mathcal{S}_N(\mathcal{C})}(\tilde{\mathcal{M}}(\mathcal{X})) = E_N(\mathcal{X})); \tag{4.122}$$

$$Var_{\mathcal{S}_N(\mathcal{C})}(\tilde{\mathcal{M}}(\mathcal{X})) = Var_N(\mathcal{X})); \tag{4.123}$$

$$T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}(\tilde{\mathcal{M}}(\mathcal{X})) = T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{C}(N)]}. \tag{4.124}$$

Proof - Note from equation (4.119) we have that $[\mathcal{S}_N(\mathcal{C})] \perp [\mathcal{T}(\mathcal{C}(N))]$ and from equation (4.120) we have that $[\tilde{\mathcal{M}}(\mathcal{C})] \perp [\mathcal{T}(\mathcal{C}(N))]$ and the results follow. □

Thus, if we want to adjust the population mean collection corresponding to a sequence of co-exchangeable finite exchangeable systems by observing individuals in each of the systems then we need only the resulting sample means to perform the adjustment. Notice that this mirrors precisely the case when the systems were judged infinitely exchangeable, see Theorem 8.

4.11 Grouped multivariate exchangeable systems: the finite analogue

We now develop the analogous results to those for the adjustment of the mean collection for grouped multivariate exchangeable systems we considered in Chapter 2 where, at least in principle, there were an infinite number of individuals available in each group. We judge that the relationships between each pair of individuals is the same as when we considered the individuals as being from an infinite population. To each individual we may apply the representation theorem to write, for each g, v, i,

$$X_{gvi} = \tilde{\mathcal{M}}^{[m_g]}(X_{gv}) + \tilde{\mathcal{R}}_i^{[m_g]}(X_{gv}). \tag{4.125}$$

The induced specifications for the $\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)$, $\tilde{\mathcal{R}}_i^{[m_g]}(\mathcal{C}_g)$ are as

$$\begin{aligned}
Cov(\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g), \tilde{\mathcal{M}}^{[m_h]}(\mathcal{C}_h)) &= \begin{cases} \alpha_{gg}C + \frac{1}{m_g}\beta_g E & \text{if } g = h; \\ \alpha_{gh}C & \text{otherwise;} \end{cases} && (4.126) \\
Cov(\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g), \tilde{\mathcal{R}}_j^{[m_h]}(\mathcal{C}_h)) &= 0 \; \forall g, h, j; && (4.127) \\
Cov(\tilde{\mathcal{R}}_i^{[m_g]}(\mathcal{C}_g), \tilde{\mathcal{R}}_j^{[m_h]}(\mathcal{C}_h)) &= \begin{cases} \frac{(m_g-1)}{m_g}\beta_g E & \text{if } g = h,\, i = j; \\ -\frac{1}{m_g}\beta_g E & \text{if } g = h,\, i \neq j; \\ 0 & \text{otherwise.} \end{cases} && (4.128)
\end{aligned}$$

Once more, we let A be the $g_0 \times g_0$ matrix with (g, h)th entry $(A)_{gh} = \alpha_{gh}$, and let B be the $g_0 \times g_0$ diagonal matrix with (g, g)th entry $(B)_{gg} = \beta_g$. We collect the population sizes together into the matrix $M = diag(m_1, \dots, m_{g_0})$.

We wish to take a sample of size $n_g > 0$ from the gth group and use the observations to adjust our beliefs over the population mean collection, $\tilde{\mathcal{M}}(\mathcal{C})$. We collect the sample sizes together into the matrix $N = diag(n_1, \dots, n_{g_0})$. Using our usual vector notation then we may combine Lemma 22 with equations (4.126) - (4.128) to obtain the following lemma.

Lemma 23 *The second-order specifications for the population mean collection, $\tilde{\mathcal{M}}(\mathcal{C})$, and the sample means, $\mathcal{S}_N(\mathcal{C})$, may be expressed as*

$$\begin{aligned}
Var(\tilde{\mathcal{M}}(\mathcal{C})) &= (A \otimes C) + (M^{-1}B \otimes E); && (4.129) \\
Cov(\tilde{\mathcal{M}}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) &= (A \otimes C) + (M^{-1}B \otimes E); && (4.130) \\
Var(\mathcal{S}_N(\mathcal{C})) &= (A \otimes C) + (N^{-1}B \otimes E). && (4.131)
\end{aligned}$$

We now consider separately the analysis of variables and groups as follows.

4.12 Underlying canonical variable problem

Recall the work at the start of this chapter for finite second-order exchangeable sequences. We showed that the adjustment of the population mean collection induced by a sample of size n drawn from the population had the same canonical resolutions for all choices of n, whilst for all choices of m, the length of the exchangeable sequence, the canonical resolutions had the same, up to a scale factor to ensure a prior variance of one, co-ordinate representation. In our current framework, we have a total of g_0 finite second-order exchangeable sequences. Suppose that we consider the adjustment in a single group from a sample, of size n, drawn exclusively from that group. We may rewrite Corollary 16 as the following corollary.

Corollary 17 *For the adjustment of $\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)$ by $\mathcal{C}_g(n_g)$, the resolution transform matrix, $T_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]}$, is calculated as*

$$T_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]} = \{\alpha_{gg}C + (1/n_g)\beta_g E\}^{-1}\{\alpha_{gg}C + (1/m_g)\beta_g E\}. \quad (4.132)$$

For each $s = 1, \cdots, v_0$, the canonical directions are given by

$$Z_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} = \sqrt{\frac{m_g\phi_s}{m_g\alpha_{gg}\phi_s + (1-\phi_s)\beta_g}}Y_s^T\tilde{\mathcal{M}}(X_g), \quad (4.133)$$

with the corresponding canonical resolutions given by

$$\lambda_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} = \frac{n_g}{m_g}\left\{\frac{m_g\alpha_{gg}\phi_s + \beta_g(1-\phi_s)}{n_g\alpha_{gg}\phi_s + \beta_g(1-\phi_s)}\right\} \quad (4.134)$$

$$= \frac{\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}}{\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(m_g)]s}}, \quad (4.135)$$

where Y_s is the sth underlying canonical variable direction and ϕ_s the sth underlying canonical variable resolution as given by Definition 12.

Corollary 17 is the finite analogue of Theorem 11. Notice that each $Z_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ has, up to a scale-factor the same co-ordinate representation as $Z_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ (as given by equation (2.97)) and that both these directions are found from the solution of the underlying canonical variable problem as given by Definition 12. As we would expect from observing Theorem 11, we have similarly strong coherence properties across the groups as we do for sample sizes and sequence lengths. Irrespective of the length of sequence in the group or the sample size observed, up to a scale-factor, the canonical resolutions have the same co-ordinate representation and thus the qualitative information of adjusting $[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)]$ by a sample of size n drawn purely from that group does not depend upon g (or m_g or n).

Notice also how both $\lambda_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ and $\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$ (see equation (2.123)) may be obtained from the solution of the underlying canonical variable problem. Recall from equation (4.94) that we may rewrite equation (4.135) as

$$\lambda_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} = \lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} + \frac{n_g}{m_g}(1 - \lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}). \quad (4.136)$$

We emphasise once more that the essential difference between the realistic judgement of an extendible finite second-order exchangeable sequence and the convenient assumption of an infinite second-order exchangeable sequence is not qualitative, but quantitative. We have a finite population correction, $(n_g/m_g)(1 - \lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s})$ for the sth canonical resolution which corresponds to the same direction in the finite case as it does in the infinite case. Notice that the size of this correction is dependent upon the proportion of the sequence we observe. Recall that from equation (4.95), we suggested that the proportion (n_g/m_g) could be used as a rule of thumb for assessing the validity of the infinite approximation. Here, we see that (n_g/m_g) will typically differ across the groups so that groups with small values of (n_g/m_g) may be perceived as being better approximated to an infinite exchangeable sequence than those with larger values of (n_g/m_g).

4.12.1 The examiner problem in this finite setting

Recall how in Chapter 2, the examiner attempted to improve his modelling in Chapter 1 by explicitly introducing the two markers into his model. Having reassessed the model in Chapter 1 in the first half of this chapter by introducing the finite exchangeable modelling, the examiner wishes to reintroduce the markers to this model. He assesses that the first marker has a potential m_1 individuals to mark and the second marker has a potential m_2 individuals to mark. For example, $m_1 + m_2$ could refer to the total number of individuals registered for the exam. The examiner specifies the same relationships between each pair of individuals as he did in the infinite case. In terms of the representation theorem he may write $X_{gvi} = \tilde{\mathcal{M}}^{[m_g]}(X_{gv}) + \tilde{\mathcal{R}}_i^{[m_g]}(X_{gv})$. The finite modelling version of equations (2.47) and (2.48) are

$$Cov(\tilde{\mathcal{M}}^{[m_g]}(X_{gv}), \tilde{\mathcal{M}}^{[m_h]}(X_{hw})) = \begin{cases} \gamma_{gh} + (\delta_{gh}/m_g)4 & \text{if } v = w, \forall g, h; \\ 0.25\gamma_{gh} + (\delta_{gh}/m_g)3.5 & \text{if } v \neq w, \forall g, h; \end{cases} \quad (4.137)$$

$$Cov(\tilde{\mathcal{R}}_i^{[m_g]}(X_{gv}), \tilde{\mathcal{R}}_j^{[m_h]}(X_{hw})) = \begin{cases} (\delta_{ij} - (1/m_g))4 & \text{if } g = h,\ v = w, \forall i, j; \\ (\delta_{ij} - (1/m_g))3.5 & \text{if } g = h,\ v \neq w, \forall i, j; \\ 0 & \text{otherwise.} \end{cases} \quad (4.138)$$

where $\gamma_{gh} = 1$ if $g = h$, and $\gamma_{gh} = \gamma$ otherwise, for $-1 \leq \gamma \leq 1$. The examiner, through use of Corollary 16, calculates the canonical structure for the adjustment for the gth marker for a sample of size n. He has already, see equation (2.118), solved the underlying canonical variable problem and so his results are immediate. The canonical directions, for each g are

$$\begin{aligned} Z_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n)]1} &\propto \tilde{\mathcal{M}}(X_{g-}) & (4.139)\\ &= \tilde{\mathcal{M}}(X_{g1}) - \tilde{\mathcal{M}}(X_{g2}); & (4.140)\\ Z_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n)]2} &\propto \tilde{\mathcal{M}}(X_{g+}) & (4.141)\\ &= \tilde{\mathcal{M}}(X_{g1}) + \tilde{\mathcal{M}}(X_{g2}), & (4.142) \end{aligned}$$

with the corresponding canonical resolutions being

$$\begin{aligned} \lambda_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n)]1} &= \lambda_{(n)g-} + \frac{n}{m_g}(1 - \lambda_{(n)g-}); & (4.143)\\ \lambda_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n)]2} &= \lambda_{(n)g+} + \frac{n}{m_g}(1 - \lambda_{(n)g+}), & (4.144) \end{aligned}$$

where $\lambda_{(n)g-}$ is as given by equation (2.52) and $\lambda_{(n)g+}$ is as given by equation (2.51). Notice the crucial difference between the finite and infinite assumptions here. For each of the two markers, the second-order exchangeable sequences $\mathcal{C}_1^*$ and $\mathcal{C}_2^*$ are judged to have the same relationships between each pair of individuals. We observe the same number of individuals' marks for each marker. When we judged $\mathcal{C}_1^*$ and $\mathcal{C}_2^*$ both to be the union of an infinite number of individuals then we observed that the canonical resolutions for the adjustment, $\lambda_{(n)g-}$ and $\lambda_{(n)g+}$ did not depend upon g. However, in the finite case, unless $m_1 = m_2$, $\lambda_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n)]1}$ and $\lambda_{[\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)/\mathcal{C}_g(n)]2}$ do depend upon g. We will learn most for the marker who has the smallest number of available individuals. This is not surprising, for in the finite case, we do physically observe part of the population mean collection, whereas in the infinite case, the mean components are unobservable. Thus, we will learn most in the marker group where we have observed the greatest proportion of the available population. This is intuitively clear, but it is worth making explicit the role of the factor (n_g/m_g) in the finite population correction and that this is the only appearance of m_g in the resolution.

4.13 Underlying canonical finite group problems

Definition 20 *The sth underlying canonical finite group directions are defined as the columns of the matrix* $\tilde{W}_s = [\tilde{W}_{1s} \dots \tilde{W}_{g_0 s}]$ *solving the generalised eigenvalue* *problem*

$$\{\phi_s A + (1-\phi_s)M^{-1}B\}\tilde{W}_s \quad = \quad \{\phi_s A + (1-\phi_s)N^{-1}B\}\tilde{W}_s\tilde{\Lambda}_{(N)s}, \quad (4.145)$$

where $\tilde{\Lambda}_{(N)s} = diag(\tilde{\lambda}_{(N)1s}, \ldots, \tilde{\lambda}_{(N)g_0s})$ *is the matrix of eigenvalues.* $\tilde{W}_s$ *is chosen so that* $\tilde{W}_s^T\{A+((1/\phi_s)-1)M^{-1}B\}\tilde{W}_s = I$, $\tilde{W}_s^T\{A+((1/\phi_s)-1)N^{-1}B\}\tilde{W}_s\Psi_s = I$. *The ordered eigenvalues* $1 > \tilde{\lambda}_{(N)1s} \geq \ldots \geq \tilde{\lambda}_{(N)g_0s} > 0$ *are termed the sth underlying canonical finite group resolutions.*

We motivate this definition on the identical fashion to the infinite case by forming, for each $s = 1, \ldots, v_0$, the analogous collection to $\mathcal{Z}_{(N)s}$ (see equation (2.124)). Thus, for each s, we form the collection

$$\tilde{\mathcal{Z}}_{(N)s} = \{Z_{[\tilde{\mathcal{M}}^{[m_1]}(C_1)/C_1(n_1)]s}, \ldots, Z_{[\tilde{\mathcal{M}}^{[m_{g_0}]}(C_{g_0})/C_{g_0}(n_{g_0})]s}\}. \quad (4.146)$$

Letting D_{Ms} be the $g_0 \times g_0$ matrix D_{Ms} with (g,g)th entry

$$(D_{Ms})_{gg} \quad = \quad \sqrt{\frac{m_g\phi_s}{m_g\alpha_{gg}\phi_s + \beta_g(1-\phi_s)}}, \quad (4.147)$$

then in our usual vector notation, we also represent $\tilde{\mathcal{Z}}_{(N)s}$ as the $g_0 \times 1$ vector

$$\tilde{\mathcal{Z}}_{(N)s} \quad = \quad D_{Ms}(I_{g_0} \otimes Y_s^T)\tilde{\mathcal{M}}(\mathcal{C}). \quad (4.148)$$

Notice the similarity with the representation in the infinite setting, given by equation (2.126).

Lemma 24 *The second-order relationships between* $\tilde{\mathcal{Z}}_{(N)s}$ *and* $\mathcal{S}_N(\mathcal{C})$ *may be expressed as follows*

$$\begin{aligned}
Var(\tilde{\mathcal{Z}}_{(N)s}) &= (1/\phi_s)D_{Ms}\{\phi_s A + (1-\phi_s)M^{-1}B\}D_{Ms}; && (4.149)\\
Cov(\tilde{\mathcal{Z}}_{(N)s}, \mathcal{S}_N(\mathcal{C})) &= (1/\phi_s)D_{Ms}[\epsilon_{v_0s}^T \otimes \{\phi_s A + (1-\phi_s)M^{-1}B\}] \times \\
&\quad [(I_{g_0} \otimes Y_1)\ldots(I_{g_0} \otimes Y_{v_0})]^{-1}; && (4.150)\\
Var(\mathcal{S}_N(\mathcal{C})) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1}\begin{bmatrix} I_{g_0} \otimes \epsilon_{v_01}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\oplus_{s=1}^{v_0}\{\phi_s A + (1-\phi_s)N^{-1}B\}] \times \\
&\quad [(I_{g_0} \otimes Y_1)\ldots(I_{g_0} \otimes Y_{v_0})]^{-1}; && (4.151)\\
Cov(\mathcal{S}_N(\mathcal{C}), \tilde{\mathcal{Z}}_{(N)s}) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1}\begin{bmatrix} I_{g_0} \otimes \epsilon_{v_01}^T \\ \vdots \\ I_{g_0} \otimes \epsilon_{v_0v_0}^T \end{bmatrix}^{-1} \times \\
&\quad [\epsilon_{v_0s} \otimes \{\phi_s A + (1-\phi_s)M^{-1}B\}]D_{Ms}. && (4.152)
\end{aligned}$$

Proof - The result follows in a similar manner to Lemma 8, making use of equation (4.148). □

Denote the resolution transform for the adjustment of $\tilde{\mathcal{Z}}_{(N)s}$ by $\mathcal{C}(N)$ as $T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{C}(N)]}$. We have the following theorem.

Theorem 32 *For the adjustment of $[\tilde{\mathcal{Z}}_{(N)s}]$ by $\mathcal{C}(N)$, the resolution transform matrix is calculated as*

$$T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{C}(N)]} = D_{Ms}^{-1}\{\phi_s A + (1-\phi_s)N^{-1}B\}^{-1}\{\phi_s A + (1-\phi_s)M^{-1}B\}D_{Ms}. \quad (4.153)$$

The canonical directions are given by

$$\tilde{Z}_{(N)ds} = (D_{Ms}^{-1}\tilde{W}_{ds})^T\tilde{\mathcal{Z}}_{(N)s} = (\tilde{W}_{ds}\otimes Y_s)^T\tilde{\mathcal{M}}(\mathcal{C}), \quad (4.154)$$

for each $d = 1, \dots g_0$, with corresponding canonical resolutions given by $\tilde{\lambda}_{(N)ds}$. $\tilde{W}_{ds}$ is the (d,s)th underlying canonical finite group direction as given in Definition 20; $\tilde{\lambda}_{(N)ds}$ is the corresponding (d,s)th underlying canonical finite group resolution.

Proof - From Theorem 31, $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(N)$ for adjusting $\tilde{\mathcal{M}}(\mathcal{C})$. For each s, we have $[\tilde{\mathcal{Z}}_{(N)s}] \subset [\tilde{\mathcal{M}}(\mathcal{C})]$, so that $\mathcal{S}_N(\mathcal{C})$ is Bayes linear sufficient for adjusting $\tilde{\mathcal{Z}}_{(N)s}$. Thus, $T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{C}(N)]} = T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}$, where $T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}$ is the resolution transform for the adjustment of $\tilde{\mathcal{Z}}_{(N)s}$ by $\mathcal{S}_N(\mathcal{C})$. From equation (1.74), $T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}$ may be computed as

$$T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{S}_N(\mathcal{C})]} = \{Var(\tilde{\mathcal{Z}}_{(N)s})\}^{-1}Cov(\tilde{\mathcal{Z}}_{(N)s}, \mathcal{S}_N(\mathcal{C})) \{Var(\mathcal{S}_N(\mathcal{C}))\}^{-1}Cov(\mathcal{S}_N(\mathcal{C}), \tilde{\mathcal{Z}}_{(N)s}). \quad (4.155)$$

Thus, by inverting equation (4.149), post multiplying by equation (4.150) and then the inversion of (4.151) and finally by (4.152) we may obtain equation (4.153).

From the solution of the underlying group problem as given by equation (4.145), it is straightforward to see that

$$T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{S}_N(\mathcal{C})]} = D_{Ms}^{-1}\tilde{W}_s\tilde{\Lambda}_{(N)s}\tilde{W}_s^{-1}D_{Ms}, \quad (4.156)$$

so that

$$T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{S}_N(\mathcal{C})]}(D_{Ms}^{-1}\tilde{W}_s) = D_{Ms}^{-1}\tilde{W}_s\tilde{\Lambda}_{(N)s}. \quad (4.157)$$

Hence, $\tilde{\Lambda}_{(N)s}$ is the matrix whose diagonal elements are the canonical resolutions of the adjustment. To confirm that the $(D_{Ms}^{-1}\tilde{W}_{ds})^T\tilde{\mathcal{Z}}_{(N)s}$ are the corresponding canonical resolutions, we verify that they are mutually uncorrelated with prior variance one. Notice from equation (4.148) we have

$$\begin{aligned}
& Cov((D_{Ms}^{-1}\tilde{W}_{ds})^T \tilde{\mathcal{Z}}_{(N)s}, (D_{s'}^{-1}\tilde{W}_{d's'})^T \tilde{\mathcal{Z}}_{(N)s'}) = \\
& \quad (\tilde{W}_{ds}^T \otimes Y_s^T) Var(\tilde{\mathcal{M}}(\mathcal{C}))(\tilde{W}_{d's'} \otimes Y_{s'}) && (4.158)\\
&= (\tilde{W}_{ds}^T \otimes Y_s^T)\{(A \otimes C) + (M^{-1}B \otimes E)\}(\tilde{W}_{d's'} \otimes Y_{s'}) && (4.159)\\
&= \{(\tilde{W}_{ds}^T A \tilde{W}_{d's'} \otimes Y_s^T C Y_{s'}) + (\tilde{W}_{ds}^T M^{-1} B \tilde{W}_{d's'} \otimes Y_s^T E Y_{s'})\} && (4.160)\\
&= \delta_{ss'}[\tilde{W}_{ds}^T\{A + ((1/\phi_s) - 1)M^{-1}B\}\tilde{W}_{d's}] && (4.161)\\
&= \delta_{ss'}\delta_{dd'}. && (4.162)
\end{aligned}$$

Equation (4.161) follows by the choice of Y in Definition 12 and equation (4.162) follows by the choice of $\tilde{W}_s$ in Definition 20. Setting $s = s'$ completes the verification. Notice that equation (4.162) also shows that, since the $\tilde{Z}_{(N)ds}$ form a basis for $[\tilde{\mathcal{Z}}_{(N)s}]$, then if $s \neq s'$, $[\tilde{\mathcal{Z}}_{(N)s}] \perp [\tilde{\mathcal{Z}}_{(N)s'}]$. □

This theorem is the finite equivalent of Theorem 12. Notice how if we let each $m_g \to \infty$ then $D_{Ms} \to D$ (as given by equation (2.125)) and also $T_{[\tilde{\mathcal{Z}}_{(N)s}/\mathcal{C}(N)]} \to T_{[\mathcal{Z}_{(N)s}/\mathcal{C}(N)]}$ (as given by equation (2.145)). The sth canonical group structure may be derived from the problem of learning about the relationships of the sth most important variable directions when considered across the groups. Notice the similarity in form of equations (4.154) and (2.146) but also the crucial difference. Equation (4.154) will, in general, depend upon the choice of s, so that the adjustment over each $[\tilde{\mathcal{Z}}_{(N)s}]$ does not share the same qualitative features for each s. This contrasts with the scenario in the finite case where the adjustment over each $[\mathcal{Z}_{(N)s}]$ does provide the same qualitative information. This result is, perhaps, not surprising if we think back to equation (4.136). From equation (2.98), we have that

$$\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s} \geq \lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s'} \quad \Leftrightarrow \quad \phi_s \geq \phi_{s'} \tag{4.163}$$

so that the $\lambda_{[\mathcal{M}(\mathcal{C}_g)/\mathcal{C}_g(n_g)]s}$s are ordered according to the ϕ_ss, irrespective of g. Thus, for large values of ϕ_s the finite population correction term will have a much smaller effect than for smaller values of ϕ_s. Recall that from equation (4.95), we perceived that the fraction (n_g/m_g) could be used as a rule of thumb for assessing the validity of an infinite approximation and that from equation (4.136), we suggested that there was a difference in the groups in that those with small values of (n_g/m_g) may be perceived as being closer to the infinite case than those with larger values of (n_g/m_g). Now, from equation (4.136), we can suggest that if $\phi_s > \phi_{s'}$ this difference between the groups in their closeness to being infinite is damped down so that if $s < s'$, we would expect, since $\phi_s \geq \phi_{s'}$, that the quantities in the collection $\tilde{\mathcal{Z}}_{(N)s}$ to be at least as well approximated by the collection $\mathcal{Z}_{(N)s}$ as the collection $\tilde{\mathcal{Z}}_{(N)s'}$ is approximated by $\mathcal{Z}_{(N)s'}$.

4.13.1 The underlying canonical finite group problem for the examiner

From equation (4.140), the examiner forms the collection

$$\begin{aligned}\tilde{\mathcal{Z}}_{(nI_2)1} &= \Big\{\sqrt{(2\lambda_{(m_1)-}/3)}(\tilde{\mathcal{M}}(X_{11}) - \tilde{\mathcal{M}}(X_{12})),\\ &\qquad \sqrt{(2\lambda_{(m_2)-}/3)}(\tilde{\mathcal{M}}(X_{21}) - \tilde{\mathcal{M}}(X_{22}))\Big\}. \end{aligned} \quad (4.164)$$

From equation (4.142), he forms the collection

$$\begin{aligned}\tilde{\mathcal{Z}}_{(nI_2)2} &= \Big\{\sqrt{(2\lambda_{(m_1)+}/5)}(\tilde{\mathcal{M}}(X_{11}) + \tilde{\mathcal{M}}(X_{12})),\\ &\qquad \sqrt{(2\lambda_{(m_2)+}/5)}(\tilde{\mathcal{M}}(X_{21}) + \tilde{\mathcal{M}}(X_{22}))\Big\}. \end{aligned} \quad (4.165)$$

$\lambda_{(m)-}$ and $\lambda_{(m)+}$ are as given in Table 1.2. The inner product spaces $[\tilde{\mathcal{Z}}_{(nI_2)1}]$ and $[\tilde{\mathcal{Z}}_{(nI_2)2}]$ are then formed. The examiner constructs the sth underlying canonical finite group problem by making use of A and B as given in equation (2.117) and that $N = nI_2$ and $M = diag(m_1, m_2)$. It should be emphasised that in this balanced design, n is constrained to be at most $min(m_1, m_2)$. The problem is then reduced to finding the eigenstructure of the matrix, $D_{Ms}T_{[\tilde{\mathcal{Z}}_{(nI_2)s}/\mathcal{C}(nI_2)]}D_{Ms}^{-1}$, where

$$\begin{aligned} D_{Ms}T_{[\tilde{\mathcal{Z}}_{(nI_2)s}/\mathcal{C}(nI_2)]}D_{Ms}^{-1} &= \begin{pmatrix} \phi_s + (1/n)(1-\phi_s) & \phi_s\gamma \\ \phi_s\gamma & \phi_s + (1/n)(1-\phi_s) \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} \phi_s + (1/m_1)(1-\phi_s) & \phi_s\gamma \\ \phi_s\gamma & \phi_s + (1/m_2)(1-\phi_s) \end{pmatrix}. \end{aligned} \quad (4.166)$$

The examiner thus solves this sth underlying canonical finite group problem to obtain

$$\begin{aligned} \tilde{\lambda}_{(nI_2)1s} &= \lambda_{(nI_2)1s} + \frac{n}{m_1}(1 - \lambda_{(nI_2)1s}) + \\ &\left(\frac{1}{m_2} - \frac{1}{m_1}\right)\frac{n}{4}\{2 - \lambda_{(nI_2)1s} - \lambda_{(nI_2)2s}\} - (1 - \delta_{m_1 m_2})g_1(\phi_s); \quad (4.167) \\ \tilde{W}_{1s} &\propto [1 \ \delta_{m_1 m_2} + ((1/m_2) - (1/m_1))g_2(\phi_s)]^T; \quad (4.168) \\ \tilde{\lambda}_{(nI_2)2s} &= \lambda_{(nI_2)2s} + \frac{n}{m_1}(1 - \lambda_{(nI_2)2s}) + \\ &\left(\frac{1}{m_2} - \frac{1}{m_1}\right)\frac{n}{4}\{2 - \lambda_{(nI_2)1s} - \lambda_{(nI_2)2s}\} + (1 - \delta_{m_1 m_2})g_1(\phi_s); \quad (4.169) \\ \tilde{W}_{2s} &\propto [1 \ -\delta_{m_1 m_2} + ((1/m_2) - (1/m_1))g_2(\phi_s)]^T; \quad (4.170) \end{aligned}$$

$g_1(\phi_s)$ and $g_2(\phi_s)$ are functions of ϕ_s obtained by regarding the solution of the sth underlying canonical finite group problem as a function of ϕ_s. $g_2(\phi_s)$ does vary with n so that $\tilde{W}_{1s}$ and $\tilde{W}_{2s}$ are not the same for each sample size, which contrasts with the infinite case. $\lambda_{(nI_2)1s}$ and $\lambda_{(nI_2)2s}$ are the canonical resolutions for the adjustment of $[\mathcal{Z}_{(nI_2)s}]$ by $\mathcal{C}(nI_2)$. For their precise forms, see equations (2.159) and (2.160), or equivalently, Table 2.1 and Table 2.2.

The important observation to make is when $m_1 = m_2$. In this case, $M \propto N$ and so we see an equal proportion of the population mean collection in each group. Notice that, in this case, $\tilde{\lambda}_{(nI_2)ds}$ may be thought of as $\lambda_{(nI_2)ds}$ plus a finite population correction term, this correction term having precisely the same form as that for the adjustment of a single finite second-order exchangeable sequence, see equation (4.94). In this case, we also see that $\tilde{W}_{ds}$ do not depend upon s and are proportional to W_d, the dth underlying canonical group direction as given in Definition 12. We shall show that these observations are not typical to this example, but a completely expected consequence.

There is one feature of this example which is a consequence of the example for it is driven by the additional second-order exchangeability of the questions and the balanced sample, but is perhaps worth noting. We shall make use of the prior uncertainty in the system and the resolved uncertainty for the collection, see Goldstein (2000; p9) for further details. We have that the resolved uncertainty for $[\tilde{\mathcal{Z}}_{(nI_2)s}]$, for each s, is

$$RU_{nI_2}(\tilde{\mathcal{Z}}_{(nI_2)s}) = \sum_{d=1}^{2} \tilde{\lambda}_{(nI_2)ds}, \tag{4.171}$$

and the prior uncertainty in the collection $\tilde{\mathcal{Z}}_{(nI_2)s}$, for each s, is $U(\tilde{\mathcal{Z}}_{(nI_2)s}) = 2$. The resolved uncertainty for $[\mathcal{Z}_{(nI_2)s}]$, for each s, is

$$RU_{nI_2}(\mathcal{Z}_{(nI_2)s}) = \sum_{d=1}^{2} \lambda_{(nI_2)ds}, \tag{4.172}$$

and the prior uncertainty in the collection $\mathcal{Z}_{(nI_2)s}$, for each s, is $U(\mathcal{Z}_{(nI_2)s}) = 2$. We define the remaining uncertainty in the collection $\tilde{\mathcal{Z}}_{(nI_2)s}$, for each s, is

$$U(\tilde{\mathcal{Z}}_{(nI_2)s}) - RU_{nI_2}(\tilde{\mathcal{Z}}_{(nI_2)s}) = \sum_{d=1}^{2}(1 - \tilde{\lambda}_{(nI_2)ds}) \tag{4.173}$$

$$= \left\{1 - \frac{n}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\right\} \sum_{d=1}^{2}(1 - \lambda_{(nI_2)ds}) \tag{4.174}$$

$$= \left\{1 - \frac{n}{2}\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\right\} \{U(\mathcal{Z}_{(nI_2)s}) - RU_{nI_2}(\mathcal{Z}_{(nI_2)s})\}. \tag{4.175}$$

This holds for any choice of m_1 and m_2 and is a means of assessing the remaining uncertainty left in each collection after the adjustment. The form of this result should be compared with equation (4.95) summed over all the canonical directions. For this overall summary, the finite and infinite cases are easily comparable.

4.14 The adjustment of the mean components by the observed sample

We now consider the adjustment of the full collection, $[\tilde{\mathcal{M}}(\mathcal{C})]$. We shall show that once more the underlying canonical variable analysis and each underlying canonical finite group analysis completely determine the adjustment of the full collection, $[\tilde{\mathcal{M}}(\mathcal{C})]$. To proceed, we re-express equations (4.129) - (4.131) as the following lemma; the proof follows by various applications of Lemma 7.

Lemma 25 *The relationships between the $\tilde{\mathcal{M}}(\mathcal{C})$ and the $\mathcal{S}_N(\mathcal{C})$ may be expressed as follows*

$$\begin{aligned} Var(\tilde{\mathcal{M}}(\mathcal{C})) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g0} \otimes \epsilon_{v_0 1}^T \\ \cdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\ & [\oplus_{s=1}^{v_0} \phi_s A + (1-\phi_s) M^{-1} B\}\tilde{W}_s][(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})]^{-1}; \quad (4.176) \\ Cov(\tilde{\mathcal{M}}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) &= Var(\tilde{\mathcal{M}}(\mathcal{C})); \quad (4.177) \\ Var(\mathcal{S}_N(\mathcal{C})) &= [I_{g_0} \otimes Y^{-1}(C+E)^{-1}]^{-1} \begin{bmatrix} I_{g0} \otimes \epsilon_{v_0 1}^T \\ \cdots \\ I_{g_0} \otimes \epsilon_{v_0 v_0}^T \end{bmatrix}^{-1} \times \\ & [\oplus_{s=1}^{v_0} \phi_s A + (1-\phi_s) N^{-1} B\}\tilde{W}_s][(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})]^{-1}; \quad (4.178) \\ Cov(\mathcal{S}_N(\mathcal{C}), \tilde{\mathcal{M}}(\mathcal{C})) &= Var(\tilde{\mathcal{M}}(\mathcal{C})). \quad (4.179) \end{aligned}$$

Denote by $T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{C}(N)]}$ the resolution transform for the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C})]$ by $\mathcal{C}(N)$. We have the following theorem.

Theorem 33 *The adjustment of $\tilde{\mathcal{M}}(\mathcal{C})$ by $\mathcal{C}(N)$ satisfies the following properties*

1. There exist v_0 g_0-dimensional orthogonal subspaces, $[\tilde{\mathcal{Z}}_{(N)1}]$, ..., $[\tilde{\mathcal{Z}}_{(N)v_0}]$ of $[\tilde{\mathcal{M}}(\mathcal{C})]$. For each s,

$$\tilde{\mathcal{Z}}_{(N)s} = \{Z_{[\tilde{\mathcal{M}}^{[m_1]}(\mathcal{C}_1)/\mathcal{C}_1(n_1)]s}, \cdots, Z_{[\tilde{\mathcal{M}}^{[m_{g_0}]}(\mathcal{C}_{g_0})/\mathcal{C}_{g_0}(n_{g_0})]s}\}. \quad (4.180)$$

2. The canonical directions for the adjustment in each $[\tilde{\mathcal{Z}}_{(N)s}]$ are given by the columns of the matrix $\tilde{Z}_{(N)s} = [\tilde{Z}_{(N)1s} \ldots \tilde{Z}_{(N)g_0 s}]$, where

$$\tilde{Z}_{(N)ds} = (D_{Ms}^{-1}\tilde{W}_{ds})^T \tilde{\mathcal{Z}}_{(N)s} = (\tilde{W}_{ds} \otimes Y_s)^T \tilde{\mathcal{M}}(\mathcal{C}). \quad (4.181)$$

3. The resolution transform matrix for the adjustment is calculated as

$$\begin{aligned} T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{C}(N)]} &= \{(A \otimes C) + (N^{-1}B \otimes E)\}^{-1}\{(A \otimes C) + (M^{-1}B \otimes E)\} \quad (4.182) \\ &= [(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})][\oplus_{s=1}^{v_0} \tilde{W}_s^T \{\phi_s A + (1-\phi_s) N^{-1} B\}^{-1} \times \\ & \{\phi_s A + (1-\phi_s) M^{-1} B\}\tilde{W}_s][(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})]^{-1}. \quad (4.183) \end{aligned}$$

4. The collection $\tilde{Z}_{(N)} = \{\tilde{Z}_{(N)ds}\}$ *for* $d = 1, \ldots, g_0$, $s = 1, \ldots, v_0$, *are the canonical directions of the adjustment with canonical resolutions given by* $\tilde{\lambda}_{(N)ds}$.

Proof - Statement 1. follows from equation (4.162).

Statement 2. follows immediately from Theorem 32.

To obtain Statement 3., we use the Bayes linear sufficiency of $\mathcal{S}_N(\mathcal{C})$ for $\mathcal{C}(N)$ for adjusting $\tilde{\mathcal{M}}(\mathcal{C})$ which follows from Theorem 31. Letting $T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$ be the resolution transform for the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C})]$ by $\mathcal{S}_N(\mathcal{C})$, we have that $T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{C}(N)]} = T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$. From equation (1.74) we have that

$$\begin{aligned} T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]} &= \{Var(\tilde{\mathcal{M}}(\mathcal{C}))\}^{-1}Cov(\tilde{\mathcal{M}}(\mathcal{C}), \mathcal{S}_N(\mathcal{C})) \times \\ &\qquad \{Var(\mathcal{S}_N(\mathcal{C}))\}^{-1}Cov(\mathcal{S}_N(\mathcal{C}), \tilde{\mathcal{M}}(\mathcal{C})). \quad (4.184) \end{aligned}$$

We may verify equation (4.182) by using equations (4.129) - (4.131) in equation (4.184). Equation (4.183) follows by using equations (4.176) - (4.179) in equation (4.184).

Using equation (4.183) and substituting in the solution of the sth underlying canonical finite group problem, as given by Definition 20, yields

$$\begin{aligned} T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]} &= [(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})][\oplus_{s=1}^{v_0} \tilde{\Lambda}_{(N)s}] \times \\ &\qquad [(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})]^{-1}. \quad (4.185) \end{aligned}$$

Thus,

$$\begin{aligned} T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}[(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})] &= \\ &\qquad [(\tilde{W}_1 \otimes Y_1) \ldots (\tilde{W}_{v_0} \otimes Y_{v_0})][\oplus_{s=1}^{v_0} \tilde{\Lambda}_{(N)s}]. \quad (4.186) \end{aligned}$$

It then follows that $(\tilde{Z}_{(N)ds}, \tilde{\lambda}_{(N)ds})$ is an eigenvector/value pair for $T_{[\tilde{\mathcal{M}}(\mathcal{C})/\mathcal{S}_N(\mathcal{C})]}$. Since $\tilde{Z}_{(N)ds}$ has prior variance one, $(\tilde{Z}_{(N)ds}, \tilde{\lambda}_{(N)ds})$ are the pairs of canonical directions/resolutions. Statement 4. thus follows. □

This is the finite version of Theorem 13. Notice the similarities. We break down the $v_0 g_0 \times v_0 g_0$ problem into one problem of size $v_0 \times v_0$ and then v_0 problems of size $g_0 \times g_0$. Each $g_0 \times g_0$ problem has a similar form and so in fact we need only to solve the sth underlying group problem as the functions $\tilde{W}_s = \tilde{W}(\phi_s)$ and $\tilde{\Lambda}_s = \tilde{\Lambda}(\phi_s)$ to reveal the complete solution. The $v_0 \times v_0$ problem consists of finding the underlying canonical variable structure, and through its motivation as being the single group adjustment, we see how and why this problem remains the same in both the finite and infinite modelling assumptions. Having solved across the variables, we then use this solution to solve across the groups. Each collection, $\tilde{\mathcal{Z}}_{(N)s}$ is analogous to the collection $\mathcal{Z}_{(N)s}$, so that our understanding of the design remains constant in the finite and infinite cases. Recall how when we dealt with a single finite exchangeable sequence, see for example Corollary 16, that the difference was quantitative as opposed to qualitative. In this grouped framework, we observe that there is a qualitative difference as well which may be observed happening in the collections $\mathcal{Z}_{(N)s}$. Intuitively, see our comments following Corollary 17 and Theorem 32, we felt that this was caused by the fact that in the finite case we actually observe part of the population mean collection and so for general choices of M and N, certain groups will be better approximated by the infinite assumption. Explicitly, we considered (n_g/m_g) to be a 'rule of thumb' for the validity of the infinite assumption. The following corollary shows that this intuition was well based.

Corollary 18 *Suppose that there exists $\theta \geq 1$ such that $M = \theta N$. Then the adjustment in $[\tilde{\mathcal{Z}}_{(N)s}]$ is qualitatively the same for each s and is qualitatively the same as the adjustment in each $[\mathcal{Z}_{(N)s}]$. The canonical directions of the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C})]$ by $\mathcal{C}(N)$ and the canonical directions of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$ thus have, up to a scale factor, the same co-ordinate representation. If $Z_{(N)ds} = (W_d \otimes Y_s)^T \mathcal{M}(\mathcal{C})$ is a canonical direction of the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(N)$, with canonical resolution $\lambda_{(N)ds}$, then*

$$\tilde{Z}_{(N)ds} = \lambda_{(M)ds}^{-\frac{1}{2}} (W_d \otimes Y_s)^T \tilde{\mathcal{M}}(\mathcal{C}) \tag{4.187}$$

is the corresponding canonical direction of the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C})]$ by $\mathcal{C}(N)$ with canonical resolution

$$\tilde{\lambda}_{(N)ds} = \lambda_{(N)ds} + \frac{1}{\theta}(1 - \lambda_{(N)ds}). \tag{4.188}$$

$\lambda_{(M)ds}$ is the canonical resolution corresponding to the canonical direction $Z_{(M)ds} = Z_{(N)ds}$ for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(M)$ and may be found from equation (2.184).

In this case, the link between the finite and infinite case is much clearer. The solution of the underlying variable and group canonical problems given in Definition 12 and Definition 13 allow us to completely determine the finite and infinite modelling solutions. Notice the similarity of equation (4.188) with equation (4.94). The finite assumption provides the same qualitative information as the adjustment using the infinite approximation and the quantitative differences are easily obtainable and comparable. This is the feature that we drew attention to in the examiner example when we take $m_1 = m_2$ in equations (4.167) - (4.170).

We may obtain the analogous form of equation (4.95) which is

$$Var_N(\tilde{Z}_{(N)ds}) = \left(1 - \frac{1}{\theta}\right) Var_N(Z_{(N)ds}) \qquad (4.189)$$

so that in this group setting, we may use $(1/\theta)$ as a 'rule of thumb' for the validity of the infinite approximation. If both N and M are balanced then we may easily adopt Corollary 4 to relate the finite and infinite cases. Although it may be rare that Corollary 18 holds, it is useful for it provides us with understanding as to how the infinite approximation in each group fits in with the more realistic modelling in each group. We see that the infinite approximation only plays a role in the underlying canonical finite group problems and we may perceive these as being perturbations of the underlying canonical group problem for the infinite case, as given by Definition 12. Thus, we can perceive the canonical resolutions deduced from the adjustment of each $[\tilde{\mathcal{Z}}_{(N)s}]$ as being rotations of the infinite equivalent directions deduced from the adjustment of $[\mathcal{Z}_{(N)s}]$. This interpretation may also allow us to appreciate differences in the choice of possible sequence lengths, M, should we be uncertain here.

In this grouped case, the infinite assumption means we solve only a single $g_0 \times g_0$ problem, so has computational advantages, but we also have a means of assessing the validity of the infinite assumption by comparing this $g_0 \times g_0$ problem with each, or indeed any, of the v_0 $g_0 \times g_0$ problems for the finite case.

4.14.1 The full adjustment for the examiner

Having already solved the underlying canonical variable problem and each underlying canonical finite group problem, from Theorem 33 the examiner has completed all the calculations he need perform for the adjustment of the full collection. The canonical directions for the adjustment are $\tilde{\lambda}_{(nI_2)ds}$, for $d = 1, 2$ and $s = 1, 2$, as given by equations (4.167) and (4.170). The corresponding canonical directions are given by $\tilde{Z}_{(nI_2)ds} = (\tilde{W}_{ds} \otimes Y_s)^T \tilde{\mathcal{M}}(\mathcal{C})$. Notice from equations (4.168) and (4.170) when we set $m_1 = m_2$, so that $M \propto N$, Corollary 18 reveals that $\tilde{W}_{ds} \propto W_d$, where W_d is the dth column of the matrix W as given by equation (2.156). Equation (4.188) is clearly seen, in this instance, in equations (4.167) and (4.170) as we observed in the comments following these two equations.

Chapter 5

Extensions: one-level hierarchical co-exchangeability

SUMMARY

In this chapter, we illustrate another avenue to co-exchangeable modelling, namely that of the one-level hierarchical model. We explore this model in the context of balanced designs where the residuals in each group do not explicitly depend upon the group. The model is specified in Section 5.1 and in Section 5.2 we show that in this balanced symmetric design, we may use three different Bayes linear sufficient statistics for the adjustment of the three components in the model: the overall population collection; the gth group residuals; and the group mean collection. In Section 5.3, we adjust the overall population collection. Section 5.4 sees the adjustment of the gth group residuals. We show how learning in the gth group residuals reduces to a problem comparable to the problems studied in Chapter 2. In Section 5.5, we show that this lower level analysis completely determines the adjustment of the group means. Thus, the adjustment of the group mean collection in this model may also be solved by considering a series of subproblems which have interpretable forms.

5.1 The one-level hierarchical co-exchangeable model

We shall consider designs where each individual can be classified as coming from one of g_0 possible groups, where g_0 is the total number of available groups; there is no concept of additional groups. For each individual, we wish to measure the same set of variables, $\mathcal{C} = \{X_1, \ldots, X_{v_0}\}$, and thus denote by $\mathcal{C}_{gi} = \{X_{g1i}, \ldots, X_{gv_0i}\}$ the collection of measurements for the ith individual in the gth group. We shall proceed as if there were a potentially infinite number of individuals that could be observed in each group and denote by $\mathcal{C}_g^*$ the union of the measurements of the individuals in the gth group. We judge that, in accordance with Theorem 7, that $\mathcal{C}_1^*, \ldots, \mathcal{C}_{g_0}^*$, are a sequence of co-exchangeable infinite exchangeable systems and hence we may write

$$X_{gi} = \mathcal{M}(\mathcal{C}_g) + \mathcal{R}_i(\mathcal{C}_g). \tag{5.1}$$

This means we are required to specify the $v_0 \times v_0$ matrix $Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{M}(\mathcal{C}_h)) = C_{gh}$ for each g, h, and the $v_0 \times v_0$ matrix $Var(\mathcal{R}_i(\mathcal{C}_g)) = E_g$, for each g. In Subsection 2.3.1, we mentioned that one approach could be that we judge that the mean components are second-order exchangeable over groups, that is $C_{gg} = F$ for all g and $C_{gh} = P$ for all $g \neq h$. We now adopt this model in this chapter. Hence, we judge that the collection $\mathcal{M}(\mathcal{C}) = \cup_{g=1}^{g_0} \mathcal{M}(\mathcal{C}_g)$ is second-order exchangeable over the groups, see Theorem 25, and we may write

$$\mathcal{M}(\mathcal{C}_g) = \tilde{\mathcal{M}}(\mathcal{C}.) + \tilde{\mathcal{R}}_g(\mathcal{C}), \tag{5.2}$$

where

$$\tilde{\mathcal{M}}(\mathcal{C}.) = \frac{1}{g_0} \sum_{g=1}^{g_0} \mathcal{M}(\mathcal{C}_g). \tag{5.3}$$

We term $\tilde{\mathcal{M}}(\mathcal{C}.)$ the overall population collection, and $\tilde{\mathcal{R}}_g(\mathcal{C})$ the gth group residuals. The full model is

$$X_{gi} = \tilde{\mathcal{M}}(\mathcal{C}.) + \tilde{\mathcal{R}}_g(\mathcal{C}) + \mathcal{R}_i(\mathcal{C}_g). \tag{5.4}$$

The requirements from Theorem 7 and Theorem 25 for the asserted judgements yields the following lemma.

Lemma 26 *The collections $\tilde{\mathcal{M}}(\mathcal{C}.)$, $\tilde{\mathcal{R}}_g(\mathcal{C})$, and $\mathcal{R}_i(\mathcal{C}_g)$ satisfy the following rela-*

tionships

$$\begin{aligned}
E(\tilde{\mathcal{M}}(\mathcal{C}.)) &= m; && (5.5)\\
E(\tilde{\mathcal{R}}_g(\mathcal{C})) &= 0 \ \forall g; && (5.6)\\
E(\mathcal{R}_i(\mathcal{C}_g)) &= 0 \ \forall g, i; && (5.7)\\
Var(\tilde{\mathcal{M}}(\mathcal{C}.)) &= P + \frac{1}{g_0}F; && (5.8)\\
Cov(\tilde{\mathcal{M}}(\mathcal{C}.), \tilde{\mathcal{R}}_h(\mathcal{C})) &= 0 \ \forall h; && (5.9)\\
Cov(\tilde{\mathcal{R}}_g(\mathcal{C}), \tilde{\mathcal{R}}_h(\mathcal{C})) &= \begin{cases} \frac{g_0-1}{g_0}F & \textit{if } g = h; \\ -\frac{1}{g_0}F & \textit{if } g \neq h; \end{cases} && (5.10)\\
Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{R}_j(X_h)) &= 0 \ \forall g, h, j; && (5.11)\\
Cov(\mathcal{R}_i(X_g), \mathcal{R}_j(X_h)) &= \begin{cases} E_g & \textit{if } g = h \textit{ and } i = j; \\ 0 & \textit{otherwise.} \end{cases} && (5.12)
\end{aligned}$$

The $v_0 \times 1$ vector m is arbitrary, whilst the $v_0 \times v_0$ matrices P, F, E_g, for each g, are merely constrained by having to provide a coherent specification.

Once more, we shall assume that P, F, and each E_g are positive definite, so that they are of full rank. This hierarchical co-exchangeable structure is similar to one-way layout multivariate analysis of variance (MANOVA) models; see Press (1989; Chapter VI). Goldstein (1988b) investigated the univariate one-way layout from a Bayes linear perspective. Suppose that we assume the the individual residuals are the same for each g; that is $E_g = E$ for each $g = 1, \ldots, g_0$.

5.2 Sampling and learning about the components of the model

We would like to observe a sample of individuals from each group in order to learn about the components of the model: the overall mean collection, the group residuals and the mean components corresponding to the gth group. Suppose that we observe n individuals in each of the g_0 groups. The total sample is then expressed as $\mathcal{C}(nI_{g_0})$. From Theorem 8 we have that the collection of sample means, $\mathcal{S}_{nI_{g_0}}(\mathcal{C})$, is Bayes linear sufficient for $\mathcal{C}(nI_{g_0})$ for adjusting $[\mathcal{M}(\mathcal{C})]$. Notice that since $[\tilde{\mathcal{M}}(\mathcal{C}.)] \subset [\mathcal{M}(\mathcal{C})]$ then $\mathcal{S}_{nI_{g_0}}(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(nI_{g_0})$ for adjusting $[\tilde{\mathcal{M}}(\mathcal{C}.)]$. Also, for each g, we have that $[\tilde{\mathcal{R}}_g(\mathcal{C})] \subset [\mathcal{M}(\mathcal{C})]$. Thus, $\mathcal{S}_{nI_{g_0}}(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(nI_{g_0})$ for adjusting $[\tilde{\mathcal{R}}_g(\mathcal{C})]$ and hence for adjusting $[\tilde{\mathcal{R}}(\mathcal{C})]$, where $\tilde{\mathcal{R}}(\mathcal{C}) = \cup_{g=1}^{g_0} \tilde{\mathcal{R}}_g(\mathcal{C})$. Notice that of the g_0 elements forming $\tilde{\mathcal{R}}(\mathcal{C})$, only $g_0 - 1$ of them are linearly independent since, for example, we have

$$\begin{aligned}
\sum_{g=1}^{g_0-1}\tilde{\mathcal{R}}_g(\mathcal{C}) &= \sum_{g=1}^{g_0-1}\mathcal{M}(\mathcal{C}_g)-\frac{g_0-1}{g_0}\sum_{g=1}^{g_0}\mathcal{M}(\mathcal{C}_g) && (5.13)\\
&= -\mathcal{M}(\mathcal{C}_{g_0})+\frac{1}{g_0}\sum_{g=1}^{g_0}\mathcal{M}(\mathcal{C}_g) && (5.14)\\
&= -\tilde{\mathcal{R}}_{g_0}(\mathcal{C}). && (5.15)
\end{aligned}$$

From equations (5.8) - (5.12), we may derive the following lemma.

Lemma 27 *The second-order relationships between the $\mathcal{M}(\mathcal{C}_g)$s and the $\mathcal{S}_n(\mathcal{C}_g)$s may be expressed as*

$$\begin{aligned}
Cov(\mathcal{M}(\mathcal{C}_g),\mathcal{M}(\mathcal{C}_h)) &= \begin{cases} P+F & \text{if } g=h;\\ P & \text{if } g\neq h;\end{cases} && (5.16)\\
Cov(\mathcal{M}(\mathcal{C}_g),\mathcal{S}_n(\mathcal{C}_h)) &= \begin{cases} P+F & \text{if } g=h;\\ P & \text{if } g\neq h;\end{cases} && (5.17)\\
Cov(\mathcal{S}_n(\mathcal{C}_g),\mathcal{S}_n(\mathcal{C}_h)) &= \begin{cases} P+F+\frac{1}{n}E & \text{if } g=h;\\ P & \text{if } g\neq h.\end{cases} && (5.18)
\end{aligned}$$

Notice that we may equivalently write equations (5.16) - (5.18) as

$$\begin{aligned}
Var(\mathcal{M}(\mathcal{C})) &= \{(I_{g_0}\otimes F)+(J_{g_0}\otimes P)\}; && (5.19)\\
Cov(\mathcal{M}(\mathcal{C}),\mathcal{S}_{nI_{g_0}}(\mathcal{C})) &= \{(I_{g_0}\otimes F)+(J_{g_0}\otimes P)\}; && (5.20)\\
Var(\mathcal{S}_{nI_{g_0}}(\mathcal{C})) &= [\{I_{g_0}\otimes(F+(1/n)E)\}+\{J_{g_0}\otimes P\}] && (5.21)\\
&= [\{I_{g_0}\otimes(F+(1/n)E)^{-1}\}+\{J_{g_0}\otimes B\}]^{-1}. && (5.22)
\end{aligned}$$

B is the matrix $-\{F+(1/n)E\}^{-1}P\{F+g_0P+(1/n)E\}^{-1}$ and multiplication of equations (5.21) and (5.22) confirms the validity of equation (5.22). Furthermore, equation (5.17) shows that $\mathcal{S}_{nI_{g_0}}(\mathcal{C})$ is second-order exchangeable over the groups. We may then use the representation theorem, see Theorem 25, to write

$$\mathcal{S}_n(\mathcal{C}_g) = \tilde{\mathcal{S}}_{nI_{g_0}}(\mathcal{C})+\tilde{\mathcal{T}}_g(\mathcal{C}(nI_{g_0})). \quad (5.23)$$

where

$$\tilde{\mathcal{S}}_{nI_{g_0}}(\mathcal{C}) = \frac{1}{g_0}\sum_{g=1}^{g_0}\mathcal{S}_n(\mathcal{C}_g). \quad (5.24)$$

The second-order relationships are

$$\begin{aligned}
Var(\tilde{\mathcal{S}}_{nI_{g_0}}(\mathcal{C})) &= P+\frac{1}{g_0}\left(F+\frac{1}{n}E\right); && (5.25)\\
Cov(\tilde{\mathcal{S}}_{nI_{g_0}}(\mathcal{C}),\tilde{\mathcal{T}}_h(\mathcal{C}(nI_{g_0}))) &= 0\ \forall g,h; && (5.26)\\
Cov(\tilde{\mathcal{T}}_g(\mathcal{C}(nI_{g_0})),\tilde{\mathcal{T}}_h(\mathcal{C}(nI_{g_0}))) &= \begin{cases} \frac{g_0-1}{g_0}\left\{F+\frac{1}{n}E\right\} & \text{if } g=h;\\ -\frac{1}{g_0}\left\{F+\frac{1}{n}E\right\} & \text{if } g\neq h.\end{cases} && (5.27)
\end{aligned}$$

By making use of equation (5.24) and the specifications given in equations (5.8) - (5.12), we may derive the following lemma.

Lemma 28 *The second-order relationships between $\tilde{\mathcal{M}}(\mathcal{C}.)$, the $\tilde{\mathcal{R}}_g(\mathcal{C})$s, $\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C})$ and the $\tilde{\mathcal{T}}_g(\mathcal{C}(nI_{g0}))$s may be expressed as*

$$\begin{aligned}
Cov(\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C}), \tilde{\mathcal{M}}(\mathcal{C}.)) &= P + \frac{1}{g_0}F; & (5.28)\\
Cov(\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C}), \tilde{\mathcal{R}}_h(\mathcal{C})) &= 0\ \forall h; & (5.29)\\
Cov(\tilde{\mathcal{T}}_g(\mathcal{C}(nI_{g0})), \tilde{\mathcal{M}}(\mathcal{C}.)) &= 0\ \forall g; & (5.30)\\
Cov(\tilde{\mathcal{T}}_g(\mathcal{C}(nI_{g0})), \tilde{\mathcal{R}}_h(\mathcal{C})) &= \begin{cases} \frac{g_0-1}{g_0}F & \text{if } g = h;\\ -\frac{1}{g_0}F & \text{if } g \neq h. \end{cases} & (5.31)
\end{aligned}$$

The orthogonalities given by equations (5.26), (5.29) and (5.30) immediately enable us to determine the following theorem.

Theorem 34 *The collection of measurements $\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C})$ is Bayes linear sufficient for $\mathcal{C}(nI_{g0})$ for adjusting the collection $\tilde{\mathcal{M}}(\mathcal{C}.)$.*

The collection of measurements $\tilde{\mathcal{T}}(\mathcal{C}(nI_{g0})) = \{\tilde{\mathcal{T}}_1(\mathcal{C}(nI_{g0})), \dots, \tilde{\mathcal{T}}_{g0}(\mathcal{C}(nI_{g0}))\}$ is Bayes linear sufficient for $\mathcal{C}(nI_{g0})$ for adjusting the collection $\tilde{\mathcal{R}}(\mathcal{C}) = \{\tilde{\mathcal{R}}_1(\mathcal{C}), \dots, \tilde{\mathcal{R}}_{g0}(\mathcal{C})\}$.

5.3 Learning about the overall population collection

Suppose that the matrix $\tilde{U} = [\tilde{U}_1 \dots \tilde{U}_{v0}]$ solves the generalised eigenvalue problem

$$(g_0P + F)\tilde{U} = (g_0P + F + E)\tilde{U}\tilde{\Lambda}, \qquad (5.32)$$

where $\tilde{\Lambda} = diag(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{v0})$ is the matrix of eigenvalues. $\tilde{U}$ is chosen so that $\tilde{U}^T(P+(1/g_0)F)\tilde{U} = I_{v0}$, $\tilde{U}^T\{P+(1/g_0)(F+E)\}\tilde{U}\tilde{\Lambda} = I_{v0}$. The resolution transform for the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C}.)]$ by $\mathcal{C}(nI_{g0})$ is denoted by $T_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]}$. We have the following theorem.

Theorem 35 *For the adjustment of the overall population collection, $[\tilde{\mathcal{M}}(\mathcal{C}.)]$, by the balanced sample, $\mathcal{C}(nI_{g0})$, the resolution transform matrix is calculated as*

$$T_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]} = \{g_0P + F + (1/n)E\}^{-1}\{g_0P + F\}. \qquad (5.33)$$

The overall population collection canonical directions for the balanced design are the same for each n and are given by

$$\tilde{Z}_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]s} = \tilde{U}_s^T \tilde{\mathcal{M}}(\mathcal{C}.), \tag{5.34}$$

for each $s = 1, \ldots, v_0$. *The corresponding overall population collection canonical resolutions are given by*

$$\tilde{\lambda}_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]s} = \frac{n\tilde{\lambda}_s}{(n-1)\tilde{\lambda}_s + 1}. \tag{5.35}$$

Proof - By the Bayes linear sufficiency of $\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C})$ for $\mathcal{C}(nI_{g0})$, we have that the resolution transform matrix, $T_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]} = T_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C})]}$, where $T_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C})]}$ is the resolution transform for the adjustment of $[\tilde{\mathcal{M}}(\mathcal{C}.)]$ by $\tilde{\mathcal{S}}_{nI_{g0}}(\mathcal{C})$. By substituting equations (5.8), (5.25) and (5.28) into equation (1.74), we obtain equation (5.33). Equations (5.34) and (5.35) thus follow immediately from equation (5.32). □

5.4 Learning about the gth group residual

We wish to investigate the adjustment of $[\tilde{\mathcal{R}}(\mathcal{C})]$ by $\mathcal{C}(nI_{g0})$. From Theorem 34, we have that the collection $\tilde{\mathcal{T}}(\mathcal{C}(nI_{g0}))$ is Bayes linear sufficient for $\mathcal{C}(nI_{g0})$ for this adjustment. Thus, to proceed, all we require is the second-order specifications for the collections $\tilde{\mathcal{R}}(\mathcal{C})$ and $\tilde{\mathcal{T}}(\mathcal{C}(nI_{g0}))$. We collect equations (5.10), (5.27) and (5.31) together as the following lemma.

Lemma 29 *The second-order specifications for the group residuals,* $\tilde{\mathcal{R}}(\mathcal{C})$, *and the sample mean group residuals,* $\tilde{\mathcal{T}}(\mathcal{C}(nI_{g0}))$, *may be expressed as*

$$Var(\tilde{\mathcal{R}}(\mathcal{C})) = (I_{g_0} - (1/g_0)J_{g_0}) \otimes F; \tag{5.36}$$

$$Cov(\tilde{\mathcal{R}}(\mathcal{C}), \tilde{\mathcal{T}}(\mathcal{C}(nI_{g_0}))) = (I_{g_0} - (1/g_0)J_{g_0}) \otimes F; \tag{5.37}$$

$$Var(\tilde{\mathcal{T}}(\mathcal{C}(nI_{g_0}))) = (I_{g_0} - (1/g_0)J_{g_0}) \otimes (F + (1/n)E). \tag{5.38}$$

Compare this lemma with Lemma 6 of Chapter 2. In terms of the adjustment, $\tilde{\mathcal{R}}(\mathcal{C})$ fills the role of $\mathcal{M}(\mathcal{C})$ and $\tilde{\mathcal{T}}(\mathcal{C}(nI_{g0}))$ fills the role of $\mathcal{S}_N(\mathcal{C})$. Notice how for both adjustments, $\tilde{\mathcal{T}}(\mathcal{C}(nI_{g0}))$ and $\mathcal{S}_N(\mathcal{C})$ are both Bayes linear sufficient for the respective adjustment by the full sample. The specifications in Lemma 29 do differ slightly though from those in Lemma 6.

Firstly, since the collection $\tilde{\mathcal{R}}(\mathcal{C})$ is not linearly independent then $(I_{g_0} - (1/g_0)J_{g_0})$ is not of full rank. Thus, for any resolution transform involving this quantity, there will be an infinite number of matrix representations of that transform, generated via equation (20) of Goldstein & Wooff (1998). In this case, equation (1.74) of this thesis should be re-expressed as equation (20) of Goldstein & Wooff (1998), namely

$$\begin{aligned} T_{\mathcal{D}} &= \{Var(\mathcal{B})\}^{\dagger}Cov(\mathcal{B},\mathcal{D})\{Var(\mathcal{D})\}^{\dagger}Cov(\mathcal{D},\mathcal{B}) + \\ &\qquad [I_p - \{Var(\mathcal{B})\}^{\dagger}Var(\mathcal{B})]T^0, \qquad (5.39) \end{aligned}$$

where $A^{\dagger}$ represents the Moore-Penrose generalised inverse of A, see Penrose (1955), and T^0 is an arbitrary $p \times p$ matrix. The natural choice, and indeed the choice we make in this thesis, is that $T^0 = 0$. Notice that since

$$(I_{g_0} - (1/g_0)J_{g_0})^2 = (I_{g_0} - (1/g_0)J_{g_0}), \qquad (5.40)$$

then we have the following Moore-Penrose inverses

$$\begin{aligned} (I_{g_0} - (1/g_0)J_{g_0})^{\dagger} &= (I_{g_0} - (1/g_0)J_{g_0}); & (5.41) \\ \{(I_{g_0} - (1/g_0)J_{g_0}) \otimes F\}^{\dagger} &= (I_{g_0} - (1/g_0)J_{g_0}) \otimes F^{-1}; & (5.42) \\ \{(I_{g_0} - (1/g_0)J_{g_0}) \otimes (F + (1/n)E)\}^{\dagger} &= (I_{g_0} - (1/g_0)J_{g_0}) \otimes & (5.43) \\ & \qquad (F + (1/n)E)^{-1}. & (5.44) \end{aligned}$$

These Moore-Penrose inverses will allow us to proceed with representations of all the resolution transforms of interest for this problem. As we have commented already, the corresponding canonical directions and resolutions are found over the linear span of the columns of the matrices of interest.

Secondly, in terms of a direct parallel, we may take $N = I_{g_0}$ (since we have a balanced design, we observe from Corollary 4 that we may solve the group problem as if we obtained a sample of size I_{g_0}; the sample size could then be displayed through the variable problem) and then $B = (I_{g_0} - (1/g_0)J_{g_0})$. However, $(I_{g_0} - (1/g_0)J_{g_0})$ is not diagonal. Again, this problem is surmountable. The results developed in Chapter 2 does not explicitly require B to be diagonal and the results all hold for any choice of B.

Thus, we may view the adjustment of $[\tilde{\mathcal{R}}(\mathcal{C})]$ by $\mathcal{C}(nI_{g_0})$ to be completely analogous to the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_{g_0})$ in Chapter 2; we solve a variable problem and a group problem and combine the two to obtain the full adjustment.

We commence with the variable problem, we make the following definition which is analogous to Definition 12.

Definition 21 *The underlying group residual canonical variable directions are defined as the columns of the matrix $Y = [Y_1 \dots Y_{v_0}]$ solving the generalised eigenvalue* *problem*

$$FY = (F+E)Y\Phi, \tag{5.45}$$

where $\Phi = diag(\phi_1, \dots, \phi_{v0})$ *is the matrix of eigenvalues. Y is chosen so that* $Y^T FY = I_{v_0}$, $Y^T(F+E)Y\Phi = I_{v_0}$. *The ordered eigenvalues* $1 > \phi_1 \geq \dots \geq \phi_{v_0} > 0$ *are termed the underlying group residual canonical variable resolutions.*

This definition is thus motivated by considering adjusting the gth group residual collection, $\tilde{\mathcal{R}}_g(\mathcal{C})$ by a sample of size n drawn from that group. Thus, we are interested in the adjustment of $[\tilde{\mathcal{R}}_g(\mathcal{C})]$ by $\mathcal{C}_g(n)$. Denoting the resolution transform for the adjustment of $[\tilde{\mathcal{R}}_g(\mathcal{C})]$ by $\mathcal{C}_g(n)$ as $T_{[\tilde{\mathcal{R}}_g(\mathcal{C})/\mathcal{C}_g(n)]}$, we have the following corollary, which follows immediately from Theorem 11.

Corollary 19 *The resolution transform matrix,* $T_{[\tilde{\mathcal{R}}_g(\mathcal{C})/\mathcal{C}_g(n)]}$, *is calculated as*

$$T_{[\tilde{\mathcal{R}}_g(\mathcal{C})/\mathcal{C}_g(n)]} = \{F + (1/n)E\}^{-1}F. \tag{5.46}$$

For each $s = 1, \dots, v_0$, *the canonical directions for the adjustment are given by*

$$Z_{[\tilde{\mathcal{R}}_g(\mathcal{C})/\mathcal{C}_g(n)]s} = \sqrt{\frac{g_0}{1-g_0}} Y_s^T \tilde{\mathcal{R}}_g(\mathcal{C}), \tag{5.47}$$

with the corresponding canonical resolutions given by

$$\lambda_{[\tilde{\mathcal{R}}_g(\mathcal{C})/\mathcal{C}_g(n)]s} = \frac{n\phi_s}{(n-1)\phi_s + 1}. \tag{5.48}$$

This completes our underlying variable work. We now move on to the underlying group problem. We make the following definition, which is analogous to Definition 13.

Definition 22 *The underlying group residual canonical group directions are defined as the columns of the matrix* $H = [H_1 \dots H_{g_0}]$ *solving the eigenvalue problem*

$$(I_{g_0} - (1/g_0)J_{g_0})H = H\Psi, \tag{5.49}$$

where $\Psi = diag(\psi_1, \dots, \psi_{g_0})$. *We choose*

$$H_d = \frac{1}{\sqrt{d(d+1)}}[1_d^T \; -d \; 0 \dots 0]^T, \tag{5.50}$$

for each $d = 1, \dots, g_0-1$, *so that if* $d, d' \in \{1, \dots, g_0-1\}$ *then* $H_d^T(I_{g_0} - (1/g_0)J_{g_0})H_{d'} = \delta_{dd'}$. $H_{g_0} = (1/g_0)1_{g_0}^T$. *For each* $d = 1, \dots, g_0 - 1$, $\phi_d = 1$ *and* $\phi_{g_0} = 0$. *The* ϕ_d*s and* ϕ_0 *are the underlying group residual canonical group resolutions.*

$H^{g_0} = [\sqrt{g_0}H_{g_0}\ H_1 \dots H_{g_0-1}]$ is the transpose of the $g_0 \times g_0$ Helmert matrix; see, for example, Searle (1982; p71). The motivation for the definition is thus that for each $s = 1, \dots, v_0$, we form the collection

$$\mathcal{Z}_{(nI_{g_0})s} \quad = \quad \{Z_{[\tilde{\mathcal{R}}_1(\mathcal{C})/\mathcal{C}_1(n)]s}, \dots, Z_{[\tilde{\mathcal{R}}_{g_0}(\mathcal{C})/\mathcal{C}_{g_0}(n)]s}\}. \tag{5.51}$$

Using our usual vector notation, we may also represent $\mathcal{Z}_{(nI_{g_0})s}$ as the $g_0 \times 1$ vector

$$\mathcal{Z}_{(nI_{g_0})s} = \sqrt{\frac{g_0}{1-g_0}}(I_{g_0} \otimes Y_s^T)\tilde{\mathcal{R}}(\mathcal{C}). \tag{5.52}$$

Notice that of the g_0 elements forming $\mathcal{Z}_{(nI_{g_0})s}$ only $g_0 - 1$ of them are linearly independent. Thus, we expect to find at most $g_0 - 1$ non-zero canonical resolutions for the adjustment of $[\tilde{\mathcal{R}}(\mathcal{C})]$ by $\mathcal{C}(nI_{g_0})$ which explains the presence of the zero underlying group residual canonical group resolution $\phi_{g_0} = 0$. For completeness, we have the following lemma.

Lemma 30 *The second-order relationships between $\mathcal{Z}_{(nI_{g_0})s}$ and $\tilde{\mathcal{T}}(\mathcal{C}(nI_{g_0}))$ may be expressed as follows*

$$\begin{aligned}
Var(\mathcal{Z}_{(nI_{g_0})s}) &= \frac{g_0}{g_0-1}\{I_{g_0} - (1/g_0)J_{g_0}\}; && (5.53)\\
Cov(\mathcal{Z}_{(nI_{g_0})s}, \tilde{\mathcal{T}}(\mathcal{C}(nI_{g_0}))) &= \sqrt{\frac{g_0}{1-g_0}}[\{I_{g_0} - (1/g_0)J_{g_0}\} \otimes Y_s^T F]; && (5.54)\\
Var(\tilde{\mathcal{T}}(\mathcal{C}(nI_{g_0}))) &= (I_{g_0} - (1/g_0)J_{g_0}) \otimes (F + (1/n)E). && (5.55)
\end{aligned}$$

Let $T_{[\mathcal{Z}_{(nI_{g_0})s}/\mathcal{C}(nI_{g_0})]}$ denote the resolution transform for the adjustment of $\mathcal{Z}_{(nI_{g_0})s}$ by $\mathcal{C}(nI_{g_0})$. We have the following corollary which follows from Theorem 12.

Corollary 20 *For the adjustment of $[\mathcal{Z}_{(nI_{g_0})s}]$ by $\mathcal{C}(nI_{g_0})$, the resolution transform matrix may be calculated as*

$$\begin{aligned}
T_{[\mathcal{Z}_{(nI_{g_0})s}/\mathcal{C}(nI_{g_0})]} &= \{I_{g_0} - (1/g_0)J_{g_0}\}^\dagger\{(I_{g_0} - (1/g_0)J_{g_0}) \otimes Y_s^T F\}\times \\
&\quad \{(I_{g_0} - (1/g_0)J_{g_0}) \otimes (F + (1/n)E)\}^\dagger\{(I_{g_0} - (1/g_0)J_{g_0}) \otimes F Y_s^T\} && (5.56)\\
&= \lambda_{[\tilde{\mathcal{R}}_g(\mathcal{C})/\mathcal{C}_g(n)]s}\{I_{g_0} - (1/g_0)J_{g_0}\}. && (5.57)
\end{aligned}$$

The canonical directions are given by

$$Z_{(nI_{g_0})ds} \quad = \quad \sqrt{\frac{g_0-1}{g_0}}H_d^T\mathcal{Z}_{(nI_{g_0})s} \quad = \quad (H_d \otimes Y_s)^T\tilde{\mathcal{R}}(\mathcal{C}), \tag{5.58}$$

for each $d = 1, \dots, g_0 - 1$, with corresponding canonical resolutions given by

$$\lambda_{(nI_{g_0})s} \quad = \quad \frac{n\phi_s}{(n-1)\phi_s + 1}. \tag{5.59}$$

Hence, $\lambda_{(nI_{g_0})s}$ is a resolution with multiplicity $g_0 - 1$. H_d is the dth underlying group residual canonical group direction.

In order to achieve the representation of $T_{[\mathcal{Z}_{(nI_{g0})s}/\mathcal{C}(nI_{g0})]}$, we have made use of the Moore-Penrose inverses given in equations (5.41) and (5.44). Equation (5.59) should be compared to equation (2.147) (or, equivalently, equation (2.185) since we are dealing with the balanced design) with $\psi_d = 1$. Notice also that $H_{g0}^T \mathcal{Z}_{(nI_{g0})s} = 0$, a quantity with prior variance zero.

The canonical variable and group analysis completely determine the adjustment of the full collection $[\tilde{\mathcal{R}}(\mathcal{C})]$ by $\mathcal{C}(nI_{g_0})$. We let $T_{[\tilde{\mathcal{R}}(\mathcal{C})/\mathcal{C}(nI_{g0})]}$ denote the resolution transform for the adjustment. The following corollary follows immediately from Theorem 13.

Corollary 21 *The resolution transform matrix, $T_{[\tilde{\mathcal{R}}(\mathcal{C})/\mathcal{C}(nI_{g0})]}$, is calculated as*

$$
\begin{aligned}
T_{[\tilde{\mathcal{R}}(\mathcal{C})/\mathcal{C}(nI_{g0})]} &= \{(I_{g_0} - (1/g_0)J_{g_0}) \otimes F\}^{\dagger}\{(I_{g_0} - (1/g_0)J_{g_0}) \otimes F\}\times \\
&\quad \{(I_{g_0} - (1/g_0)J_{g_0}) \otimes (F + (1/n)E)\}^{\dagger}\{(I_{g_0} - (1/g_0)J_{g_0}) \otimes F\} \qquad (5.60)\\
&= (I_{g_0} - (1/g_0)J_{g_0}) \otimes (F + (1/n)E)^{-1}F. \qquad (5.61)
\end{aligned}
$$

The collection $Z_{(nI_{g0})} = \{Z_{(nI_{g0})ds}\}$ for $d = 1, \ldots, g_0 - 1$, $s = 1, \ldots, v_0$ are the canonical directions of the adjustment with the corresponding canonical resolution to $Z_{(nI_{g0})ds}$ being given by $\lambda_{(nI_{g0})s}$.

Notice that from the representation of $T_{[\tilde{\mathcal{R}}(\mathcal{C})/\mathcal{C}(nI_{g0})]}$, given by equation (5.61), the results about the canonical resolutions being the product of the eigenvalues of $(I_{g_0} - (1/g_0)J_{g_0})$ and $(F + (1/n)E)^{-1}$ are not surprising. The eigenvalues of $A \otimes B$, for general matrices A and B, are well known to be the product of eigenvalues of A with those of B, see Searle (1982; p266). The link with the work of Chapter 2 and the corresponding understanding thus provided in terms of understanding the canonical structure was less apparent.

5.5 Learning about the group means

We now show that the analysis of the overall mean collection and the analysis of the group residuals completely determine the adjustment of the full collection, $\mathcal{M}(\mathcal{C}) = \{\mathcal{M}(\mathcal{C}_1), \ldots, \mathcal{M}(\mathcal{C}_{g_0})\}$. We denote the resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{C}(nI_{g_0})$ as $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(nI_{g0})]}$. We have the following theorem.

Theorem 36 *The resolution transform matrix for the adjustment is calculated as*

$$\begin{aligned} T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(nI_{g0})]} &= [\{I_{g_0} \otimes (F + (1/n)E)\} + \{J_{g_0} \otimes P\}]^{-1} \times \\ &\qquad [\{I_{g_0} \otimes F\} + \{J_{g_0} \otimes P\}] \quad (5.62) \\ &= [\{I_{g_0} \otimes (F + (1/n)E)^{-1}F\} + \\ &\{J_{g_0} \otimes (F + (1/n)E)^{-1}P\{I_{v_0} - (F + g_0P + (1/n)E)^{-1}(F + g_0P)\}\}]. \quad (5.63) \end{aligned}$$

The canonical directions of the adjustment are given by

$$Z_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(nI_{g0})]ds} = \begin{cases} (H_d \otimes Y_s)^T \mathcal{M}(\mathcal{C}) & \text{for } d = 1, \ldots, g_0 - 1; \forall s; \\ (H_{g_0} \otimes \tilde{U}_s)^T \mathcal{M}(\mathcal{C}) & \text{for } d = g_0; \forall s \end{cases} \quad (5.64)$$

$$= \begin{cases} Z_{(nI_{g0})ds} & \text{for } d = 1, \ldots, g_0 - 1; \forall s; \\ Z_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]s} & \text{for } d = g_0; \forall s. \end{cases} \quad (5.65)$$

The corresponding canonical resolutions of the adjustment are given by

$$\lambda_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(nI_{g0})]ds} = \begin{cases} \lambda_{(nI_{g0})s} & \text{for } d = 1, \ldots, g_0 - 1; \forall s; \\ \lambda_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]s} & \text{for } d = g_0; \forall s. \end{cases} \quad (5.66)$$

Proof - By the Bayes linear sufficiency of $\mathcal{S}_{nI_{g0}}(\mathcal{C})$ for $\mathcal{C}(nI_{g0})$, we have that the resolution transform, $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(nI_{g0})]} = T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_{nI_{g0}}(\mathcal{C})]}$, where $T_{[\mathcal{M}(\mathcal{C})/\mathcal{S}_{nI_{g0}}(\mathcal{C})]}$ is the resolution transform for the adjustment of $[\mathcal{M}(\mathcal{C})]$ by $\mathcal{S}_{nI_{g0}}(\mathcal{C})$. By substituting equations (5.19) - (5.21) into equation (1.74), we obtain equation (5.62). Equation (5.63) follows by using the inverse of $Var(\mathcal{S}_{nI_{g0}}(\mathcal{C}))$ given in equation (5.22).

Noting that $J_{g_0}H_d = 0$ for each $d = 1, \ldots, g_0 - 1$, it follows immediately that $H_d \otimes Y_s$ is an eigenvector of $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(nI_{g0})]}$, with eigenvalue $\lambda_{(nI_{g0})s}$ for $d = 1, \ldots, g_0 - 1$; $\forall s$. We then observe that

$$\begin{aligned} (H_d \otimes Y_s)^T \tilde{\mathcal{R}}(\mathcal{C}) &= \frac{1}{\sqrt{d(d+1)}} Y_s^T \left\{ \sum_{g=1}^{d} \tilde{\mathcal{R}}_g(\mathcal{C}) - d\tilde{\mathcal{R}}_{d+1}(\mathcal{C}) \right\} && (5.67) \\ &= \frac{1}{\sqrt{d(d+1)}} Y_s^T \left\{ \sum_{g=1}^{d} \mathcal{M}(\mathcal{C}_g) - d\mathcal{M}(\mathcal{C}_{d+1}) \right\} && (5.68) \\ &= (H_d \otimes Y_s)^T \mathcal{M}(\mathcal{C}), && (5.69) \end{aligned}$$

so that $(H_d \otimes Y_s)^T \mathcal{M}(\mathcal{C}) = Z_{(nI_{g0})ds}$.

Simple algebra shows that for each $s = 1, \ldots, v_0$, $H_{g_0} \otimes \tilde{U}_s$ is an eigenvector of $T_{[\mathcal{M}(\mathcal{C})/\mathcal{C}(nI_{g0})]}$, with eigenvalue $\lambda_{[\tilde{\mathcal{M}}(\mathcal{C}.)/\mathcal{C}(nI_{g0})]s}$ and the result follows since $\tilde{\mathcal{M}}(\mathcal{C}.) = H_{g_0}^T \mathcal{M}(\mathcal{C})$. □

5.6 Comments and conclusions

Thus, our lower level analysis allows us to completely determine the analysis of the higher level analysis; we may consider separately the adjustment of the components parts of $\mathcal{M}(\mathcal{C})$, $\tilde{\mathcal{M}}(\mathcal{C}.)$ and $\tilde{\mathcal{R}}(\mathcal{C})$ and the union of these canonical components yield the canonical components for the adjustment of $\mathcal{M}(\mathcal{C})$. Once more, we have the breakdown of the problem into interpretable subproblems. This result is not surprising for the second-order exchangeable statements we made for this problem produce an invariant covariance matrix for the relationships between individuals,

$$Cov(X_{gi}, X_{hj}) = \begin{cases} P+F+E & \text{if } g=h,\ i=j; \\ P+F & \text{if } g=h,\ i\neq j; \\ P & \text{if } g\neq h. \end{cases} \tag{5.70}$$

So our modelling is formally analogous to models considered in Consonni & Dawid (1985) and Dawid (1988). We could consider our example as being the multivariate version of Example 1 of Consonni & Dawid (1985). In this multivariate version, we have $\Gamma_0 = P + F + E$, $\Gamma_1 = P + F$ and $\Gamma_2 = P$, where Γ_0, Γ_1, Γ_2 are the respective matrix versions of γ_0, γ_1, γ_2, as given by Consonni & Dawid (1985; p632). Alternatively, we could view our specification as analogous to the model $(\underline{M} * W)/\underline{R}$ of Dawid (1988; p6) so that to compare our modelling with his, the groups are the machines, the variables are the workers, and the individuals are the runs. It may be shown, in an analogous way to the work here, that if we consider more levels in our model, and maintain an invariant covariance matrix which is a judgement of hierarchical co-exchangeability (we need not consider the variables to be second-order exchangeable), then for balanced design, the adjustment of the mean components corresponding to any level may be performed by solving the adjustment for all the components on the lower levels.

The multi-way layouts discussed by Goldstein & Wooff (1997) and the construction of the canonical structure for the mean components may also be viewed as deriving from the solution of a series of interpretable subproblems. For example in the two-way layout, see Section 6 of Goldstein & Wooff (1997), solving the overall mean problem yields equation (15), the column problem yields equation (16) and the row problem yields equation (17), which then combine together to yield the whole solution.

Chapter 6

Conclusions, suggestions for further work

In the introduction to the work in this thesis, given in Chapter 1, we outlined the reasons behind our viewpoint that the revision of belief should be viewed from a subjective standpoint and that the most widely used methodology for organising and analysing our subjective beliefs, namely the familiar Bayesian approach, is handicapped by our inability to both envisage and quantify every possible outcome. As such, we advocate a system of partial belief adjustment based upon the revision of prevision; the system is frequently known as the Bayes linear approach.

Throughout this thesis, we have used the canonical resolutions and directions as the central focus for understanding, both qualitatively and quantitatively, the information we expect to gain from the adjustment of a belief structure, $[\mathcal{B}]$, following the observation of a data collection, $\mathcal{D}$. The canonical structure may be calculated from the eigenstructure of the resolution transform matrix; the eigenvectors corresponding to the directions and the eigenvalues to the resolutions. The reason for our interest in them is that the collection $Z = \{Z_1, Z_2, \dots\}$ of canonical resolutions form an orthogonal grid of directions which summarise the effects of the adjustment. The Z_i form a basis for $[\mathcal{B}]$ and so the prior variance of any $\mathcal{A} \in [\mathcal{B}]$ may be expressed as a linear combination of the prior variances of the Z_i, but crucially, as equation (1.62) shows, the adjusted variance of any $\mathcal{A} \in [\mathcal{B}]$ may be expressed as a linear combination of the adjusted variances of the Z_i and hence the resolutions also; see equation (1.63). Constrained by being uncorrelated with $(Z_1, \dots, Z_j)$, Z_{j+1} is the element of $[\mathcal{B}]$ maximising the resolution. Thus, the grid summarises the qualitative information we expect to receive, with the canonical resolutions providing the quantitative information. By comparing the canonical structure for different data collections, we may begin the process of design, for we may understand the differences between different data collections in the types of information we expect to learn and thus can begin to compare the benefits of each collection weighed against the cost.

The work of Goldstein & Wooff (1998) elegantly shows this in operation for the adjustment of infinitely second-order exchangeable sequences: for learning about the mean components, the canonical directions are the same for all sample sizes and so our qualitative learning in unaffected by sample size; only the quantitative information is dependent upon the sample size. The same statements apply to the prediction of future individuals. It is this beauty and level of understanding that we sought to build upon in Chapter 2 when we widened the infinite second-order exchangeability to encompass a form of co-exchangeability.

We introduced the co-exchangeability through having second-order exchangeable sequences labelled as a group. In each group, the sequence measured the same set of variables and our specification, as given by equations (2.67) - (2.69), was such that we could separate the specification as the product of a group part and a variable part. The consequence of this was that the adjustment of the mean components of each second-order exchangeable sequence by a sample drawn from that group had the same quantitative form provided by the underlying canonical variable problem as given by Definition 12. We may then collect the analogous directions from each group together to form v_0 orthogonal collections and adjust each of these collections by the complete data collection. The adjustment of each of these collections had the same quantitative form provided by the underlying canonical group problem, given by Definition 13. The collection of all the canonical directions and resolutions from these v_0 g_0-dimensional collections formed the canonical structure for the full adjustment of the mean components.

Notice how we achieved our three aims given in Chapter 1. Adoption of a Bayes linear approach enables us to work with a second-order specification, whilst our scaled approach across groups reduces the specification further. In the most general co-exchangeable setting, we are free, up to maintaining coherency, to specify the $(1/2)g_0(g_0+1)$ $v_0 \times v_0$ matrices $Cov(\mathcal{M}(\mathcal{C}_g), \mathcal{M}(\mathcal{C}_h))$ as we see fit: a total of $(1/4)g_0(g_0+1)v_0(v_0+1)$ covariances. Similarly, specifying $Var(\mathcal{R}_i(\mathcal{C}_g))$ requires a further $(1/2)g_0v_0(v_0+1)$ covariance statements. The specifications we adopted required only the specifications of the $v_0 \times v_0$ symmetric matrices C and E, the $g_0 \times g_0$ symmetric matrix A and the $g_0 \times g_0$ diagonal matrix B. Of course, we choose to work with this model when it does, to the level of specification we are willing to make, conform with our beliefs and we would not seek to forcibly impose it. The second aim is accomplished through the reduction in the dimension of problems we have to solve. The $g_0 v_0 \times g_0 v_0$ problem is reduced to one of size $v_0 \times v_0$ and a second of size $g_0 \times g_0$. Moreover, only the latter problem is changed when we consider different samples. As both of these problems are interpretable then we gain insights into how changing features of the design will effect the qualitative information provided by the canonical directions. Thus, we reach the third aim. As we suggested in Chapter 2, future work could be made in viewing the subproblems as perturbation problems and looking at the rotations of the canonical directions as a consequence of these perturbations.

One illuminating method of assessing the change in quantitative information provided by different designs and specifications could be via the use of Gerschgorin's theorem, Gerschgorin (1931), which states that the eigenvalues of a matrix lie in the union of certain disks in the complex plane. For applications to perturbation theory, see Stewart & Sun (1990; p180-187).

We concluded Chapter 2 by examining the prediction of future individuals, both from the previously observed groups and also from unobserved groups. Although, in general, qualitatively different, the predictive problem follows the same breakdown into interpretable problems as for the adjustment of the mean components which aids our understanding of the design process. In Chapter 3 we took the specifications to their natural limit by considering that the group structure could be broken down into a series of factors, and we showed that we could interpret the design as one of solving a series of subproblems related to the individual factors, as well as the same variable problem. This allowed us to consider both slicing and marginalisation of the factor settings, and these problems could be handled without the need to solve additional problems.

Having reached the end of Chapter 3, one is faced with the question of 'where do we go from here?'. As ever, we check back through our modelling and begin to consider whether the model has weaknesses, areas where we have not truly reflected our beliefs and made pragmatic simplifications. One area of attention that we are aware of, and would like to know the consequences of, is the use of infinite second-order exchangeability in the model; a convenient approximation to the reality of only finite sequences.

Examining the adjustment of finitely second-order exchangeable sequences also opens up the scope of our analysis, for we may now handle sequences which have no notion of being potentially infinite in length. We commence Chapter 4 by looking at the adjustment of finite second-order exchangeable sequences in the same way as Goldstein & Wooff (1998) considered the adjustment of infinite second-order exchangeable sequences. We showed that the adjustment of such sequences exhibited the same coherence properties as for the infinite case, in that the resolution transform for the population structure induced by the finite second-order exchangeable sequence has the same form whatever the sample size; the canonical directions are the same for each sample size, with simple modifications for the canonical resolutions. We then explore the comparisons between sequences where only the lengths change and show that, reassuringly so that our qualitative understanding remains unaffected by sequence length, the canonical resolutions of each sequence share, up to a scale factor, the same co-ordinate representation. We may view infinite second-order exchangeability as being the infinite extension of a finitely second-order sequence and our results mean that we can view the adjustment, in terms of the canonical structure, of a infinitely second-order sequence as being the limit of the finite adjustment as the sequence length goes to infinity. Thus, the essential difference between the adjustment of the finite sequence and the infinite sequence is not qualitative but quantitative; the finite canonical resolutions require a simple finite population correction term.

We then moved on to consider our co-exchangeable model from the finite perspective. In this case, we may still separate the problem into problems over the variables and over the groups; in the finite case we get a different group problem for each direction of the variable problem. This differs from the infinite case and so there are qualitative differences between the finite and infinite cases in that the finite case causes rotations of the canonical resolutions of the infinite case. However, it is easy to see that as we become closer to the infinite case in all of the groups, these rotations get smaller and so the adjustment may still be viewed as the limit of the finite adjustment.

This reaches the end of the core work in this thesis. However, it does not reach the end of our investigations. We examined one particular type of co-exchangeability; there are many other avenues we could take before reaching the most general form of co-exchangeability and it may be profitable to examine these to see what insights they provide us with. One such development is that discussed in Chapter 5 where we judged second-order exchangeability over the groups, as well as the individuals. This left us with a one-level hierarchical co-exchangeable model and in the balanced design with no group dependence for the residuals, the analysis broke down once more into a series of interpretable subproblems over the components of the model.

The types of problems discussed in this thesis, and those outlined in Section 5.6, all have one thing in common, and that is the presence of invariant subspaces. In the theory developed in Chapter 2, for example, it is straightforward to see that the adjustment of $[\mathcal{M}(\mathcal{C}_g)]$ by $\mathcal{C}(N)$ has the same qualitative form for each g and is the same as that for the adjustment of $[\mathcal{M}(\mathcal{C}_g)]$ by $\mathcal{C}(n_g)$. It is, perhaps, this avenue that we should explore: that of an exchangeability of the canonical directions across the second-order exchangeable sequences that may provide the most illuminating use of co-exchangeable second-order exchangeable sequences.

Appendix A

Some useful notation

A.1 Matrices and sets

- I_p denotes the $p \times p$ identity matrix. It has the column representation $I_p = [\epsilon_{p1} \dots \epsilon_{pp}]$.
- J_{pq} denotes the $p \times q$ matrix of ones. It has the column representation $J_{pq} = [1_{p1} \dots 1_{pq}]$.
- $(A)_{pq}$ denotes the (p,q)th element of the matrix A.
- $\otimes_{i=1}^{k} A_i = A_k \otimes A_{k-1} \otimes \cdots \otimes A_1$.
- $\oplus_{i=1}^{k} A_i = A_1 \oplus A_2 \oplus \cdots \oplus A_k$.
- $\Delta = \{1, \dots, k\}$ denotes a finite set of classification criteria, or factors.
- $\mathcal{I}_q = \{1, \dots, l_{q,0}\}$ denotes a finite set of possible levels of the qth factor.
- $\mathcal{I} = \times_{q \in \Delta} \mathcal{I}$ denotes the available cells.
- $l_{[k]}$ denotes the cell $(l_k, \dots, l_1)$.
- $l_{[k;q+1]} l_q l_{[q-1;1]}$ denotes the cell $(l_k, \dots, l_{q+1}, l_q, l_{q-1}, \dots, l_1)$, and so makes the level of the qth factor explicit.

A.2 Collections of random quantities

- $\mathcal{C} = \{X_1, X_2, \dots\}$, finite or infinite, is a series of measurements to be made upon a sequence of individuals.

- $\mathcal{C}_i = \{X_{1i}, X_{2i}, \dots\}$ denotes the measurements for the ith individual.

- $\mathcal{C}^*$ denotes the full collection of measurements for all of the individuals. In a single sequence, it is the union of all of the elements in all of the collections $\mathcal{C}_i$.

- $\mathcal{C}_g = \{X_{g1}, X_{g2}, \dots\}$, finite or infinite, is a series of measurements to be made upon a sequence of individuals, where each individual may be classified as belonging to the gth system (typically a group).

- $\mathcal{C}_{gi} = \{X_{g1i}, X_{g2i}, \dots\}$ denotes the measurements for the ith individual in the gth system.

- $\mathcal{C}_g^*$ denotes the full collection of measurements for all of the individuals in the gth system.

- $\mathcal{C}(n) = \cup_{i=1}^{n} \mathcal{C}_i$, the union of the first n individuals' collections.

- $\mathcal{C}_g(n_g) = \cup_{i=1}^{n_g} \mathcal{C}_{gi}$, the union of the first n_g individuals' collections in the gth system.

- $\mathcal{C}(N) = \cup_{g \in G} \mathcal{C}_g(n_g)$, the union of systems g, belonging to a set G, of the union of the first n_g individuals' collections in the gth system.

- $\mathcal{C}(n; r) = \cup_{i=n+1}^{n+r} \mathcal{C}_i$, the union of the $(n+1)$st to $(n+r)$th individuals' collections.

- $\mathcal{C}_g(n_g; r_g) = \cup_{i=n_g+1}^{n_g+r_g} \mathcal{C}_{gi}$, the union of the (n_g+1)st to (n_g+r_g)th individuals' collections in the gth system.

- $\mathcal{C}(N; R) = \cup_{g \in G} \mathcal{C}_g(n_g; r_g)$, the union of systems g, belonging to a set G, of the union of the (n_g+1)st to (n_g+r_g)th individuals' collections in the gth system.

- $\mathcal{C}(R/N) = \cup_{h \in H} \mathcal{C}_h(r_h)$, the union of systems h, belonging to a set H, of the union of the first r_h individuals' collections in the hth system. We impose the constraint that $\mathcal{C}(R/N) \cap \mathcal{C}(N) = \emptyset$.

- $\mathcal{M}(\mathcal{C})$ denotes the total collection of underlying population mean quantities relating to systems of infinitely second-order exchangeable sequences. If we have a single system, we may write $\mathcal{M}(\mathcal{C}) = \{\mathcal{M}(X_1), \mathcal{M}(X_2), \dots\}$. If we have a set, G, of systems then we denote by $\mathcal{M}(\mathcal{C}_g) = \{\mathcal{M}(X_{g1}), \mathcal{M}(X_{g2}), \dots\}$ the total collection of underlying population mean quantities for the gth system and then we have, $\mathcal{M}(\mathcal{C}) = \cup_{g\in G}\mathcal{M}(\mathcal{C}_g)$.

- $\tilde{\mathcal{M}}(\mathcal{C})$ denotes the total collection of underlying population mean quantities relating to systems of finitely second-order exchangeable sequences. If we have a single system, of length m, we write $\tilde{\mathcal{M}}^{[m]}(\mathcal{C}) = \{\tilde{\mathcal{M}}^{[m]}(X_1), \tilde{\mathcal{M}}^{[m]}(X_2), \ldots\}$. If there is no confusion, we may omit the explicit reference to m, the sequence length. If we have a set, G, of systems then we denote by $\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g) = \{\tilde{\mathcal{M}}^{[m_g]}(X_{g1}), \tilde{\mathcal{M}}^{[m_g]}(X_{g2}), \ldots\}$ the total collection of underlying population mean quantities for the gth system, which contains m_g individuals, and $\tilde{\mathcal{M}}(\mathcal{C}) = \cup_{g\in G}\tilde{\mathcal{M}}^{[m_g]}(\mathcal{C}_g)$.

- $\mathcal{R}_i(\mathcal{C}) = \{\mathcal{R}_i(X_1), \mathcal{R}_i(X_2), \ldots\}$ denotes the collection of residuals for the ith individual in a single infinitely second-order exchangeable sequence. If we have a set, G, of infinitely second-order exchangeable sequences then $\mathcal{R}_i(\mathcal{C}_g) = \{\mathcal{R}_i(X_{g1}), \mathcal{R}_i(X_{g2}), \ldots\}$ denotes the collection of residuals for the ith individual in the gth sequence.

- $\tilde{\mathcal{R}}_i^{[m]}(\mathcal{C}) = \{\tilde{\mathcal{R}}_i^{[m]}(X_1), \tilde{\mathcal{R}}_i^{[m]}(X_2), \ldots\}$ denotes the collection of residuals for the ith individual in a single finitely second-order exchangeable sequence, of length m. If we have a set, G, of finitely second-order exchangeable sequences then $\tilde{\mathcal{R}}_i^{[m_g]}(\mathcal{C}_g) = \{\tilde{\mathcal{R}}_i^{[m_g]}(X_{g1}), \tilde{\mathcal{R}}_i^{[m_g]}(X_{g2}), \ldots\}$ denotes the collection of residuals for the ith individual in the gth sequence, which is of length m_g.

- $\mathcal{S}_n(\mathcal{C}) = \{\mathcal{S}_n(X_1), \mathcal{S}_n(X_2), \ldots\}$ denotes the collection of sample averages for the individuals in $\mathcal{C}(n)$.

- $\mathcal{S}_{n_g}(\mathcal{C}_g) = \{\mathcal{S}_{n_g}(X_{g1}), \mathcal{S}_{n_g}(X_{g2}), \ldots\}$ denotes the collection of sample averages for the individuals in $\mathcal{C}_g(n_g)$.

- $\mathcal{S}_N(\mathcal{C}) = \cup_{g\in G}\mathcal{S}_{n_g}(\mathcal{C}_g)$ denotes the collection of sample averages for the individuals in $\mathcal{C}(N)$.

- $\mathcal{T}_i(\mathcal{C}(n)) = \{\mathcal{T}_i(X_1), \mathcal{T}_i(X_2), \ldots\}$ denotes the collection of residuals from the sample mean for the ith individual in $\mathcal{C}(n)$.

- $\mathcal{T}_i(\mathcal{C}_g(n_g)) = \{\mathcal{T}_i(X_{g1}), \mathcal{T}_i(X_{g2}), \ldots\}$ denotes the collection of residuals from the sample mean for the ith individual in $\mathcal{C}_g(n_g)$.

Bibliography

Aldous, D. J. (1985). Partial exchangeability. In S. Kotz, N. L. Johnson, and C. B. Read (Eds.), *Encyclopedia of Statistical Sciences*, Volume 6, pp. 579–581. New York: Wiley.

Bayes, T. (1764). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc. 53*, 370–418. Reprinted in *Biometrika* **45** (1958), 293-315.

Bernardo, J. M. and A. F. M. Smith (1994). *Bayesian Theory.* Chichester: Wiley.

Box, J. F. (1978). *R. A. Fisher, The Life of a Scientist.* Chichester: Wiley.

Bruno, A. (1964). On the notion of partial exchangeability. *Giorn. Ist. It. Attuari 27*, 174–196. English translation in B. de Finetti, *Probability, Induction and Statistics*, pp. 229-249. London: Wiley, 1972.

Chaloner, K. and I. Verdinelli (1995). Bayesian Experimental Design: A Review. *Statistical Science 10*, 273–304.

Cifarelli, D. M. and E. Regazzini (1996). De Finetti's Contribution to Probability and Statistics. *Statistical Science 11*, 253–282.

Cochran, W. G. and G. M. Cox (1957). *Experimental Designs, 2nd edition.* New York: Wiley.

Consonni, G. and A. P. Dawid (1985). Invariant Normal Bayesian Linear Models and Experimental Designs. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 2*, pp. 629–644. Amsterdam: North-Holland.

Dale, A. I. (1985). A study of some early investigations of exchangeability. *Historia Mathematica 12*, 323–336.

Dawid, A. (1988). Symmetry models and hypotheses for structured data layouts (with discussion). *J.R.Statist. Soc. B 50*, 1–34.

de Finetti, B. (1931). Funzioni carattaristica di un fenomeno aleatorio. *Academia Nazionale dei Lincei 4*, 86–133.

de Finetti, B. (1937). La prévision, ses lois logiques, ses sources subjectives. *Annales de L'Institute Henri Poincaré 7*, 1–68. English translation in H. E. Kyburg, jr. and H. E. Smokler, editors, *Studies in Subjective Probability*, pp. 93-158. Wiley, New York, 1964.

de Finetti, B. (1938). Sur la condition d'equivalence partielle. *Actualities Scientifiques et Industrielles 739*, 5–18. English translation in R. Jeffrey, editor, *Studies in Inductive Logic and Probability II*, pp. 193-205. University of California Press, Berkeley, 1980.

de Finetti, B. (1959). La probabilità e la statistica nei rapporti con l'induzione, secondo i diversi punti di vista. In *Corse C.I.M.E su Induzione e Statistica.* Cremones. English translation in B. de Finetti, *Probability, Induction and Statistics*, pp. 147-227. Wiley, London, 1972.

de Finetti, B. (1972). *Probability, Induction and Statistics.* London: Wiley.

de Finetti, B. (1973). Bayesianism: its unifying role for both the foundations and the applications of statistics. *Bulletin of the International Statistical Institute 4*, 349–368. Also appears in B. de Finetti, *Probabilità e Induzione*, original articles and English translations, P. Monari and D. Cocchi (eds.), pp. 467-490. Biblioteca di Statistica, Editrice CLUEB, Bologna, 1993.

de Finetti, B. (1974). *Theory of Probability*, Volume 1. London: Wiley.

de Finetti, B. (1975). *Theory of Probability*, Volume 2. London: Wiley.

Deely, J. (1992). Comment on "Advances in Bayesian Experimental Design" by I. Verdinelli . In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 475–477. Oxford: University Press. see Verdinelli(1992).

Diaconis, P. (1977). Finite forms of de Finetti's theorem on exchangeability. *Synthese 36*, 271–281.

Diaconis, P. (1988). Recent progress on de Finetti's notion of exchangeability (with discussion). In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 111–125. Oxford: University Press.

Diaconis, P. and D. Freedman (1980a). de Finetti's generalizations of exchangeability. In R. A. Jeffrey (Ed.), *Studies in Inductive Logic and*

Probability II, pp. 233–249. Berkeley: University of California Press.

Diaconis, P. and D. Freedman (1980b). Finite exchangeable sequences. *Ann. Prob. 8*, 745–764.

Everitt, B. S. (1992). *The Analysis of Contingency Tables, 2nd edition.* London: Chapman & Hall.

Farrow, M. and M. Goldstein (1992). Reconciling Costs and Benefits in Experimental Design. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 607–615. Oxford: University Press.

Fishburn, P. C. (1986). The Axioms of Subjective Probability (with discussion). *Statistical Science 1*, 335–358.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.

Fisher, R. A. (1935). *The Design of Experiments.* Edinburgh: Oliver and Boyd.

Freedman, D. A. (1962). Invariants under mixing which generalise de Finetti's theorem. *Ann. Math. Statist. 33*, 916–923.

Freedman, D. A. (1963). Invariants under mixing which generalise de Finetti's theorem: continuous time parameter. *Ann. Math. Statist. 34*, 1194–1216.

Galambos, J. (1982). Exchangeability. In S. Kotz, N. L. Johnson, and C. B. Read (Eds.), *Encyclopedia of Statistical Sciences*, Volume 2, pp. 573–577. New York: Wiley.

Gerschgorin, S. A. (1931). Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk SSSR, Ser. Fiz.-Mat. 6*, 749–754.

Goel, P. K. and M. H. DeGroot (1980). Only Normal Distributions Have Linear Posterior Expectations in Linear Regression. *J. Amer. Statist. Ass. 75*, 895–900.

Goldstein, M. (1976). Bayesian analysis of regression problems. *Biometrika 63*, 51–58.

Goldstein, M. (1981). Revising previsions: a geometric interpretation (with discussion). *J. R. Statist. Soc. B 43*, 105–130.

Goldstein, M. (1983). The prevision of a prevision. *J. Amer. Statist. Ass. 78*, 817–819.

Goldstein, M. (1986a). Exchangeable belief structures. *J. Amer. Statist. Ass. 81*, 971–976.

Goldstein, M. (1986b). Separating beliefs. In P. K. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques*, pp. 197–215. Amsterdam: North Holland.

Goldstein, M. (1988a). Adjusting belief structures. *J. R. Statist. Soc B 50*, 133–154.

Goldstein, M. (1988b). The data trajectory. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics 3*, pp. 189–209. Oxford: University Press.

Goldstein, M. (1990). Influence and Belief Adjustment. In R. M. Oliver and J. Q. Smith (Eds.), *Influence Diagrams, Belief Nets and Decision Analysis*, pp. 143–174. Chichester: Wiley.

Goldstein, M. (1991). Belief transforms and the comparison of hypotheses. *Ann. Statist. 19*, 2067–2089.

Goldstein, M. (1992). Comment on "Is our view of Bayesian Statistics too Narrow?" by D. V. Lindley. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 8–9. Oxford: University Press.

Goldstein, M. (1994a). Belief revision: subjectivist principles and practice. In D. Prawitz and D. Westerstahl (Eds.), *Logic and Philosophy of Science in Uppsala*, pp. 117–130. Dordrecht: Kluwer Academic Publishers.

Goldstein, M. (1994b). Revising exchangeable beliefs: subjectivist foundations for the inductive argument. In P. Freeman and A. F. M. Smith (Eds.), *Aspects of Uncertainty: A Tribute to D. V. Lindley*, pp. 201–222. Chichester: Wiley.

Goldstein, M. (1997). Prior inferences for posterior judgements. In M. L. D. Chiara, K. Doets, D. Mundici, and J. van Benthem (Eds.), *Structures and Norms in Science, Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995*, pp. 55–71. Dordrecht: Kluwer Academic Publishers.

Goldstein, M. (1999). Bayes linear analysis. In S. Kotz, C. B. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences Update Volume 3*, pp. 29–34. Chichester: Wiley.

Goldstein, M. (2000). Bayes linear methods i - adjusting beliefs: concepts and properties. *Journal of statistical software 5*, 1–19. Also available at www.stat.ucla.edu/journals/jss/v05/i02.

Goldstein, M. and A. O'Hagan (1996). Bayes Linear Sufficiency and Systems of Expert Posterior Assessments. *J. R. Statist. Soc. B 58*, 301–316.

Goldstein, M. and D. A. Wooff (1994). Robustness measures for Bayes linear analyses. *Journal of Statistical Planning and Inference 40*, 261–277.

Goldstein, M. and D. A. Wooff (1997). Choosing sample sizes in balanced experimental designs: a bayes linear approach. *The Statistician 46*, 1–17.

Goldstein, M. and D. A. Wooff (1998). Adjusting exchangeable beliefs. *Biometrika 85*, 39–54.

Hartigan, J. A. (1969). Linear Bayesian methods. *J. R. Statist. Soc. B 31*, 446–454.

Kendall, M. G. and A. Stuart (1979). *The Advanced Theory of Statistics Volume 2, 4th edition.* High Wycombe: Griffin.

Kyburg, H. E. and H. E. Smokler (1964). *Studies in Subjective Probability.* New York: Wiley.

Kyburg, H. E. and H. E. Smokler (1980). *Studies in Subjective Probability, 2nd edition.* New York: Dover.

Lad, F. (1996). *Operational Subjective Statistical Methods.* New York: Wiley.

Lad, F. and J. Deely (1994). A subjective utilitarian view of experimental design. In P. Freeman and A. F. M. Smith (Eds.), *Aspects of Uncertainty: A Tribute to D.V. Lindley*, pp. 267–281. Chichester: Wiley.

Laplace, P. S. (1774). Mémoire sur la probabilité des causes par les évenements. *Mem. Acad. Sci. Paris 6*, 621–656.

Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statist. Sci. 1*, 359–378. English translation of Laplace (1774), with an introduction by S. M. Stigler.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford: Clarendon Press.

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint, 2 volumes.* Cambridge: University Press.

Lindley, D. V. (1972). *Bayesian statistics, a review.* Philadelphia: SIAM.

Mouchart, M. and L. Simar (1980). Least squares approximation in Bayesian analysis (with discussion). In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Eds.), *Bayesian Statistics,* pp. 207–222. Valencia: University Press. The discussion appears on pp. 237-245.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics : Bayesian Inference, Vol 2B.* London: Edward Arnold.

Penrose, R. A. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc. 51,* 406–413.

Press, S. J. (1989). *Bayesian Statistics.* New York: Wiley.

Raiffa, H. and R. Schlaifer (1961). *Applied statistical decision theory.* Boston: Division of Research, Graduate School of Business Adminitration, Harvard University.

Ramsey, F. P. (1926). Truth and probability. Unpublished during Ramsey's lifetime, the essay was first published in *The Foundations of Mathematics and other Logical Essays* by F. P. Ramsey. Edited by R. B. Braithwaite. London. Kegan Paul, Trench, Trubner & Co (1931), pp. 156-198. It is reprinted in Kyburg & Smokler (1964, 1980).

Savage, L. J. (1954). *The Foundations of Statistics.* New York: Wiley.

Savage, L. J. (1981). *The Writings of Leonard Jimmie Savage - A Memorial Selection.* Washington, DC: American Statistical Association.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics.* Chichester: Wiley.

Searle, S. R., G. Casella, and C. E. McCulloch (1992). *Variance Components.* Chichester: Wiley.

Shafer, G. (1986). Savage Revisited (with discussion). *Statistical Science 1,* 463–501.

Shaw, S. C. and M. Goldstein (1999). Simplifying Complex Designs: Bayes Linear Experimental Design for Grouped Multivariate Exchangeable Systems. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 6,* pp. 839–848. Oxford: University Press.

Stewart, G. W. and J.-g. Sun (1990). *Matrix perturbation theory.* San Diego: Academic Press.

Stone, M. (1963). Robustness of non-ideal decision procedures. *J. Amer. Statist. Ass. 58*, 480–486.

Toman, B. (1999). Bayesian experimental design. In S. Kotz, C. B. Read, and D. L. Banks (Eds.), *Encyclopedia of Statistical Sciences Update Volume 3*, pp. 35–39. Chichester: Wiley.

Verdinelli, I. (1992). Advances in Bayesian Experimental Design (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 467–481. Oxford: University Press.

von Plato, J. (1994). *Creating modern probability.* Cambridge: University Press.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* London: Chapman & Hall.

Whittle, P. (1992). *Probability via Expectation, 3rd edition.* New York: Springer-Verlag.

Wilkinson, D. J. (1995). *Bayes linear covariance matrix adjustment.* Ph. D. thesis, University of Durham, Durham, DH1 3LE, UK.

Wooff, D. A. (1992). [B/D] works. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics 4*, pp. 851–859. Oxford: University Press.

Wooff, D. A. and M. Goldstein (1994). Adjusting co-exchangeable beliefs. University of Durham Technical Report.