UNIVERSIDADE DA BEIRA INTERIOR

Error orthogonal models: Structure, Operations and Inference

Carla Maria Lopes da Silva Afonso dos Santos

Tese para obtenção do Grau de Doutor em

Matemática Aplicada

(3º ciclo de estudos)

Orientador: Prof. Doutor João Tiago Praça Nunes Mexia
Co-orientador: Prof. Doutora Célia Maria Pinto Nunes

Covilhã, Julho de 2012

Acknowledgements

First I would like to express my very special thanks to my supervisor, Professor João Tiago Mexia. His continuous support, encouragement, friendship and endless patience allowed me to overcome the many obstacles that came along this path.

I would also like to thank my co-supervisor Professor Célia Nunes who always supported me with kindness, encouragement and friendship and with whom I could always count despite the physical distance.

I am so fortunate to have two such wonderful persons as supervisors. It has been a very long journey and I could not have reached the goal without you two. It was an honour to work with you.

I would also like to acknowledge the huge amount of support and encouragement from my wonderful loving family. Especially to my little girl, Inês, your smile and your love light up my life and give me strength to overcome all odds.

Thank you all for helping me to make this dream come true.

Abstract

In this thesis, we develop the theory of Error-Orthogonal Models availing ourselves of the identity of these models and those with Commutative Orthogonal Block Structure.

Thus our treatment will rest on the algebraic structure of the models.

In our development we consider: the estimation of variance components; crossing and nesting of models; model joining, in which observations vectors obtained separately are jointly analyzed; step nesting which require much less observations than the corresponding usual models.

To broaden our treatment we also consider L Extensions of Error-Orthogonal models. In this way, we may consider interesting cases such as models otherwise balanced with different numbers of replicates for the treatments.

Last we include normality. We will be interested in obtaining sufficient statistics as well as conditions for them to be complete. We will carry out inference and consider orthogonal L extensions.

Keywords

Orthogonal block structure, commutative orthogonal block structure, error-orthogonal model, segregation, matching, binary operations, L extensions.

Resumo

Nesta tese é desenvolvida a teoria dos modelos *Error-orthogonal* recorrendo à identidade entre estes modelos e os modelos com estrutura ortogonal de blocos comutativos.

Desta forma, o tratamento apresentado irá assentar na estrutura algébrica dos modelos. No desenvolvimento considera-se: a estimação das componentes de variância; o cruzamento e aninhamento de modelos; a junção de modelos, na qual vectores das observações obtidos separadamente são analisados conjuntamente; aninhamento em escada, que requer muito menos observações do que os modelos correspondentes.

Para alargar o tratamento apresentado consideram-se também Extensões L de modelos *Error-orthogonal*. Desta forma, poderemos considerar casos interessantes como o dos modelos com número diferente de repetições para os vários tratamentos.

Por fim, inclui-se o caso normal. Com base no pressuposto da normalidade pretende-se obter estatísticas suficientes assim como condições para que estas sejam completas. É realizada inferência e consideram-se extensões L ortogonais.

Palavras-chave

Estrutura ortogonal em blocos, estrutura ortogonal em blocos comutativos, modelo *error-orthogonal*, segregação, emparelhamento, operações binárias, extensões L.

Contents

List of Figures

Notations and acronyms

$\underline{v}$	:	Vector
$\underline{0}$	:	Null vector
$\underline{1}$	:	Vector of 1's
$\underline{Y}$	:	Random vector
A	:	Matrix
$\lVert \cdot \rVert$	:	Euclidean norm
$\underline{\underline{X}}$	:	Random matrix
0_n	:	Null matriz of order n
I_n	:	Identity matrix of order n
J_n	:	Matrix of 1's of order n
A^T	:	Transpose of the matrix A
A^{-1}	:	Inverse of the matrix A
A^+	:	Moore-Penrose inverse of the matrix A
$A \perp B$	:	Matrices A and B are pairwise orthogonal
$\text{rank}(A)$	:	Rank of matrix A
$\det(A)$	:	Determinant of matrix A
$R(A)$	:	Range of matrix A
$Q(A)$	:	Ortogonal projection matrix on the range space of matrix A
$N(A)$	:	Nullspace of matrix A
$\otimes$	:	Kronecker matrix product
$\dim(V)$	:	Dimension of the sub-space V
$V^\perp$	:	Orthogonal complement of V
$\underline{V}_\nabla$	:	orthogonal projection of $\underline{V}$ on the subspace ∇

$\mathcal{A}$	:	The algebra (CJAS) $\mathcal{A}$
$\dim(\mathcal{A})$	:	Dimension of the algebra (CJAS) $\mathcal{A}$
$pb(\mathcal{A})$	:	Principal base of the algebra(CJAS) $\mathcal{A}$
$\mathcal{A}(\underline{M})$	:	The algebra (CJAS) generated by $\underline{M}$
$\underline{M} \setminus \underline{Q}$	:	Transition matrix between the families $\underline{M}$ and $\underline{Q}$
$\oplus$	:	orthogonal direct sum
$\mathcal{A}_1 \otimes \mathcal{A}_2$	:	Kronecker product between the CJAS $\mathcal{A}_1$ and $\mathcal{A}_2$
$\mathcal{A}_1 * \mathcal{A}_2$	:	Restricted Kronecker product between the CJAS $\mathcal{A}_1$ and $\mathcal{A}_2$
$\mathcal{A}_1 \underset{(C)}{*} \mathcal{A}_2$	:	Generalized Kronecker product between the CJAS $\mathcal{A}_1$ and $\mathcal{A}_2$
$\mathcal{A}_1 \times \mathcal{A}_2$	:	Cartesian product between the CJAS $\mathcal{A}_1$ and $\mathcal{A}_2$
$P(\cdot)$	:	Probability
$E(X)$	:	Expected value of the random variable X
$V(X)$	:	Variance of the random variable X
$COV(X,Y)$	:	Covariance between random variables X and Y
$E(\underline{Y})$	:	Expected value of the random vector $\underline{Y}$
$V(\underline{Y})$	:	Variance-covariance matrix of the random vector $\underline{Y}$
$V(\underline{X},\underline{Y})$	:	Cross-covariance matrix between the random vectors $\underline{X}$ and $\underline{Y}$
$\underline{\mu}$	:	Mean vector
$\underline{X} \sim \mathcal{N}(\cdot \mid \underline{\mu}, V)$	:	$\underline{X}$ is a normal random vector with mean vector $\underline{\mu}$ and variance-covariance matrix V
χ^2_n	:	Central chi-square random variable with n degrees of freedom
$\chi^2_{n,\delta}$	:	Chi-square random variable with n degrees of freedom and non-centrality parameter δ
OPM	:	Orthogonal projection matrix
POOPM	:	Pairwise orthogonal orthogonal projection matrices
FPOOPM	:	Family of pairwise orthogonal orthogonal projection matrices

JA	:	Jordan Algebra
CJA	:	Commutative Jordan Algebra
CJAS	:	Commutative Jordan Algebra of symmetric matrices
LSE	:	Least squares estimator
UMVUE	:	Uniformly minimum variance unbiased estimator
BLUE	:	Best linear unbiased estimator
UBLUE	:	Uniformly best linear unbiased estimator
OBS	:	Model with orthogonal block structure
COBS	:	Model with commutative orthogonal block structure
CCOBS	:	Model with completely commutative orthogonal block structure
EO	:	Error-orthogonal model
CEO	:	Complete error-orthogonal model
SEO	:	Error-orthogonal model with segregation
MEO	:	Error-orthogonal model with matching
EEO	:	Expanding error-orthogonal model
SCEO	:	Complete error-orthogonal model with segregation
MCEO	:	Complete error-orthogonal model with matching

1. Introduction

Linear models can be considered the core of linear statistical inference constituting the foundation of much of statistical practice.

Using the matrix notation we can represent a linear model by

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \ ,$$

where $\underline{Y}$ is the observations vector, X is the design matrix, $\underline{\varepsilon}$ is the errors vector and $\underline{\beta}$ is a vector of unknown parameters β_j, $j = 1, \dots, k$, that can be all constants, all random variables or a combination of both. When some of the parameters $\beta_0, \beta_1, \dots, \beta_k$ are considered as constants and others as random variables we have a mixed model.

Mixed models are a versatile and powerful tool for analysing data collected in experiments and, over the years, they have been applied to several areas such as biological and medical research, animal and human genetics, agriculture or industry.

For a general presentation of the theory of mixed models we can consult, for instance, Khuri et al (1998).

In our work, the mixed model

$$\underline{Y} = \sum_{i=0}^{w} X_i \underline{\beta}_i \quad ,$$

where $\underline{\beta}_0$ is fixed and $\underline{\beta}_1, \dots, \underline{\beta}_w$ are independent random vectors with null mean vectors and variance-covariance matrices $\sigma_1^2 I_{c_1}, \dots, \sigma_w^2 I_{c_w}$, where $c_i = \text{rank}(X_i)$, $i = 1, \dots, w$, plays a central part. More precisely, we will focus on those who constitute a special class within the models with Orthogonal Block Structure (OBS), this is, a particular case of the mixed models whose structure have variance-covariance matrix

$$V = \sum_{j=1}^{m} \gamma_j Q_j$$

where the $Q_1, \dots, Q_m$ are known pairwise orthogonal orthogonal projection matrices such that

$$\sum_{j=1}^{m} Q_j = I_n \, .$$

OBS were introduced by J. A. Nelder , see Nelder (1965a)(1965b). These models have been intensively studied, see for instance Houtman & Speed (1983) and Mejza (1992) and continue to play an important role in the theory of randomized block designs, see Calinski & Kageyama (2000, 2003).

Error-orthogonal models (EO), introduced by VanLeeuwen et al (1998, 1999), are models with orthogonal block structure, OBS, where the least squares estimators, LSE, for estimable vectors are uniformly best linear unbiased estimators, UBLUE, this is to say that whatever $\underline{\gamma}=(\gamma_1,\dots,\gamma_m)$ they are best linear unbiased estimators, BLUE. Thus, given the LSE $\underline{\tilde{\Psi}}$ and another unbiased estimator $\underline{\Psi}^*$ of an estimable vector $\underline{\Psi}$, the difference $V(\underline{\Psi}^*)-V(\underline{\tilde{\Psi}})$ of their variance-covariance matrices is, whatever $\underline{\gamma}$, a positive semi-definite matrix.

It is now convenient to recall a version of the Gauss-Markov theorem due to Zmyslony (1978) which refers that " If the orthogonal projection matrix on the space spanned by the mean vector of the model commutes with the variance-covariance matrix, V, the LSE of estimable vectors are BLUE". We point out that to apply this theorem it's not necessary that the models has orthogonal block structure and moreover that the space, Ω, spanned by $\underline{\mu}=X_0\underline{\beta}_0$ is the range space, $R(X_0)$. Actually this result motivated the introduction of COBS by Fonseca et al. (2008), as a special class of OBS in which matrix T, the orthogonal projection matrix on the space spanned by the mean vector, commutes with the matrices $Q_1,\dots,Q_m$. Thus, whatever $\underline{\gamma}$, T and V will commute ensuring that COBS are EO.

VanLeeuwen et al (1998) showed that EO and COBS are identical classes of models. In studying EO we will favour the COBS approach which besides leading directly to UBLUE, according to Zmyslony (1978), is, as we shall see, interesting in:

- Estimating variance components;
- Building up complex models from simpler ones using, for instance, model crossing and model nesting;
- Discussing sufficient and complete statistics once normality is assumed.

Besides this introduction this thesis comprises three more chapters.

The preliminary results chapter will be on matrices and on estimation. We start by presenting some important results on Matrix Algebra where we emphasize the Kronecker matrix product and the commutative Jordan Algebras of symmetric matrices, CJAS. These algebras are linear spaces constituted by symmetric matrices that commute and containing the squares of its matrices. Each algebra, $\mathcal{A}$, has, see Seely (1971), an unique basis, the principal basis pb($\mathcal{A}$), whose elements are pairwise orthogonal orthogonal projection matrices. These algebras play a central part when we use the COBS approach.

The results on estimation will refer to LSE and to the use of sufficient and complete statistics to obtaining good pointwise estimators. These last results will be useful in the study of the normal models.

In the third chapter we start by considering mixed models whose mean vector is

$$\underline{\mu} = X_0 \underline{\beta}_0$$

and whose variance-covariance matrix is

$$V = \sum_{i=1}^{w} \sigma_i^2 M_i \ ,$$

with $M_i = X_i X_i^T$, $i = 1,\ldots,w$. Assuming, with $R(U)$ the range space of matrix U, that

$$R\left(\begin{bmatrix} X_1 & \cdots & X_w \end{bmatrix}\right) = R^n \ ,$$

when matrices $M_1,\ldots,M_w$ [and T] commute the model, as we will see, is OBS [COBS]. The first of these results is also established in VanLeeuwen et al (1998).

Next we present an independent proof of the identity of EO and COBS showing that, if LSE are UMVUE, the OPM T commutes with the $Q_1, \ldots, Q_m$.

When the model has commutative orthogonal block structure, the matrices $T, M_1,\ldots,M_w$ will belong to a CJAS with principal basis constituted by the $Q_1, \ldots, Q_m$. It may be shown that

$$T = \sum_{j=1}^{z} Q_j \ ,$$

with $z < m$, and

$$M_i = \sum_{j=1}^{m} b_{i,j} Q_j \ , \ i = 1,\ldots,w \ ,$$

so that

$$V = \sum_{j=1}^{m} \gamma_j Q_j$$

with the canonical variance components

$$\gamma_j = \sum_{i=1}^{w} b_{i,j} \sigma_i^2 \ , \ j = 1,\ldots,m \, .$$

As we shall see, the relations between the usual and the canonical variance components will be very useful in estimating.

Next we consider building up complex models from simple ones. Namely we will consider model crossing and model nesting. In model crossing the treatments of the new model, resulting from the crossing, are all the combinations of treatments of the initial models. In model nesting each treatment of the first model nests all treatments of the other. Our techniques for building these models will rest on binary operations on CJAS, namely the Kronecker product and restricted Kronecker product of CJAS. A third operation on CJAS, the Cartesian product, will be used in the study of model joining. Then we superimpose observation vectors and carry out joint inference. This operation also is relevant in connection with a special class of models, those derived through step nesting. This class is interesting since it leads to great economy in the number of observations.

In this third chapter we consider L extensions in which the observations vector is given by

$$\underline{Y} = L\underline{Y}^{0} + \underline{\varepsilon}.$$

Here, $\underline{Y}^{0}$ will be the observations vector of an EO, independent from the error vector, $\underline{\varepsilon}$, and the matrix L will have linearly independent column vectors. These extension are interesting since they include, for instance, models unbalanced in the last step.

In fourth chapter we present the normal case. When considering normality our previous treatment leads directly to sufficient statistics. As for completeness a very specific problem arises when we consider mixed models, since linear restrictions on the $\underline{\tilde{\eta}}_1, \dots, \underline{\tilde{\eta}}_z$ or the canonical variance components may arise and we will only have sufficient but not complete statistics.

We include in this chapter a section on inference in which we avail ourselves of the normality of the observation vectors. We also study orthogonal L extensions in which the column vectors of matrix L are pairwise orthogonal with norm 1.

Our option of using the COBS approach, in studying the models, rests on the point that in this way the algebraic structure of the models plays a central part leading to interesting results on the estimation of variance components and on the building up of models. Moreover, as pointed above, when normality is assumed this approach leads directly to sufficient statistics. The VanLeeuwen et al (1998) definition of EO is, in our opinion, strongly connected with LSE. Now with the COBS approach the Zmyslony (1978) version of the Gauss-Markov theorem gives directly the same optimal property for the LSE as in considered in the VanLeeuwen et al (1998). So using the COBS approach we are considering both estimable vectors and variance components. It may be interesting to point out that in model build-up we obtain complex models which, when they are EO, have LSE which have optimal properties.

Finally, the fifth chapter summarizes the results obtained in the preceding chapters and presents the direction in future research.

2. Preliminary results

Matrix Algebra plays an important role in many areas of Statistics. In particular in linear statistical models it's usual to use Matrix Algebra in the presentation and verification of results, because it allows us to handle efficiently the complexity of multiple observed variables.

In this chapter we include important results on Matrix Algebra and estimation that will be needed on the remainders chapters.

We present some results in matrix theory, on topics such as orthogonal projection matrices, Moore-Penrose inverse and Kronecker matrix product, and pay special attention to commutative Jordan Algebras of symmetric matrices, CJAS, that will be used to express the algebraic structure of the models we will study.

The proofs not included in this first section can be found, for instance, in Schott (1997).

The results on estimation will refer to least squares estimators, LSE, and to sufficient and complete statistics. These last results, and conjunction with the results presented for normal vectors, will be useful in the study of the normal models.

2.1. Matrices

We will restrict ourselves to real matrices, this is, matrices whose elements are all real.

Let A be an $n\times m$ matrix. We will use the notation $a_{i,j}$ to refer to the element in the i-th row and j-th column, $i=1,\dots,n$, $j=1,\dots,m$, of matrix A and we write $A=\lfloor a_{i,j}\rfloor$.

Computing the Euclidean vector norm on the stacked columns of an $n\times m$ matrix, $A=\lfloor a_{i,j}\rfloor$, the Euclidean norm of A is defined by

$$\|A\|=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}a_{i,j}^{2}}\ . \qquad (2.1.1)$$

When a matrix has the same number of rows as columns it is called a square matrix. An $n\times n$ square matrix is said to be of order n.

In a square matrix the elements $a_{i,i}$, $i=1,\dots,n$, are called diagonal elements. If all other elements of this matrix are zero the matrix is said to be diagonal and we write,

$$A=D\left(a_{1,1},\dots,a_{n,n}\right). \qquad (2.1.2)$$

When a matrix is presented as a partition of several blocks (sub-matrices) we call it a blockwise matrix.

A square blockwise matrix of the form

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{bmatrix} , \qquad (2.1.3)$$

where $1 \le r \le n$, in which the off-diagonal blocks are null matrices, is called a block diagonal matrix and we write

$$A = D(A_1 \ldots A_r). \qquad (2.1.4)$$

2.1.1. Symmetric and orthogonal projection matrices

Definition 2.1. A square matrix, M, is said to be symmetric if

$$M^T = M ,$$

this is, if the element in i-th row and j-th column equals the element in j-th row and i-th column, for all i and j .

Definition 2.2. A square matrix, P, is said to be an orthogonal matrix if

$$P P^T = P^T P = I ,$$

where I denotes the identity matrix.

The previous and the following definitions are equivalent.

Definition 2.3. If the matrix P is invertible

$$P^{-1} = P^T ,$$

where P^{-1} is the inverse of matrix P.

Definition 2.4. Let A and B be square matrices of order n. Matrix A is said to be similar to matrix B, we write, $A \sim B$ if there exists an invertible matrix P, of order n, such that

$$P^{-1} A P = B .$$

Many times we need to replace a matrix with another similar to it that is simpler, or in some way easier to deal. Being diagonal matrices the simplest matrices, we can replace the matrix by a diagonal matrix similar to it. When a matrix is similar to a diagonal matrix we say it is diagonalizable.

A symmetric matrix, M, is orthogonally diagonalizable if there is an orthogonal matrix P, whose columns are the (linear independent) eigenvectors, $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$, of M, such that

$$P^T M P = D(\theta_1, \ldots, \theta_k), \tag{2.1.5}$$

where $D(\theta_1, \ldots, \theta_k)$ is the diagonal matrix whose principal elements are the eigenvalues, $\theta_1, \ldots, \theta_k$, of M.

The inverse of a matrix is defined for square invertible matrices but often, in the study of Statistics, we need to use a matrix that behaves like an inverse for rectangular or singular matrices. Moore, in 1920, and Penrose, in 1955, developed a generalized inverse, for any $m \times n$ matrix, that possesses four properties that the inverse of a square invertible matrix has.

Given an $m \times n$ matrix A there is an unique $n \times m$ matrix A^+, the Moore-Penrose inverse of A, satisfying the conditions:

$$AA^+ A = A \quad , \tag{2.1.6}$$

$$A^+ AA^+ = A^+ \quad , \tag{2.1.7}$$

$$(AA^+)^T = AA^+ \quad , \tag{2.1.8}$$

$$(A^+ A)^T = A^+ A \quad . \tag{2.1.9}$$

For any invertible square matrix A,

$$A^+ = A^{-1}. \tag{2.1.10}$$

Since $\left(A^T\right)^+ = \left(A^+\right)^T$, if A is symmetric, A^+ will be symmetric.

With M symmetric, we will have

$$\begin{cases} M = P^T D(\theta_1, \ldots, \theta_k) P \\ M^+ = P^T D\left(\theta_1^+, \ldots, \theta_k^+\right) P \end{cases}, \tag{2.1.11}$$

where

$$\theta_j^+ = \begin{cases} \theta_j^{-1}, \text{when } \theta_j \neq 0, \ j = 1, \ldots, k \\ 0, \text{when } \theta_j = 0, \ j = 1, \ldots, k \end{cases}. \tag{2.1.12}$$

Proposition 2.1. *A matrix P is an orthogonal projection matrix (OPM) if and only if it is symmetric and idempotent.*

Proof. Given a vector space V any vector $\underline{x} \in V$, can be uniquely expressed as $\underline{x} = \underline{x}_1 + \underline{x}_2$ where $\underline{x}_1$ is in a subspace $S \subseteq V$ and $\underline{x}_2$ is in the orthogonal complement, $S^{\perp}$. If P is an orthogonal projection matrix of $\underline{x}$ onto S, $P\underline{x} = \underline{x}_1$ and $P\underline{x}_1 = \underline{x}_1$, that means that further projections to S should have no effect on $P\underline{x}_1$. So $P\underline{x} = \underline{x}_1 = P\underline{x}_1 = P(P\underline{x}) = P^2\underline{x}$ and $\left(P - P^2\right)\underline{x} = 0$. Once $\underline{x}$ is arbitrary, we have $P = P^2$, which mean that P is an idempotent matrix.

Being $\underline{x}_1$ the orthogonal projection of $\underline{x}$, noting that $\underline{x}_2 = \underline{x} - \underline{x}_1 = (I - P)\underline{x}$, we have $0 = \underline{x}_1^T \underline{x}_2 = \underline{x}^T P^T (I - P)\underline{x}$ hence $P^T(I - P) = 0$ so that $P^T = P^T P$ and $P = P P^T$ then P is symmetric. Conversely, if P is a symmetric and idempotent matrix,

$$\underline{x}_1^T \underline{x}_2 = \underline{x}^T P^T (\underline{x} - \underline{x}_1) = \underline{x}^T\left(P\underline{x} - P^2\underline{x}\right) = \underline{x}^T\left(P - P^2\right)\underline{x} = 0,$$

hence P is an orthogonal projection matrix.

□

Proposition 2.2. If P *is an OPM its eigenvalues will be equal to 0 or to 1.*

Proof. Being P an OPM and $\underline{x}$ an eigenvector of matrix P for the eigenvalue λ, we have $\lambda\underline{x} = P\underline{x} = P^2\underline{x} = P(P\underline{x}) = P(\lambda\underline{x}) = \lambda P\underline{x} = \lambda^2\underline{x}$. Since eigenvectors are non-null vectors $\lambda\underline{x} = \lambda^2\underline{x} \Leftrightarrow \lambda(1 - \lambda)\underline{x} = 0$ only occurs when $\lambda = 0$ or $\lambda = 1$.

□

From Proposition 2.1 it follows that, if Q is an OPM then

$$Q^{+} = Q \quad . \tag{2.1.13}$$

Definition 2.5. Two orthogonal projection matrices, Q_1 and Q_2, are pairwise orthogonal, we put $Q_1 \perp Q_2$, when

$$Q_2 Q_1 = 0_{k\times k}^T = 0_{k\times k} \ ,$$

where $0_{r\times s}$ is the $r \times s$ null matrix.

Proposition 2.3. *If* Q_1 *and* Q_2 *are pairwise orthogonal orthogonal projection matrices, POOPM, then* $Q_1 + Q_2$ *is an orthogonal projection matrix.*

Proof. Since the sum of symmetric matrices gives a symmetric matrix and

$$(Q_1 + Q_2)(Q_1 + Q_2) = Q_1Q_1 + Q_1Q_2 + Q_2Q_1 + Q_2Q_2 = Q_1 + Q_2 \ ,$$

since Q_1 and Q_2 are idempotent and pairwise orthogonal matrices, the sum of two pairwise orthogonal orthogonal projection matrices is an orthogonal projection matrix.

□

In what follows, families of pairwise orthogonal orthogonal projection matrices, FPOOPM, will play a central part.

Moreover, see Mexia (1995), the orthogonal projection matrix on the range space of matrix X , $\Omega = R(X)$, will be

$$Q(\Omega) = Q(X) = X(X^T X)^+ X^T = XX^+ \ . \tag{2.1.14}$$

We point out that

$$X^+ = (X^T X)^+ X^T , \tag{2.1.15}$$

which reduces the problem of obtaining Moore-Penrose inverses to getting them for symmetric matrices.

2.1.2. $\mathscr{B}$-matrices

These matrices are relevant in connection with least square estimator, LSE, as we shall see.

Definition 2.6. An $r \times s$ matrix $C = [c_{i,j}]$ is a $\mathscr{B}$-matrix when

$$\begin{cases} \frac{1}{s}\sum_{j=1}^{s} c_{i,j} = c & i = 1,\dots,r \\ \frac{1}{r}\sum_{i=1}^{r} c_{i,j} = c & j = 1,\dots,s \end{cases} ,$$

where

$$c = \frac{1}{r \times s}\sum_{i=1}^{r}\sum_{j=1}^{s} c_{i,j} \ .$$

Proposition 2.4. *Let* C *be a* $\mathscr{B}$*-matrix* , then

$$\begin{cases} \frac{1}{r}J_rC = \begin{bmatrix} \frac{1}{r}\sum_{i=1}^{r} c_{i,1} & \cdots & \frac{1}{r}\sum_{i=1}^{r} c_{i,s} \\ \vdots & & \vdots \\ \frac{1}{r}\sum_{i=1}^{r} c_{i,1} & \cdots & \frac{1}{r}\sum_{i=1}^{r} c_{i,s} \end{bmatrix} \\ C\frac{1}{s}J_s = \begin{bmatrix} \frac{1}{s}\sum_{j=1}^{s} c_{1,j} & \cdots & \frac{1}{s}\sum_{j=1}^{s} c_{1,j} \\ \vdots & & \vdots \\ \frac{1}{s}\sum_{j=1}^{s} c_{r,j} & \cdots & \frac{1}{s}\sum_{j=1}^{s} c_{r,j} \end{bmatrix} \end{cases},$$

with $J_n = \underline{1}_n \underline{1}_n^T$ *, where* $\underline{1}_n$ *is the* $n \times 1$ *vector with all components equal to 1, thus*

$$\frac{1}{r}J_rC = C\frac{1}{s}J_s .$$

Proof. Since C is a $\mathscr{B}$-matrix, we have

$$\begin{cases} \frac{1}{r}\sum_{i=1}^{r} c_{i,1} = \frac{1}{s}\sum_{j=1}^{s} c_{1,j} & \cdots & \frac{1}{r}\sum_{i=1}^{r} c_{i,s} = \frac{1}{s}\sum_{j=1}^{s} c_{1,j} \\ \vdots & & \vdots \\ \frac{1}{r}\sum_{i=1}^{r} c_{i,1} = \frac{1}{s}\sum_{j=1}^{s} c_{r,j} & \cdots & \frac{1}{r}\sum_{i=1}^{r} c_{i,s} = \frac{1}{s}\sum_{j=1}^{s} c_{r,j} \end{cases},$$

thus $\frac{1}{r}J_rC = C\frac{1}{s}J_s$.

□

We now establish the following lemma,

Lemma 2.1. *We have* $\frac{1}{r}J_rC = C\frac{1}{s}J_s$ *if and only if* C *is a* $\mathscr{B}$*-matrix.*

Proof. If $\frac{1}{r}J_rC = C\frac{1}{s}J_s$ we have $\frac{1}{r}\sum_{i=1}^{r} c_{i,1} = \cdots = \frac{1}{r}\sum_{i=1}^{r} c_{i,s}$ as well as $\frac{1}{s}\sum_{j=1}^{s} c_{1,j} = \cdots = \frac{1}{s}\sum_{j=1}^{s} c_{r,j}$, so $\sum_{i=1}^{r} c_{i,j} = \frac{1}{s}\sum_{i=1}^{r}\sum_{j=1}^{s} c_{i,j}$, $j = 1,\ldots,s$ and $\sum_{j=1}^{s} c_{i,j} = \frac{1}{r}\sum_{i=1}^{r}\sum_{j=1}^{s} c_{i,j}$, $i = 1,\ldots,r$ and C is a $\mathscr{B}$-matrix. Inversely, if C is a $\mathscr{B}$-matrix it is straightforward to see that the conditions of having $\frac{1}{r}J_rC = C\frac{1}{s}J_s$ holds.

□

2.1.3. Kronecker matrix products

The Kronecker matrix product is a special type of matrix multiplication without size restrictions. This product gives the possibility to obtain a composite matrix of the elements of any pair of matrices.

The Kronecker product has important applications in Statistics, namely on the representation of variance-covariance matrices. We will repeatedly use this operation which has been widely studied, see, for instance, Steeb (1991), Graham (1981) and Steeb & Hardy (2011).

Definition 2.7. Given $A = [a_{i,j}]$, an $m \times n$ matrix, and $B = [b_{k,l}]$, an $p \times q$ matrix, the Kronecker product between A and B , denoted by $A \otimes B$, is defined as the $mp \times nq$ matrix

$$A \otimes B = \begin{bmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{bmatrix} .$$

The Kronecker product is not commutative but it satisfies the associative law, whatever matrices A , B and C, since

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C = A \otimes B \otimes C \; . \tag{2.1.16}$$

Let A and B be $m \times n$ matrices and C and D $p \times q$ matrices, we have

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D \; , \tag{2.1.17}$$

which means that the Kronecker product satisfies the distributive law.

Let A be an $m \times n$ matrix and B a $p \times q$ matrix, for scalar α , we have

$$(\alpha A) \otimes B = \alpha(A \otimes B) = A \otimes (\alpha B). \tag{2.1.18}$$

The next proposition, about the mixed product property, provides a very important and useful fact regarding the interchangeability of the conventional matrix product and the Kronecker product.

Proposition 2.5. *Let* A, B, C *and* D *be* $m \times n$, $r \times s$, $n \times p$ *and* $s \times t$ *matrices, respectively. If the usual matrices products* AC *and* BD *are defined, then*

$$(A \otimes B)(C \otimes D) = AC \otimes BD \ .$$

Proof.

$$(A \otimes B)(C \otimes D) = \begin{bmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{bmatrix} \begin{bmatrix} c_{1,1}D & \cdots & c_{1,p}D \\ \vdots & \ddots & \vdots \\ c_{n,1}D & \cdots & c_{n,p}D \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^{n} a_{1k}c_{k1}BD & \cdots & \sum_{k=1}^{n} a_{1k}c_{kp}BD \\ \vdots & \ddots & \vdots \\ \sum_{k=1}^{n} a_{mk}c_{k1}BD & \cdots & \sum_{k=1}^{n} a_{mk}c_{kp}BD \end{bmatrix}$$

$$= AC \otimes BD \, .$$

□

Proposition 2.6. *For all matrices* A *and* B,

$$(A \otimes B)^T = A^T \otimes B^T \ .$$

Proof.

$$(A \otimes B)^T = \begin{bmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{bmatrix}^T = \begin{bmatrix} a_{1,1}B & \cdots & a_{m,1}B \\ \vdots & \ddots & \vdots \\ a_{1,n}B & \cdots & a_{m,n}B \end{bmatrix} = A^T \otimes B^T \, .$$

□

Corollary 2.1. *Given* A *and* B *symmetric matrices,*

$$(A \otimes B)^T = A \otimes B \ ,$$

this is, the Kronecker product of symmetric matrices gives symmetric matrices.

Proof. Since A and B are symmetric matrices, from Proposition 2.6,

$$(A \otimes B)^T = A^T \otimes B^T = A \otimes B \, .$$

□

Proposition 2.7. *The Kronecker product of idempotent matrices gives idempotent matrices.*

Proof. Defined the usual matrices products, with A and B idempotent, we will have from mixed product property (Proposition 2.5.),

$$(A \otimes B)(A \otimes B) = (AA) \otimes (BB) = A \otimes B \ .$$

□

From Corollary 2.1. and Proposition 2.7 it follows that, the Kronecker product of orthogonal projection matrices gives orthogonal projection matrices. Moreover, if $Q_1 \perp Q_3$ and $Q_2 \perp Q_4$, with Q_1, Q_2, Q_3 and Q_4 OPM we will have

$$(Q_1 \otimes Q_2)(Q_3 \otimes Q_4) = (Q_1 Q_3) \otimes (Q_2 Q_4) = 0 \ ,$$

with 0 a null matrix, and so

$$(Q_1 \otimes Q_2) \perp (Q_3 \otimes Q_4) \quad .$$

Proposition 2.8. *Whatever matrices* A *and* B*, we have*

$$(A \otimes B)^+ = A^+ \otimes B^+ \quad .$$

Proof. We have

$$(A \otimes B)\left(A^+ \otimes B^+\right)(A \otimes B) = \left(A A^+ A\right) \otimes \left(B B^+ B\right) = A \otimes B$$

and

$$\left(A^+ \otimes B^+\right)(A \otimes B)\left(A^+ \otimes B^+\right) = \left(A^+ A A^+\right) \otimes \left(B^+ B B^+\right) = A^+ \otimes B^+$$

thus, the first and second conditions for $A^+ \otimes B^+$ to be Moore-Penrose inverse of $A \otimes B$ hold. Once

$$\left[(A \otimes B)(A \otimes B)^+\right]^T = \left[(A \otimes B)\left(A^+ \otimes B^+\right)\right]^T = \left[\left(AA^+\right) \otimes \left(BB^+\right)\right]^T = \left(AA^+\right)^T \otimes \left(BB^+\right)^T$$
$$= \left(AA^+\right) \otimes \left(BB^+\right) = (A \otimes B)\left(A^+ \otimes B^+\right) = (A \otimes B)(A \otimes B)^+$$

thus, also the third condition for $A^+ \otimes B^+$ to be Moore-Penrose inverse of $A \otimes B$ holds. The fourth condition for $A^+ \otimes B^+$ to be Moore-Penrose inverse of $A \otimes B$ can be proved analogously.

□

Next, with $Q(A)$ $[Q(B)]$ the OPM on the range space, $R(A)$ $[R(B)]$, of A $[B]$, we can establish

Proposition 2.9. *Whatever matrices* A *and* B*, we have*

$$Q(A \otimes B) = Q(A) \otimes Q(B).$$

Proof. From (2.1.14) and Propositions 2.5 and 2.8., we have

$$Q(A \otimes B) = (A \otimes B)(A \otimes B)^{+} = (A \otimes B)(A^{+} \otimes B^{+}) = (AA^{+}) \otimes (BB^{+}) = Q(A) \otimes Q(B) \; .$$

□

2.1.4. Jordan algebras

Jordan algebras were introduced by Pascual Jordan, in 1933, in his paper devoted to the axiomatic foundation of quantum mechanics and developed one year later in partnership with John von Neumann and Eugene Wigner, see Jordan et al (1934). Later on Seely (1970a) rediscover these structures and used them to solve problems in statistical inference and estimation area, calling them quadratic vector spaces. For priority sake we will call them Jordan Algebras. With Seely was initiated a very fruitful research line with relevant developments of linear statistical inference, see Seely (1970b, 1971, 1977) and Seely & Zyskind (1971).

Later, in Michalski & Zmyslony (1996) and (1999), the Jordan algebras have been used in hypothesis test, first for variance components and later for linear combinations of parameters in mixed linear models.

More recently the papers of Vanleuwen et al (1998, 1999) are highly interesting opening new research areas which we will pursue.

We can also quote Fonseca et al (2003, 2006, 2007, 2008, 2009), Rodrigues & Mexia (2006) and Jesus et al (2007, 2009a, 2009b).

For completeness sake, the definition of algebra is stated.

Definition 2.8. An algebra, $\mathcal{A}$, is a linear space provided with a binary operation, denoted here by $*$, usually called product, that satisfies the following conditions, for all $\alpha \in \mathrm{IR}$ and all $\underline{x}, \underline{y}, \underline{z} \in \mathcal{A}$:

$$\begin{cases} \underline{x} * (\underline{y} + \underline{z}) = \underline{x} * \underline{y} + \underline{x} * \underline{z} \\ (\underline{x} + \underline{y}) * \underline{z} = \underline{x} * \underline{z} + \underline{y} * \underline{z} \\ \alpha \cdot (\underline{x} * \underline{y}) = (\alpha \underline{x}) * \underline{y} = \underline{x} * (\alpha \cdot \underline{y}) \end{cases}$$

This product also enjoys the associative and commutative properties, defined below, however these properties are not necessary for a linear space to be an algebra.

Definition 2.9. If, for all $\underline{x}$, $\underline{y}$, $\underline{z} \in \mathcal{A}$,

$$(\underline{x} * \underline{y}) * \underline{z} = \underline{x} * (\underline{y} * \underline{z}) \ ,$$

the algebra $\mathcal{A}$ is said to be an associative algebra.

Definition 2.10. If, for all $\underline{x}$, $\underline{y} \in \mathcal{A}$,

$$\underline{x} * \underline{y} = \underline{y} * \underline{x} \ ,$$

the algebra $\mathcal{A}$ is said to be a commutative algebra.

Definition 2.11. A Jordan algebra (JA) is a commutative algebra, $\mathcal{A}$, whose product satisfies the Jordan identity

$$\underline{x}^2 * \left(\underline{y} * \underline{x}\right) = \left(\underline{x}^2 * \underline{y}\right) * \underline{x}$$

with $\underline{x}^2 = \underline{x} * \underline{x}$, for all $\underline{x}$, $\underline{y} \in \mathcal{A}$.

Definition 2.12. When the matrices of a JA commute it is called a commutative Jordan algebra, CJA.

Definition 2.13. When a CJA is constituted by symmetric matrices it is called a commutative Jordan algebra of symmetric matrices, CJAS.

In order to summarize what was previously set, we can say that a commutative Jordan algebra of symmetric matrices is a linear space constituted by symmetric matrices that commute containing the squares of their matrices.

To avoid going beyond the objectives of our study we will restrict ourselves to CJAS. For a deeper study of Jordan algebras see for instance Jacobson (1953).

Let $\underline{Q} = \{Q_1, \ldots, Q_m\}$ be the principal basis of the CJAS $\mathcal{A}$, pb($\mathcal{A}$). Given M a matrix belonging to a CJAS $\mathcal{A}$, we have

$$M = \sum_{j=1}^{m} b_j Q_j = \sum_{j \in C(M)} b_j Q_j \ , \qquad (2.1.19)$$

with $C(M) = \{j : b_j \neq 0\}$.

Then the Moore-Penrose inverse of M is

$$M^{+} = \sum_{j=1}^{m} b_j^{+} Q_j \ , \tag{2.1.20}$$

where $b_j^{+} = b_j^{-1}$, $\forall b_j \neq 0 \quad j = 1, \dots, m$, and so

$$C\left(M^{+}\right) = C(M)$$

thus, a CJAS contains the Moore-Penrose inverses of any of its matrices.

With

$$\nabla_j = R(Q_j), \ j = 1, \dots, m$$

and

$$g_j = \text{rank}\,(Q_j), \ j = 1, \dots, m,$$

representing by $\oplus$ the orthogonal direct sum of subspaces, we have

$$\begin{cases} R(M) = \underset{j \in C(M)}{\oplus} \nabla_j \\ r(M) = \text{rank}\,(M) = \sum_{j \in C(M)} g_j \end{cases} .$$

Moreover the orthogonal projection matrix on $R(M)$ will be

$$Q(M) = \sum_{j \in C(M)} Q_j \ . \tag{2.1.21}$$

Proposition 2.10. *The orthogonal projection matrices belonging to a CJAS, A, are sums of matrices of the* $\text{pb}(A)$.

Proof. Given Q , an orthogonal projection matrix belonging to A, we have

$$Q = \sum_{j=1}^{m} b_j Q_j \ .$$

Since Q is idempotent and $Q_1, \dots, Q_m$ are idempotent and pairwise orthogonal,

$$Q = \sum_{j=1}^{m} b_j Q_j = \sum_{j=1}^{m} b_j^2 Q_j = Q^2 \ ,$$

coming $b_j^2 = b_j$ and so $b_j = 0$ or $b_j = 1$, $j = 1, \dots, m$, then ,with $C(Q) = \{j : b_j \neq 0\}$,

$$Q = \sum_{j \in C(Q)} Q_j \ .$$

□

Since $\underline{Q} = \{Q_1, \ldots, Q_m\} = pb(\mathcal{A})$ has m matrices, $\mathcal{A}$, as a linear subspace, has dimension $\dim(\mathcal{A}) = m$. Thus there can be 2^m OPM in $\mathcal{A}$, as such as the distinct sums of matrices of $pb(\mathcal{A})$, once each of the sums corresponds to a sub-set of $\overline{\overline{m}} = \{1,\ldots,m\}$. Given $C \subseteq \overline{\overline{m}}$,

$$Q(C) = \sum_{j \in C} Q_j$$

so that, with $r(C) = \text{rank}\,(Q(C))$, we will have

$$r(C) = \sum_{j \in C} g_j \ .$$

We point out that we are considering the $0_{n \times n}$ matrices as an OPM on $\{\underline{0}^n\}$.

We also see that if, with $Q \in \mathcal{A}$, we have $r(Q) = 1$ then we must have $Q \in \underline{Q} = pb(\mathcal{A})$. Namely, with $J_n = 1_n 1_n^T$, if

$$Q = \frac{1}{n} J_n \ ,$$

we put $Q_1 = Q$ and say that $\mathcal{A}$ is a regular CJAS.

We are assuming that the matrices in $\mathcal{A}$ are $n \times n$.

Definition 2.14. When a CJAS, $\mathcal{A}$, contains invertible matrices we say that it is complete.

If $\mathcal{A}$ contains invertible matrices we must have

$$\sum_{j=1}^{m} Q_j = I_n \ , \qquad (2.1.22)$$

since we must have $\sum_{j=1}^{m} g_j = n$, then the matrices in the principal basis of a complete CJAS add up to I_n.

Let M be a matrix belonging to a CJAS, we say M is regular if and only if $C(M) = \overline{\overline{n}} = \{1,\ldots,n\}$.

Given

$$M = \sum_{j=1}^{m} b_j Q_j \ , \qquad (2.1.23)$$

with $b_j \neq 0$, $j = 1,\ldots,m$, the b_j, $j = 1,\ldots,m$, will be the eigenvalues of M with multiplicities g_j, $j = 1,\ldots,m$, so the determinant of matrix M will be

$$\det(M) = \prod_{j=1}^{m} b_j^{g_j} \tag{2.1.24}$$

and

$$M^{-1} = \sum_{j=1}^{m} b_j^{-1} Q_j .$$

Given the family $\underline{M} = \{M_1,\ldots,M_w\}$ of matrices of A, we will have

$$M_i = \sum_{j=1}^{m} b_{i,j} Q_j \ , \ i = 1,\ldots,w$$

and $B = \lfloor b_{i,j} \rfloor$ will be the transition matrix between $\underline{M}$ and $\underline{Q}$, $\underline{M} \setminus \underline{Q}$. The matrices in $\underline{M}$ are linearly independent when and only when the row vectors of B are linearly independent. Since $\dim(A) = m$, if $w = m$ and the matrices $M_1,\ldots,M_m$ are linearly independent the m row vectors of B will be linearly independent, thus B will be $m \times m$ and $\operatorname{rank}(B) = m$. Then B will be invertible and with $B^{-1} = \lfloor b^{l,h} \rfloor$ we will have

$$Q_l = \sum_{h=1}^{m} b^{l,h} M_m \ , \ l = 1,\ldots,m$$

and $\underline{M} = \{M_1,\ldots,M_w\}$ will be a basis for A.

Now, the matrices of $\underline{M} = \{M_1,\ldots,M_w\}$ commute if and only if they are diagonalized by the same matrix, P^o. We then have

$$\underline{M} \subset V(P^o),$$

with $V(P^o)$ the family of matrices diagonalized by P^o. Since $V(P^o)$ is a CJAS, we see that a family of $n \times n$ symmetric matrices is contained in a CJAS if and only if they commute. Since the intersection of CJAS gives CJAS there will be a minimum CJAS containing $\underline{M}$, whose matrices commute, this will be the CJAS $A(\underline{M})$ generated by $\underline{M}$.

Namely if $\underline{D}$ is a FPOOPM, $A(\underline{D})$ will have $\underline{D}$ as principal basis since the CJAS containing $\underline{D}$ must contain the CJAS constituted by the linear combinations of the matrices.

If the $M_1,\ldots,M_w$ commute and are diagonalized by the orthogonal matrix P^o the row vectors $\underline{\alpha}_1,\ldots\underline{\alpha}_n$ of P^o will be eigenvectors for the matrices of $\underline{M}$.

Definition 2.15. Let $\underline{\alpha}_1,\dots\underline{\alpha}_n$ be the eigenvectors of matrix $\underline{M}$. We say that exists an equivalence relation, τ, on $\{\underline{\alpha}_1,\dots\underline{\alpha}_n\}$, writing $\underline{\alpha}_h \, \tau \, \underline{\alpha}_l$, when and only when

$$\underline{\alpha}_h^T M_i \, \underline{\alpha}_h = \underline{\alpha}_l^T M_i \, \underline{\alpha}_l \quad i = 1,\dots,w,$$

this is, when $\underline{\alpha}_h$ and $\underline{\alpha}_l$, $h \neq l$, $h,l = 1,\dots,n$, are associated to identical eigenvalues for all matrices in $\underline{M}$.

Definition 2.16. A τ equivalence class is of the first type if its vectors are associated to a non null eigenvalue for at least one matrix in $\underline{M}$. The number of classes of first type will be the eigenindex of $\underline{M}$.

Besides the first type classes there may be a second type class constituted by the eigenvectors associated to null eigenvalues for all matrices in $\underline{M}$.

With $C_1,\dots,C_m$ the sets of indexes of the $\underline{\alpha}_1,\dots,\underline{\alpha}_n$ belonging to the first type τ equivalency classes the

$$Q_j = \sum_{i \in C_j} \underline{\alpha}_i \underline{\alpha}_i^T \quad, j = 1,\dots,m$$

constitute a FPOOPM, which will be the principal basis of a CJAS, $A(\underline{Q})$ with $\underline{Q} = \{Q_1,\dots,Q_m\}$.

It is easily seen that equality in (2.1.22) holds if and only if there is no second type τ equivalency class. Thus $A(\underline{D})$ is complete when and only when there is no second type τ equivalency class.

Let us establish

Proposition 2.11. *We have* $A(\underline{M}) = A(\underline{Q})$.

Proof . Since $M_i = \sum_{j=1}^{m} b_{i,j} Q_j$, $i = 1,\dots,w$, with $b_{i,1},\dots,b_{i,m}$ the eigenvalues of M_i, for the vectors in the different first type τ equivalency classes, $i = 1,\dots,w$, we have $\underline{M} \subseteq A(\underline{Q})$ so $A(\underline{M}) \subseteq A(\underline{Q})$ as well as $\underline{Q}^0 = \{Q_1^0,\dots,Q_{m^0}^0\} = \mathrm{pb}(A(\underline{M})) \subset A(\underline{Q})$, thus $Q_l^0 = \sum_{j \in D(l)} Q_j$, $l = 1,\dots,m^0$.

Moreover $M_i = \sum_{l=1}^{m^0} b_{i,l}^0 Q_l^0$, $i = 1,\dots,w$, so the $\underline{\alpha}_1,\dots,\underline{\alpha}_n$ with indexes in each of the sets $\bigcup_{j \in D(l)} C_j$ are associated to identical eigenvalues for all matrices in $\underline{M}$ which is impossible unless all sets $D(l)$, $l = 1,\dots,m^0$ contain an unique index. This will imply that the matrices in $\underline{Q}$ are the possibly reordered, matrices in $\underline{Q}^0$.

□

Corollary 2.2. *$A(\underline{M})$ is complete if and only if there is no second type τ equivalency class.*

Corollary 2.3. *The eigenindex of $\underline{M}$ equals* $d(\underline{M}) = \dim(A(\underline{M}))$

Corollary 2.4. *A family $\underline{M}$ of commuting symmetric matrices is a basis for $A(\underline{M})$ is and only if its eigenindex equals its cardinal.*

When $\underline{M}$ is a basis for $A(\underline{M})$ it is a perfect family of symmetric matrices. These families were studied by Ferreira et al (2007).

From the previous results we have an important result established in Seely (1971).

Theorem 2.1. *Every CJAS, A, has an unique basis, the principal basis*

$$\underline{Q} = \{Q_1, \ldots, Q_m\} = pb(A)$$

which is a FPOOPM. Inversely every FPOOPM is the principal basis of a CJAS.

As a parting remark we point out that, given a CJAS, A, polynomials in matrices of A belongs to A.

2.1.5. Binary operations on CJAS

We now consider binary operations on CJAS, more precisely on its principal basis. These operations will be very useful in deriving complex models from simple ones.

The first two operations, the Kronecker product of CJAS and the restricted Kronecker product of CJAS, were introduced on Fonseca et al (2006) and will be relevant for model crossing and model nesting, respectively. The last operation, the Cartesian product of CJAS, was introduced on Fernandes et al. (2010) and will be useful in considering models obtained through joining and step nesting.

2.1.5.1. Kronecker product

Proposition 2.12. *Given $\underline{Q}(l) = \left\{Q_1(l), \ldots, Q_{m(l)}(l)\right\}$, the principal basis of the CJAS $A(l)$, $l = 1, 2$, the Kronecker product between $A(1)$ and $A(2)$, $A(1) \otimes A(2)$, will be the CJAS with principal basis*

$$\underline{Q} = \underline{Q}(1) \otimes \underline{Q}(2) = \{Q_{j'}(1) \otimes Q_{j''}(2);\ j' = 1, \cdots, m(1),\ j'' = 1, \ldots, m(2)\}.$$

Proof. Being $\{Q_1(l), \ldots, Q_{m(l)}(l)\}$, the principal basis of the CJAS $A(l)$, $l = 1, 2$, constituted by pairwise orthogonal orthogonal projection matrices, the matrices

$$Q_{j'}(1) \otimes Q_{j''}(2), \; j' = 1, \cdots, m(1) \;, \; j'' = 1, \ldots, m(2)$$

are also orthogonal projection matrices because, as we saw, the Kronecker product of orthogonal projection matrices is an orthogonal projection matrix. Besides this, using the distributive law of Kronecker matrix product, we have

$$\left(\sum_{j'=1}^{m(1)} Q_{j'}(1)\right) \otimes \left(\sum_{j''=1}^{m(2)} Q_{j''}(2)\right) = \sum_{j'=1}^{m(1)} \sum_{j''=1}^{m(2)} \left(Q_{j'}(1) \otimes Q_{j''}(2)\right)$$

thus, the kronecker of CJAS,

$$A(1) \otimes A(2) \;,$$

is a linear space constituted by symmetric matrices that commute.

On the other hand, being M a matrix belonging to $A(1) \otimes A(2)$, there are two matrices $M_1 \in A(1)$ and $M_2 \in A(2)$ such that $M = M_1 \otimes M_2$.

Since $A(1)$ and $A(2)$ are CJAS, $M_1^2 \in A(1)$ and $M_2^2 \in A(2)$, thus $M^2 \in A(1) \otimes A(2)$, because, from Proposition 2.5, $M^2 = (M_1 \otimes M_2)^2 = M_1^2 \otimes M_2^2$. Therefore $A(1) \otimes A(2)$ contains the square of their matrices, it is a CJAS.

Now, follows from the Theorem 2.1 that, if $Q_{j'}(1) \otimes Q_{j''}(2)$ are pairwise orthogonal orthogonal projection matrices then they constitute the principal basis of the CJAS $A(1) \otimes A(2)$.

□

Given the families of symmetric matrices

$$\underline{M}(l) = \{M_1(l), \ldots, M_{w(l)}(l)\} \subseteq A(l), \; l = 1,2 \qquad (2.1.25)$$

and $B(l) = [b_{i,j}(l)]$, $l = 1, 2$, the transition matrices between $\underline{M}(l)$ and $\underline{Q}(l)$, $\underline{M}(l) \backslash \underline{Q}(l)$, $l = 1,2$, for

$$\underline{M} = \underline{M}(1) \otimes \underline{M}(2) = \{M_{i'}(1) \otimes M_{i''}(2);\; i' = 1, \cdots, m(1),\; i'' = 1, \ldots, m(2)\} \qquad (2.1.26)$$

and $\underline{Q}$ the transition matrix will be

$$B = B(1) \otimes B(2) \qquad (2.1.27)$$

once we order the matrices in $\underline{M}$ and $\underline{Q}$ according to the indexes

$$\begin{cases} i=(i'-1)\,w(2)+i'' \quad , \quad i'=1,\dots,w(1) \quad , \quad i''=1,\dots,w(2) \\ j=(j'-1)\,m(2)+j'' \quad , \quad j'=1,\dots,m(1) \quad , \quad j''=1,\dots,m(2) \end{cases}.$$

Since the Kronecker product of matrices is associative it is easy to see that

$$A_1 \otimes (A_2 \otimes A_3) = (A_1 \otimes A_2) \otimes A_3 , \qquad (2.1.28)$$

this is, the kronecker product of CJAS is associative.

Proposition 2.13. *If A_1 and A_2 are regular CJAS then $A_1 \otimes A_2$ is a regular CJAS.*

Proof. If A_1 and A_2 are regular CJAS , constituted by matrices of order $n(1)$ and $n(2)$, respectively, we have $\frac{1}{n(1)} J_{n(1)} \in A_1$ and $\frac{1}{n(2)} J_{n(2)} \in A_2$. Then

$$\frac{1}{n(1)} J_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)} = \frac{1}{n(1)\,n(2)} J_{n(1)\,n(2)} \in A_1 \otimes A_2 ,$$

thus $A_1 \otimes A_2$ is a regular CJAS.

□

Proposition 2.14. *If A_1 and A_2 are complete CJAS then $A_1 \otimes A_2$ is a complete CJAS.*

Proof. Being A_1 and A_2 complete CJAS, constituted by matrices of order $n(1)$ and $n(2)$, respectively, we have $\sum_{j'=1}^{m(1)} Q_{j'}(1) = I_{n(1)}$ and $\sum_{j''=1}^{m(2)} Q_{j''}(2) = I_{n(2)}$. Then

$$\sum_{j'=1}^{m(1)} \sum_{j''=1}^{m(2)} Q_{j'}(1) \otimes Q_{j''}(2) = I_{n(1)} \otimes I_{n(2)} = I_{n(1)\,n(2)} ,$$

which means that $A_1 \otimes A_2$ is a complete CJAS.

□

2.1.5.2. Restricted Kronecker product

Proposition 2.15. *Let* $\underline{Q}(l) = \{Q_1(l), \dots, Q_{m(l)}(l)\}$, *be the principal basis of the CJAS* A_l , $l = 1,2$. *Putting* $Q_1(2) = \frac{1}{n(2)} J_{n(2)}$, *the restricted Kronecker product between* A_1 *and* A_2, $A_1 * A_2$, *will be the CJAS with principal basis*

$$\underline{Q} = \underline{Q}_1 * \underline{Q}_2 = \left\{ Q_1(1) \otimes \frac{1}{n(2)} J_{n(2)}, \ldots, Q_{m(1)}(1) \otimes \frac{1}{n(2)} J_{n(2)} \right\} \cup \left\{ I_{n(1)} \otimes Q_2(2), \ldots, I_{n(1)} \otimes Q_{m(2)}(2) \right\}$$

Proof. Once the Kronecker product-$\otimes$ of OPM is an OPM, then $Q_* = Q_*(1) \cup Q_*(2)$, where $Q_*(1) = \{Q_1(1) \otimes Q_1(2), \ldots, Q_{m(1)}(1) \otimes Q_1(2)\}$ and $Q_*(2) = \{I_{n(1)} \otimes Q_2(2), \ldots, I_{n(1)} \otimes Q_{m(2)}(2)\}$, will be a family of orthogonal projection matrices.

1) Given two matrices $Q_i(1) \otimes Q_1(2)$ and $Q_{i'}(1) \otimes Q_1(2)$ belonging to $Q_*(1)$, we have

$$(Q_i(1) \otimes Q_1(2))(Q_{i'}(1) \otimes Q_1(2)) = (Q_i(1) Q_{i'}(1)) \otimes (Q_1(2) Q_1(2)) = 0 .$$

2) Given two matrices $I_{n(1)} \otimes Q_j(2)$ and $I_{n(1)} \otimes Q_{j'}(2)$, $j \neq j'$, belonging to $Q_*(2)$, we have

$$(I_{n(1)} \otimes Q_j(2))(I_{n(1)} \otimes Q_{j'}(2)) = (I_{n(1)} I_{n(1)}) \otimes (Q_j(2) Q_{j'}(2)) = I_{n(1)} \otimes (Q_j(2) Q_{j'}(2)) = 0 .$$

3) With $Q_i(1) \otimes Q_1(2) \in Q_*(1)$ and $I_{n(1)} \otimes Q_j(2) \in Q_*(2)$, we have

$$(Q_i(1) \otimes Q_1(2))(I_{n(1)} \otimes Q_j(2)) = (Q_i(1) I_{n(1)}) \otimes (Q_1(2) Q_j(2)) = 0$$

Thus, $\underline{Q}_1 * \underline{Q}_2$ will be the principal basis of the CJAS $\mathcal{A}_1 * \mathcal{A}_2$.

□

When $\mathcal{A}_1$ is regular $\mathcal{A}_1 * \mathcal{A}_2$ is regular. Besides this, if $\mathcal{A}_1$ is complete and $\mathcal{A}_2$ is regular

$$\sum_{j=1}^{m(1)} \left(Q_j(1) \otimes \frac{1}{n(2)} J_{n(2)} \right) = \left(\sum_{j=1}^{m(1)} Q_j(1) \right) \otimes \frac{1}{n(2)} J_{n(2)} = I_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)} = I_{n(1)} \otimes Q_1(2) \quad (2.1.29)$$

This result will be used later on.

If $\mathcal{A}_1$ is complete and $\mathcal{A}_2$ is regular and complete, $\mathcal{A}_1 * \mathcal{A}_2$ is complete since

$$\sum_{j=1}^{m(1)} \left(Q_j(1) \otimes \frac{1}{n(2)} J_{n(2)} \right) + \sum_{j=2}^{m(2)} (I_{n(1)} \otimes Q_j(2)) = I_{n(1)} \otimes Q_1(2) + I_{n(1)} \otimes \sum_{j=2}^{m(2)} Q_j(2)$$

$$= I_{n(1)} \otimes \sum_{j=1}^{m(2)} Q_j(2) = I_{n(1)} \otimes I_{n(2)} = I_{n(1)n(2)} \quad (2.1.30)$$

As well as the Kronecker product, the restricted Kronecker product also satisfies the associative law. Given another CJAS, $\mathcal{A}_3$, with principal basis $\{Q_1(3), \ldots, Q_{m(3)}(3)\}$, putting $Q_1(3) = \frac{1}{n(3)} J_{n(3)}$ we have

$$A_1 * (A_2 * A_3) = (A_1 * A_2) * A_3 \quad , \qquad (2.1.31)$$

see Fonseca et al (2006).

The Kronecker product and the restricted Kronecker product can be applied jointly, for instance,

$$(\mathcal{A}_1 \otimes \mathcal{A}_2) * (\mathcal{A}_3 \otimes \mathcal{A}_4)$$

would be the restricted Kronecker product between $\mathcal{A}_1 \otimes \mathcal{A}_2$ and $\mathcal{A}_3 \otimes \mathcal{A}_4$.

2.1.5.3. Generalized Kronecker product

Given two CJAS $\mathcal{A}_1$ and $\mathcal{A}_2$ the generalized Kronecker product between $\mathcal{A}_1$ and $\mathcal{A}_2$, denoted as $\mathcal{A}_1 \underset{(C)}{*} \mathcal{A}_2$, will be a CJAS with principal basis

$$pb\left(\mathcal{A}_1 \underset{(C)}{*} \mathcal{A}_2\right) = \left[\bigcup_{h \in C} \{Q_1(1) \otimes Q_h(2), \ldots, Q_{m(1)}(1) \otimes Q_h(2)\}\right] \cup \{I_{n(1)} \otimes Q_h(2); h \notin C\} \quad (2.1.32)$$

The two binary operations introduced before are special instances of $\underset{(C)}{*}$, where

$$\{1\} \subseteq C \subseteq \overline{\overline{m_2}} = \{1, \ldots, m_2\}.$$

Putting

$$\mathcal{A}_{1/2}(C) = \mathcal{A}_1 \underset{(C)}{*} \mathcal{A}_2 \qquad (2.1.33)$$

we have the kronecker product between $\mathcal{A}_1$ and $\mathcal{A}_2$,

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \mathcal{A}_{1/2}\left(\overline{\overline{m_2}}\right)$$

and the restricted Kronecker product between $\mathcal{A}_1$ and $\mathcal{A}_2$,

$$\mathcal{A}_1 * \mathcal{A}_2 = \mathcal{A}_{1/2}(\{1\}) \ .$$

2.1.5.4. Cartesian product

The Cartesian product

$$\mathcal{A}_1 \times \mathcal{A}_2 = \mathop{\mathsf{X}}_{l=1}^{2} \mathcal{A}_l \qquad (2.1.34)$$

of the CJAS $\mathcal{A}_l$, $l = 1,2$, will be the CJAS with principal basis

$$\underset{l=1}{\overset{2}{\times}} \underline{Q}(l) = \{D(Q_1(1), 0_{n(2)\times n(2)}) \cdot \ldots \cdot D(Q_{m(1)}(1), 0_{n(2)\times n(2)})\} \cup \{D(0_{n(1)\times n(1)}, Q_1(2)) \cdot \ldots \cdot D(0_{n(1)\times n(1)}, Q_{m(2)}(2))\}$$

(2.1.35)

where we assume that

$$\underline{Q}(l) = pb(\mathcal{A}_l) = \{Q_1(l), \ldots, Q_{m(l)}(l)\},\ l = 1,2$$

and that the matrices in $\mathcal{A}_l$ are $n(l)\times n(l)$, $l = 1,2$.

Given another CJAS, $\mathcal{A}_3$, we have

$$\mathcal{A}_1\times(\mathcal{A}_2\times\mathcal{A}_3) = (\mathcal{A}_1\times\mathcal{A}_2)\times\mathcal{A}_3, \quad (2.1.36)$$

thus the Cartesian product is associative.

2.2. Estimation

In this section we present important results on least squares estimators, LSE, among which we highlight a version of the Gauss-Markov theorem due to Zmyslony.

Devote special attention to the commutativity of the matrices T and V which is a sufficient condition for a linear mixed model be an error-orthogonal model.

We also present an example where we consider a balanced mixed model.

2.2.1. Least squares estimators

In what follows we are interested in models with mean vector

$$\underline{\mu} = E(\underline{Y}) = X\underline{\beta}\ . \quad (2.2.1)$$

Definition 2.17. A vector $\underline{\tilde{\beta}}$ is the least squares estimator, LSE, of $\underline{\beta}$ if it minimizes

$$s(\underline{\beta}) = \left\| \underline{Y} - X\underline{\beta} \right\|^2.$$

Proposition 2.16. *The vector $\underline{\tilde{\beta}}$ is the least squares estimator of $\underline{\beta}$ if and only if $X\underline{\tilde{\beta}} = T\underline{Y}$, where T is the OPM on $\Omega = R(X)$, the space spanned by $\underline{\mu}$.*

Proof. With $\underline{V}_\nabla$ the orthogonal projection of $\underline{V}$ on the subspace ∇, and $\Omega^\perp$ the orthogonal complement of Ω, we have

$$s(\underline{\beta}) = \left\| \underline{Y} - X\underline{\beta} \right\|^2 = \left\| \underline{Y}_{\Omega^\perp} + \underline{Y}_\Omega - X\underline{\beta} \right\|^2 = \left\| \underline{Y}_{\Omega^\perp} \right\|^2 + \left\| \underline{Y}_\Omega - X\underline{\beta} \right\|^2 .$$

Since $\left\| \underline{Y}_{\Omega^\perp} \right\|^2$ does not depend on $\underline{\beta}$, $s(\underline{\beta})$ is minimized by minimizing $\left\| \underline{Y}_\Omega - X\underline{\beta} \right\|^2$, this is, since $T\underline{Y} = \underline{Y}_\Omega$, $s(\underline{\beta})$ is minimized by minimizing $\left\| T\underline{Y} - X\underline{\beta} \right\|^2$, where T is the OPM on $\Omega = R(X)$, the space spanned by $\underline{\mu}$. Thus the squared distance between $T\underline{Y}$ and $X\underline{\beta}$ is zero if and only if $X\underline{\tilde{\beta}} = T\underline{Y}$.

□

Corollary 2.5. $\underline{\tilde{\beta}} = (X^T X)^+ X^T \underline{Y}$ *is the least squares estimator of* $\underline{\beta}$.

Proof. As we saw, the OPM on $\Omega = R(X)$, the space spanned by $\underline{\mu}$, is

$$T = X(X^T X)^+ X^T .$$

So the minimum of $s(\underline{\beta})$ is attained for

$$\underline{\tilde{\beta}} = (X^T X)^+ X^T \underline{Y} .$$

□

Besides this

Definition 2.18. $\underline{\Psi} = G\underline{\beta}$ is estimable if $\underline{\Psi}^* = U\underline{Y}$ is a linear unbiased estimator for $\underline{\Psi}$, for some matrix U.

Then, for every $\underline{\beta}$, we get

$$UX\underline{\beta} = G\underline{\beta} ,$$

so that $G = UX$ which is equivalent to $G^T = X^T U^T$ and, see Mexia (1990),

$$R(G^T) \le R(X^T).$$

If the

$$\underline{\Psi}_l^* = U_l \underline{Y} \quad , \; l = 1,2 , \tag{2.2.2}$$

are unbiased estimators for $\underline{\Psi}$ we have

$$U_1 X = U_2 X = G , \tag{2.2.3}$$

so the row vectors of $U_1 - U_2$ will be orthogonal to the column vectors of X, thus to Ω.

Let us establish

Lemma 2.2. *The* $\underline{\Psi}_l^* = U_l \underline{Y}$, $l = 1,2$, *are unbiased estimators of the same estimable vector if and only if* $U_1 T = U_2 T$.

Proof. If $\underline{\Psi}_l^*$ and $\underline{\Psi}_2^*$ are unbiased estimators of the same estimable vector $(U_1 - U_2)T = (U_1 - U_2)X(X^T X)^+ X^T = 0$, since $(U_1 - U_2)X = 0$, and so $U_1 T = U_2 T$. Inversely, the mean vectors, of $\underline{\Psi}_l^*$, $E(\underline{\Psi}_l^*) = U_l X \underline{\beta} = U_l T X \underline{\beta}$, $l = 1,2$, will be equal when $U_1 T = U_2 T$, and so the proof is complete.

□

The LSE of the estimable vector $\underline{\Psi} = A\underline{\beta}$ will be

$$\underline{\tilde{\Psi}} = A\underline{\tilde{\beta}} = U^0 \underline{Y}, \qquad (2.2.4)$$

with

$$U^0 = A(X^T X)^+ X^T.$$

We now have the

Proposition 2.17. *The LSE of estimable vectors of models with mean vector* $\underline{\mu} = X\underline{\beta}$ *are the* $M\underline{Y}$, *with* $MT = M$.

Proof. Whenever $\underline{\mu} = X\underline{\beta}$ we have $T = Q(X) = X(X^T X)^+ X^T$ the OPM on $R(X)$ the space spanned by $\underline{\mu}$.

If $\underline{\Psi} = G\underline{\beta}$ is estimable there is an unbiased estimator for $\underline{\Psi}$, say $M\underline{Y}$ for some matrix M, and its LSE will be $\underline{\tilde{\Psi}} = G\underline{\tilde{\beta}}$ with $\underline{\tilde{\beta}} = (X^T X)^+ X^T \underline{Y}$. So

$$M\underline{Y} = G\underline{\tilde{\beta}} = G(X^T X)^+ X^T \underline{Y}$$

thus

$$M = G(X^T X)^+ X^T .$$

Since $\left(X^T X\right)^+ X^T X \left(X^T X\right)^+$ is the Moore- Penrose inverse of $X^T X$ we have,

$$MT = G\left(X^T X\right)^+ X^T X \left(X^T X\right)^+ X^T = M \ .$$

Inversely, if $MT = M$, we have

$$M\underline{Y} = MT\underline{Y} = MX\left(X^T X\right)^+ X^T \underline{Y} = MX\underline{\tilde{\beta}} \ .$$

and $M\underline{Y}$ is the LSE of $\underline{\Psi} = G\underline{\beta}$ with $G = MX$.

□

Corollary 2.6. *If $\underline{\Psi}^* = U\underline{Y}$ is an unbiased estimator for $\underline{\Psi}$, the LSE for $\underline{\Psi}$ will be $\underline{\tilde{\Psi}} = UT\underline{Y}$.*

Proof. Since $UTT = UT$, it suffices to point out that $\underline{\tilde{\Psi}}$ is an unbiased estimator of $\underline{\Psi}$ of the type indicated in the thesis of Proposition 2.17.

□

According to Zmyslony (1978), we get the following relevant version of the Gauss-Markov theorem. Before, we present some remarks which may be significant.

Remarks:

- $\underline{\tilde{\Psi}}$ is BLUE for $\underline{\Psi}$ if and only if, whatever the unbiased linear estimator for $\underline{\Psi}$, $\underline{\Psi}^*$, the difference $V\left(\underline{\Psi}^*\right) - V\left(\underline{\tilde{\Psi}}\right)$ of their variance-covariance matrices is positive semi-definite.
- If V depends on parameters, say variance components, the condition $TV = VT$ is assumed to hold for all possible choices of these parameters.

Theorem 2.2. (Gauss-Markov): *If the model has mean vector $\underline{\mu} = X\underline{\beta}$ and variance-covariance matrix V that commute with the OPM T the LSE of estimable vectors are the best linear unbiased estimators, BLUE.*

Proof. According to the Corollary 2.6, given the unbiased estimator $\underline{\Psi}^* = U\underline{Y}$ of $\underline{\Psi}$, the LSE of $\underline{\Psi}$ will be $\underline{\tilde{\Psi}} = UT\underline{Y}$. Now, with $V\left(\underline{\Psi}^*\right)$ and $V\left(\underline{\tilde{\Psi}}\right)$ the variance-covariance matrices of these estimators and $T^C = I_n - T$, we have, since V and T commute and $TT^C = T^C T = 0$,

$$
\begin{aligned}
V(\underline{\Psi}^*) &= UVU^T = U(T+T^C)V(T+T^C)U^T \\
&= UTVTU^T + UTVT^CU^T + UT^CVTU^T + UT^CVT^CU^T \\
&= UTVTU^T + UTT^CVU^T + UT^CTVU^T + UT^CVT^CU^T \\
&= UTVTU^T + UT^CVT^CU^T = V(\underline{\tilde{\Psi}}) + UT^CVT^CU^T
\end{aligned}
$$

The thesis now follows from $UT^CVT^CU^T$ being positive semi-definite and from, according to Lemma 2.2, getting the same LSE, $UT\underline{Y}$, whatever the unbiased estimator $U\underline{Y}$ for $\underline{\Psi}$.

□

Given $\underline{Q} = \{Q_1,\ldots,Q_m\} = pb(\mathcal{A})$, when $T \in \mathcal{A}$ we can reorder the matrices in $\underline{Q}$ to get, for $z < m$,

$$
\begin{cases} T = \sum_{j=1}^{z} Q_j \\ V = \sum_{j=1}^{m} \gamma_j Q_j \end{cases}, \qquad (2.2.5)
$$

when V is known up to the $\gamma_1,\ldots,\gamma_m$, these will be the canonic variance components.
When the row vectors of A_j constitute an orthonormal basis to $\nabla_j = R(Q_j)$, we will have

$$
\begin{cases} A_j^T A_j = Q_j & , j = 1,\ldots m \\ A_j A_j^T = I_{g_j} & , j = 1,\ldots,m \end{cases}, \qquad (2.2.6)
$$

with $g_j = rank(Q_j) = rank(A_j) = dim(\nabla_j)$, $j = 1,\ldots,m$, see Silvey(1975). Taking

$$
\begin{cases} \underline{\tilde{\eta}}_j = A_j \underline{Y} & , j = 1,\ldots,m \\ \underline{\eta}_j = A_j \underline{\mu} & , j = 1,\ldots,m \\ S_j = \|Q_j \underline{Y}\|^2 = \|A_j \underline{Y}\|^2 = \|\underline{\tilde{\eta}}_j\|^2 & , j = 1,\ldots,m \end{cases}, \qquad (2.2.7)
$$

$\underline{\eta}_j$ will be the mean vector of $\underline{\tilde{\eta}}_j$, $E(\underline{\tilde{\eta}}_j)$, $j = 1,\ldots,m$. Moreover $\underline{\eta}_j = 0$ and S_j has mean vector $E(S_j) = g_j \gamma_j$, $j = z+1,\ldots,m$. Thus we have the unbiased estimator $\underline{\tilde{\eta}}_j$, $j = 1,\ldots,m$, for the $\underline{\eta}_j$, $j = 1,\ldots,m$ and

$$
\tilde{\gamma}_j = \frac{S_j}{g_j}, \; j = z+1,\ldots,m . \qquad (2.2.8)
$$

Later on we will consider the estimation of the γ_j, $j = 1,\ldots,z$. The estimation of the usual variance components $\sigma_1^2,\ldots,\sigma_w^2$ may present some problem. For instance, see Khuri et al (1998), in a three factors random effects balanced model in which a first factor crosses with a second one that nests a third one there is no unbiased estimator for the variance components associated to the second factor given by the difference of two mean squares. Actually unbiased estimators, for the variance components of balanced cross-nested models, given by linear combinations of mean squares were obtained in Fonseca et al (2003).

Returning to estimable vectors, since

$$\underline{\mu} = T\underline{\mu} = \left(\sum_{j=1}^{z} A_j^T A_j\right)\underline{\mu} = \sum_{j=1}^{z} A_j^T \underline{\eta}_j \quad , \; j = 1,\dots,z, \tag{2.2.9}$$

we get

$$\underline{\Psi} = G\underline{\beta} = UX\underline{\beta} = U\underline{\mu} = \sum_{j=1}^{z} C_j \underline{\eta}_j \quad , \text{with } C_j = UA_j^T \; , \; j = 1,\dots,z, \tag{2.2.10}$$

thus the estimable vectors will be generalized linear combinations, $\sum_{j=1}^{z} C_j \underline{\eta}_j$, of the canonical estimable vectors $\underline{\eta}_1,\dots,\underline{\eta}_z$. We point out that $\underline{\tilde{\eta}}_1,\dots,\underline{\tilde{\eta}}_z$ are linear unbiased estimators of $\underline{\eta}_1,\dots,\underline{\eta}_z$, which are thus estimable. Moreover

$$A_j T = A_j \sum_{j'=1}^{z} A_{j'}^T A_{j'} = A_j \; , \; j = 1,\dots,z \tag{2.2.11}$$

since $A_j A_{j'}^T = 0$, with $j \neq j'$, $A_j A_j^T = I_{g_j}$, $j = 1,\dots,z$. Then the $\underline{\tilde{\eta}}_j$ are the LSE of $\underline{\eta}_j$. Besides this

$$\underline{\tilde{\Psi}} = G\underline{\tilde{\beta}} = UX\underline{\tilde{\beta}} = UT\underline{Y} = U\sum_{j=1}^{z} A_j^T A_j \underline{Y} = \sum_{j=1}^{z} C_j \underline{\tilde{\eta}}_j \, , \tag{2.2.12}$$

with the same matrix coefficients $C_1,\dots,C_z$ that we have when we write $\underline{\Psi}$ as a generalized linear combination of the $\underline{\eta}_1,\dots,\underline{\eta}_z$.

2.2.2. Commutativity

We now obtain a general condition for the orthogonal projection matrix on the space spanned by the mean vector, T, to commute with the variance-covariance matrix of $\underline{Y}$, V. This is a sufficient condition for a model to be an error orthogonal model, as we will see later.

Let us consider a mixed model, which will be studied in Chapter 3,

$$\underline{Y} = \sum_{i=0}^{w} X_i \underline{\beta}_i \ , \qquad (2.2.13)$$

where $\underline{\beta}_0$ is fixed and the $\underline{\beta}_1,\ldots,\underline{\beta}_w$ are random vectors with null mean vectors.

The mean vector of $\underline{Y}$ will be

$$\underline{\mu} = X_0 \underline{\beta}_0 \ . \qquad (2.2.14)$$

Assuming the rows of X_0 to correspond to the sets of levels of the fixed effects factors, the mean values of the observations will be determined by those sets. Let us consider that there will be n^o sets of the levels associated to $r_1,\ldots,r_{n^o}$, contiguous rows of X_0. If the components of $\underline{\beta}_0$, $\beta_{0,1},\ldots,\beta_{0,n^o}$, are the corresponding mean values, we can reorder the observations to have the block diagonal matrix

$$X_0 = D\left(1_{r_1} \ \ldots \ 1_{r_{n^o}}\right). \qquad (2.2.15)$$

So the orthogonal projection matrix on the range space , Ω , spanned by the mean vector, is given by

$$T = D\left(\frac{1}{r_1} J_{r_1} \ \ldots \ \frac{1}{r_{n^o}} J_{r_{n^o}}\right), \qquad (2.2.16)$$

where $J_r = 1_r 1_r^T$.

The fundamental partition of $\underline{Y}$ will be constituted by the sub-vectors $\underline{Y}_1,\ldots,\underline{Y}_{n^o}$, corresponding to the n^o sets of levels for the fixed effects factors. Then the variance-covariance matrix can be defined by

$$V = \begin{bmatrix} V_{1,1} & \cdots & V_{1,n^o} \\ \vdots & & \vdots \\ V_{n^o,1} & \cdots & V_{n^o,n^o} \end{bmatrix}, \qquad (2.2.17)$$

with $V_{l,l}$ the variance-covariance matrix of $\underline{Y}_l$, $l = 1,\ldots,n^o$, and $V_{l,h}$ the cross-covariance matrix of Y_l and Y_h, $l \neq h$.

So, since

$$\begin{cases} TV = \begin{bmatrix} \frac{1}{r_1}J_{r_1}V_{1,1} & \cdots & \frac{1}{r_1}J_{r_1}V_{1,n^o} \\ \vdots & & \vdots \\ \frac{1}{r_{1n^o}}J_{r_{n^o}}V_{n^o,1} & \cdots & \frac{1}{r_{1n^o}}J_{r_{n^o}}V_{n^o,n^o} \end{bmatrix} \\ \\ VT = \begin{bmatrix} V_{1,1}\frac{1}{r_1}J_{r_1} & \cdots & V_{1,n^o}\frac{1}{r_{1n^o}}J_{r_{n^o}} \\ \vdots & & \vdots \\ V_{n^o,1}\frac{1}{r_1}J_{r_1} & \cdots & V_{n^o,n^o}\frac{1}{r_{1n^o}}J_{r_{n^o}} \end{bmatrix} \end{cases} \tag{2.2.18}$$

the matrices T and V commute, this is,

$$TV = VT$$

if and only if

$$\begin{cases} \frac{1}{r_1}J_{r_1}V_{1,1} = V_{1,1}\frac{1}{r_1}J_{r_1} & \cdots & \frac{1}{r_1}J_{r_1}V_{1,n^o} = V_{1,n^o}\frac{1}{r_{1n^o}}J_{r_{n^o}} \\ \vdots & & \vdots \\ \frac{1}{r_{1n^o}}J_{r_{n^o}}V_{n^o,1} = V_{n^o,1}\frac{1}{r_1}J_{r_1} & \cdots & \frac{1}{r_{1n^o}}J_{r_{n^o}}V_{n^o,n^o} = V_{n^o,n^o}\frac{1}{r_{1n^o}}J_{r_{n^o}} \end{cases}. \tag{2.2.19}$$

According to Lemma 2.1, establish in Section 2.1.2, and to the conditions for having $TV = VT$ we get the following proposition.

Proposition 2.18. *Matrices* T *and* V *commute when and only when the matrices* $V_{l,h}$, $l = 1,\dots,n^o$, $h = 1,\dots,n^o$ *are* $\mathcal{B}$ *-matrices.*

When $V = D\left(\sigma_1^2 I_{r_1} \; \dots \; \sigma_{1^0}^2 I_{r_{n^o}}\right)$ we clearly have $TV = VT$ and the LSE of estimable vectors will be BLUE.

2.2.3. An example of a balanced mixed model

We now consider a balanced mixed model, this is, a model with equal numbers of observations for all combinations of levels of the factors. Let

$$y_{i,j} = a_1 + a_2 x_i + a_3 x_i^2 + \gamma_j + \varepsilon_{i,j} \quad i = 1,\dots,n_1 \;,\; j = 1,\dots,n_2 \tag{2.2.20}$$

be the observations, where the $x_1,\dots,x_{n_1}$ are known and fixed and the $\gamma_1,\dots,\gamma_{n_2}$ are independent, with null mean value and variance σ_1^2 ,independent from $\varepsilon_{i,j}$, $i = 1,\dots,n_1$,

$j=1,\ldots,n_2$. The $\varepsilon_{1,1},\cdots,\varepsilon_{n_1,n_2}$ are themselves independent with null mean values and variance σ^2.

Ordering the observations, $y_{i,j}$, and the errors, $\varepsilon_{i,j}$, according to the indexes

$$l=(i-1)n_2+j \qquad i=1,\ldots,n_1 \ , \quad j=1,\ldots,n_2$$

we have

$$\begin{aligned}
y_1 &= a_1+a_2x_1+a_3x_1^2+\gamma_1+\varepsilon_{1,1}\\
&\vdots\\
y_{n_2} &= a_1+a_2x_1+a_3x_1^2+\gamma_{n_2}+\varepsilon_{1,n_2}\\
y_{n_2+1} &= a_1+a_2x_2+a_3x_2^2+\gamma_1+\varepsilon_{2,1}\\
&\vdots\\
y_{2n_2} &= a_1+a_2x_2+a_3x_2^2+\gamma_{n_2}+\varepsilon_{2,n_2}\\
&\vdots\\
y_{(n_1-1)n_2+1} &= a_1+a_2x_{n_1}+a_3x_{n_1}^2+\gamma_1+\varepsilon_{n_1,1}\\
&\vdots\\
y_{n_1n_2} &= a_1+a_2x_{n_1}+a_3x_{n_1}^2+\gamma_{n_2}+\varepsilon_{n_1,n_2}
\end{aligned}$$

so we can rewrite the model as

$$\underline{Y}=X_0\underline{\beta}+X_1\underline{\gamma}+\underline{\varepsilon} \ , \tag{2.2.21}$$

where

$$\begin{cases} X_0=I_{n_1}\otimes 1_{n_2} \\ X_1=1_{n_1}\otimes I_{n_2} \end{cases} , \tag{2.2.22}$$

the vector $\underline{\beta}$ has components

$$\beta_l=a_1+a_2x_i+a_3x_i^2 \quad , \ i=1,\ldots,n_1 \ , \tag{2.2.23}$$

and the components of the vectors $\underline{\gamma}$ and $\underline{\varepsilon}$ are , respectively $\gamma_1,\ldots,\gamma_{n_2}$ and $\varepsilon_{1,1},\cdots,\varepsilon_{n_1,n_2}$.

The variance-covariance matrix of $\underline{Y}$ is

$$V=\sigma_1^2J_{n_1}\otimes I_{n_2}+\sigma^2 I_{n_1n_2}=\sigma_1^2J_{n_1}\otimes I_{n_2}+\sigma^2 I_{n_1}\otimes I_{n_2} \ , \tag{2.2.24}$$

with sub-matrices

$$\begin{cases} V_{i,i} = \sigma_1^2 \; I_{n_2} + \sigma^2 I_{n_2} & i = 1, \ldots, n_1 \\ V_{i,i'} = \sigma_1^2 \; I_{n_2} & i \neq i' \end{cases},$$

which are clearly $\mathscr{B}$ -matrices.

Thus

$$\underline{\tilde{\beta}} = \left(X_0^T X_0\right)^{-1} X_o^T \underline{Y} \tag{2.2.25}$$

will be BLUE. It is easy to see that the components of $\underline{\tilde{\beta}}$ are

$$\tilde{\beta}_i = \frac{1}{n_2} \sum_{j=1}^{n_2} y_{(i-1)n_2+j} \; , \; i = 1, \ldots, n_1 .$$

Moreover, we have

$$\underline{\beta} = X \, \underline{a} \; , \tag{2.2.26}$$

with

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{n_1} & x_{n_1}^2 \end{bmatrix}$$

and so, from (2.2.25), we get the

$$\underline{\tilde{a}} = \left(X^T X\right)^{-1} X^T \underline{\tilde{\beta}}$$

which will be also BLUE, whenever $\text{rank}(X) = 3$.

These models can be extended in many ways, for instance increasing the degree of the polynomial regressions. In this way we obtain Error orthogonal models. These models will be studied in the next chapter. Another possible extension is to unbalance mixed models with a regressional fixed effects part.

2.2.4. Sufficient and complete statistics

Let the sample space $\mathcal{E}$ be the set of all possible n-dimensional samples. The n-dimensional samples, represented by $\underline{x}$, must be considerate as the realizations of random vector $\underline{X}$. The space $\mathcal{E}$ is the support of the distribution of the observations vectors. We will assume this distribution to be known up to a parameter vector $\underline{\theta}$ and it is denoted by $F(\cdot \mid \underline{\theta})$, while $f(\cdot \mid \underline{\theta})$ will be the corresponding density [probability] function in the continuous [discrete] case. Given a partition,

$$\mathcal{E}=\bigcup_{i=1} C_i \tag{2.2.27}$$

of the sample space, $\mathcal{E}$, we have the conditional distribution

$$F\left(\underline{x} \mid C_i;\underline{\theta}\right)=\Pr\left(\underline{X}\leq\underline{x} \mid \left(\underline{X}\in C_i\right);\underline{\theta}\right) \quad , \tag{2.2.28}$$

thus, according to the total probability theorem we have

$$\begin{aligned} F(\underline{x} \mid \underline{\theta}) &= \sum_i \Pr\left(\underline{X}\in C_i \mid \underline{\theta}\right)\Pr\left(\underline{X}\leq\underline{x} \mid \left(\underline{X}\in C_i\right);\underline{\theta}\right) \\ &= \sum_i \Pr\left(\underline{X}\in C_i \mid \underline{\theta}\right)F\left(\underline{x} \mid C_i;\underline{\theta}\right) \end{aligned} \quad . \tag{2.2.29}$$

Following we present some important definitions.

Definition 2.19. The partition is sufficient if the conditional distribution do not depend on $\underline{\theta}$, thus

$$F(\underline{x} \mid \underline{\theta})=\sum_i \Pr\left(\underline{X}\in C_i \mid \underline{\theta}\right)F\left(\underline{x} \mid C_i\right) \ .$$

Definition 2.20. A sufficient partition is minimal if any set of whatever sufficient partition is contained in a set of that partition. Namely, a sufficient partition is minimal if its sets are unions of sets of whatever sufficient partition.

Definition 2.21. A statistic is a scalar or vector function defined on the sample space that does not depend on any unknown parameter.

It is important to note that any statistic, being a function with domain $\mathcal{E}$, generates a partition of $\mathcal{E}$. If that partition is sufficient the statistic is sufficient. If the partition is minimal sufficient so is the statistic.

Sufficient statistics summarizes the whole of the relevant information in a sample, about $\underline{\theta}$, but to achieve the maximum possible data reduction, without any loss of information, it is desirable to have a statistic as condensed as possible, that is, a minimal sufficient statistic.

Definition 2.22. A sufficient statistic, $m(\underline{x})$, is minimal if and only if for every other sufficient statistic , $n(\underline{x})$, there exists a function, h , such that

$$m(\underline{x})=h(n(\underline{x})) \ .$$

Now we have, see Fraser (1957), the

Theorem 2.3. (Factorization) *A statistic,* $T(\underline{x})$, *is sufficient if and only if there are two non-negatives functions* $g(\underline{x})$ *and* $h(\underline{x})$ *such that*

$$f(\underline{x} \mid \underline{\theta}) = g(T(\underline{x}) \mid \underline{\theta})\, h(\underline{x})$$

where $h(\underline{x})$ *does not depend on* $\underline{\theta}$.

Proof. The proof that we present is for the discrete case. A general proof may be found in Fraser(1957).

Suppose that $f(\underline{x} \mid \underline{\theta}) = g(T(\underline{x}) \mid \theta)\, h(\underline{x})$. With $T(\underline{x}) = t$ we have, for all $\underline{x}$ and for all $\underline{\theta}$,

$$\Pr(T(\underline{X}) = t \mid \underline{\theta}) = \sum_{\underline{x} \in T^{-1}(t)} f(\underline{x} \mid \underline{\theta}) = \sum_{\underline{x} \in T^{-1}(t)} g(T(\underline{x}) \mid \theta)\, h(\underline{x}) = g(t \mid \theta) \sum_{\underline{x} \in T^{-1}(t)} h(\underline{x}) .$$

So, for $\underline{x} \in T^{-1}(t)$ and since $\{X = x\} \subseteq \{T(X) = t\}$,

$$\Pr(\underline{X} = \underline{x} \mid T(\underline{X}) = t\,;\theta) = \frac{\Pr[(\underline{X} = \underline{x}) \cap (T(\underline{X}) = t) \mid \theta]}{\Pr(T(\underline{X}) = t \mid \theta)} = \frac{\Pr(\underline{X} = \underline{x} \mid \theta)}{\Pr(T(\underline{X}) = t \mid \theta)} =$$
$$= \frac{g(t \mid \theta)\, h(\underline{x})}{g(t \mid \theta) \sum_{\underline{x}' \in T^{-1}(t')} h(\underline{x}')} = \frac{h(\underline{x})}{\sum_{\underline{x}' \in T^{-1}(t')} h(\underline{x}')} ,$$

which not depend on $\underline{\theta}$, hence T is a sufficient statistic. Conversely, if we assume that $T(\underline{x})$ is a sufficient statistic, which mean that the conditional distribution of $\underline{X}$, given $T(\underline{x}) = t$, does not depend on $\underline{\theta}$, so

$$\Pr(\underline{X} = \underline{x} \mid T(\underline{x}) = t\,;\theta) = P(\underline{X} = \underline{x} \mid T(\underline{x}) = t) = h(\underline{x})$$

for all $\underline{\theta}$ in the parameter space. Since,

$$\{\underline{X} = \underline{x}\} \subseteq \{T(\underline{x}) = t\}$$

we have

$$f(\underline{x} \mid \underline{\theta}) = \Pr(\underline{X} = \underline{x} \mid \underline{\theta}) = \Pr((\underline{X} = \underline{x}) \cap (T(\underline{x}) = t)) =$$
$$= \Pr(T(\underline{x}) = t \mid \underline{\theta}) \Pr(\underline{X} = \underline{x} \mid T(\underline{x}) = t)$$
$$= g(T(\underline{x}) \mid \underline{\theta})\, h(\underline{x}).$$

□

Sufficient statistics have interesting properties connected to unbiased estimators. To establish these properties we need the notion of convex function.

Definition 2.23. A function $c(\cdot)$ is convex if, whatever u_1 and u_2 in its domain and $a \in [0,1]$, we have

$$c(au_1 + (1-a)u_2) \le a\,c(u_1) + (1-a)c(u_2) \ .$$

Let us now establish

Theorem 2.4. (Jensen´s inequality) *Given the random variable* Y *and a convex function* $c(\cdot)$ *we have*

$$E(c(Y)) \ge c(E(Y))$$

whenever, the mean values of $c(Y)$ *and of* Y , $E(c(Y))$ *and* $E(Y)$ *are defined.*

Proof. The proof we present is for the continuous case, a general proof may be found in Fraser (1957).

Let $r(Y) = a + bY$ be the straight line tangent to the graph of $c(\cdot)$ at $\mu = E(Y)$. Since $c(\cdot)$ is a convex function, we have $c(Y) \ge a + bY$ and so $E(c(Y)) \ge a + bE(Y) = r(E(Y))$. Since $r(Y)$ is tangent to $c(\cdot)$ at $\mu = E(Y)$, we have $r(E(Y)) = c(E(Y))$, thus $E(c(Y)) \ge c(E(Y))$, and the thesis is established.

□

We now get the

Theorem 2.5. (Rao-Blackwell) *Let* $l^*(\underline{X})$ *be an unbiased estimator of* $l(\underline{\theta})$ *with* $l(\cdot)$ *a bounded function. Given a sufficient statistic* $\underline{T}(\underline{X})$,

$$\tilde{l}(t) = E\left(l^*(\underline{X}) \mid \underline{T}(\underline{X}) = t \right)$$

(i) $\tilde{l}(t)$ *is function of* t *but not of* $\underline{\theta}$;

(ii) $\tilde{l}(t)$ *is an unbiased estimator of* $l(\underline{\theta})$;

(iii) $V\left(\tilde{l}\right) \le V\left(l^*\right)$.

Proof.

(i) Since T is a sufficient statistic it generates a sufficient partition of $\mathcal{E}$, thus $\tilde{l}(t)$ will be function of t but not of $\underline{\theta}$.

(ii) $l(\underline{\theta}) = E\left(l^*(\underline{X})\right) = E\left(E\left(l^*(\underline{X}) \mid T(\underline{X}) = t\right) \mid \underline{\theta}\right) = E\left(\tilde{l}(\underline{\theta}) \mid \underline{\theta}\right)$, so $\tilde{l}(t)$ is an unbiased estimator of $l(\theta)$.

(iii) Consider the convex function $c(X) = (X - E(X))^2$. Since the variance is the mean value of this function, by Jensen's inequality, we have

$$\begin{aligned} V\left(\tilde{l}(t) \mid \theta\right) &= E\left(c\left(\tilde{l}(t) \mid \theta\right)\right) = E\left(c\left(E\left(l^*(\underline{X}) \mid T(\underline{X}) = t\,;\theta\right)\right)\right) \\ &\le E\left(E\left(c\left(l^*(\underline{X}) \mid T(\underline{X}) = t\,;\theta\right)\right)\right) = E\left(c\left(l^*(\underline{X}) \mid \theta\right)\right) = V\left(l^*(\underline{X}) \mid \theta\right) \end{aligned}$$

□

Besides sufficient we are interested in complete statistics.

Definition 2.24. A statistic $T(\underline{x})$ is complete for $\underline{\theta} \in \Theta$, with Θ the parametric space spanned by $\underline{\theta}$, if

$$\forall \underline{\theta} \in \Theta : E\left(l(t) \mid \underline{\theta}\right) = 0 \ ,$$

that implies

$$\forall \underline{\theta} \in \Theta : \mathrm{pr}\left(l(t) = 0\right) = 1 .$$

Using jointly the concepts of sufficiency and completeness for statistics, we get the

Theorem 2.6. (Blackwell-Lehman-Sheffé) *If* $T(\underline{x})$ *is a complete and sufficient statistic, and there exists* l^**, an unbiased estimator of* $l(\theta)$ *with* $l(\cdot)$ *a bounded function,*

$$\tilde{l}(t) = E\left(l^*(\underline{X}) \mid T(\underline{X}) = t\right)$$

is an unbiased estimator of $l(\underline{\theta})$ *with uniformly minimum variance, UMVUE.*

Proof. According to the Rao-Blackwell theorem, $\tilde{l}$ is an unbiased estimator of $l(\theta)$ and $V(\tilde{l}) \le V(l^*)$. Given another unbiased estimator ${}^*l$ of $l(\theta)$ we could use again the Rao-Blackwell theorem to show that $V({}^*\tilde{l}) \le V({}^*l)$, with ${}^*\tilde{l} = E\left({}^*l(\underline{X}) \mid T(\underline{X}) = t\right)$. But since $\underline{T}$ is complete and $\tilde{l}$ and ${}^*\tilde{l}$ are unbiased estimators of $l(\theta)$ we will have $\forall \underline{\theta} \in \Theta : \mathrm{pr}\left({}^*\tilde{l} = \tilde{l}\right) = 1$, which establishes the thesis.

□

Let s be the number of components of $\underline{\theta}$, if Θ contains the Cartesian product of s non-degenerated intervals and $f(x \mid \underline{\theta})$ belongs to the exponential family, this is, if

$$f(\underline{x} \mid \underline{\theta}) = \exp\left[\sum_{i=1}^{s} \nu_i(\underline{\theta}) T_i(\underline{x}) - B(\underline{\theta})\right] h(\underline{x}) \quad . \tag{2.2.30}$$

the statistic $\underline{T}(\underline{x})$, with components $T_i(\underline{x})$, $i = 1,\ldots,s$, will be complete, see Silvey (1975, pg 31).

If

$$\underline{\theta} = \begin{bmatrix} \underline{\theta}_1 \\ \underline{\theta}_2 \end{bmatrix}$$

where $\underline{\theta}_1$ and $\underline{\theta}_2$ have s_1 and s_2 components, respectively, and $\underline{\theta}_2 \in \nabla$, with $\dim(\nabla) < s_2$, Θ will not contain the product of $s = s_1 + s_2$ non degenerate intervals and $\underline{T}$ will not be complete.

If there is no linear constraints on the $\nu_1,\ldots,\nu_s$ or on the $T_1,\ldots,T_s$ and moreover Θ contains the Cartesian product of s non degenerate intervals, $f(\underline{x} \mid \underline{\theta})$ will belong , see Lehmann & Casella (1998, pg24), to a full rank exponential family and , see Lehmann & Casella (1998, pg39) will be minimal sufficient and complete statistic.

2.3. Normal vectors

The purpose of this section is to present some important results about normal vectors. First we introduce moments and moment generating functions of random vectors in general. We turn our attention to the particular case of normal vectors when we introduce the linear transformations.

2.3.1. Moments and generating functions.

In this sub-section we will follow Mexia (1995) were can be found the proofs not included in this text.

As usual in Linear Algebra, we will write vectors as column matrices whenever convenient.

Definition 2.25. Given a random vector

$$\underline{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_r \end{bmatrix} ,$$

the mean vector of X is

$$\underline{\mu} = E(\underline{X}) = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_r) \end{bmatrix} ,$$

whenever the mean values $E(X_1),\ldots,E(X_r)$ of the random variables $X_1,\ldots,X_r$ are defined.

Let $\underline{X}$ be a random vector, A a constant matrix and $\underline{b}$ a constant vector. Since the operator $E(\cdot)$ is linear, we will have

$$E(A\underline{X} + \underline{b}) = A\,E(\underline{X}) + \underline{b} \ . \tag{2.3.1}$$

Let us define

Definition 2.26. Given an $r \times s$ random matrix, $\underline{\underline{X}} = \left[X_{i,j}\right]$, the mean matrix of $\underline{\underline{X}}$ is

$$E(\underline{\underline{X}}) = \left[E(X_{i,j})\right],$$

whenever the mean values of the random variables $X_{i,j}$, $E(X_{i,j})$, $i = 1,\ldots,r$, $j = 1,\ldots,s$, are defined.

Let $\underline{\underline{X}}$ be a random matrix and A and B be constant matrices, once more due to the linearity of the operator $E(\cdot)$, we have

$$E(A\underline{\underline{X}}B) = A\,E(\underline{\underline{X}})B\,. \tag{2.3.2}$$

Proposition 2.19. *The variance-covariance matrix (or simply, covariance matrix) of a random vector $\underline{X}$ is given by*

$$\begin{aligned} V(\underline{X}) &= E\left[(\underline{X} - E(\underline{X}))(\underline{X} - E(\underline{X}))^T\right] \\ &= \begin{bmatrix} V(X_1) & \cdots & COV(X_1,X_r) \\ \vdots & \ddots & \vdots \\ COV(X_r,X_1) & \cdots & V(X_r) \end{bmatrix} , \end{aligned}$$

which is defined whenever the variances, $V(X_i)$ $i = 1,\ldots,r$, and the covariances, $COV(X_i,X_j)$, $i \neq j$, $i,j = 1,\ldots,r$, of the components of $\underline{X}$ are defined.

Proof.

$$
\begin{aligned}
V(\underline{X}) &= E\left[(\underline{X}-E(\underline{X}))(\underline{X}-E(\underline{X}))^T\right] \\
&= E\left(\begin{bmatrix} X_1-E(X_1) \\ \vdots \\ X_r-E(X_r) \end{bmatrix}[X_1-E(X_1)\ \ldots\ X_r-E(X_r)]\right) \\
&= E\left(\begin{bmatrix} (X_1-E(X_1))^2 & \cdots & (X_1-E(X_1))(X_r-E(X_r)) \\ \vdots & \ddots & \vdots \\ (X_r-E(X_r))(X_1-E(X_1)) & \cdots & (X_r-E(X_r))^2 \end{bmatrix}\right) \\
&= \begin{bmatrix} E(X_1-E(X_1))^2 & \cdots & E[(X_1-E(X_1))(X_r-E(X_r))] \\ \vdots & \ddots & \vdots \\ E[(X_r-E(X_r))(X_1-E(X_1))] & \cdots & E(X_r-E(X_r))^2 \end{bmatrix} \\
&= \begin{bmatrix} V(X_1) & \cdots & COV(X_1,X_r) \\ \vdots & \ddots & \vdots \\ COV(X_r,X_1) & \cdots & V(X_r) \end{bmatrix}
\end{aligned}
$$

□

Proposition 2.20. *Let* $\underline{X}$ *be a random vector,* A *a constant matrix and* $\underline{b}$ *a constant vector*

$$V(A\underline{X}+\underline{b}) = A\,V(\underline{X})\,A^T \ .$$

Proof.

$$
\begin{aligned}
V(A\underline{X}+\underline{b}) &= E\left[((A\underline{X}+\underline{b})-E(A\underline{X}+\underline{b}))((A\underline{X}+\underline{b})-E(A\underline{X}+\underline{b}))^T\right] \\
&= E\left[(A(\underline{X}-E(\underline{X})))(A(\underline{X}-E(\underline{X})))^T\right] \\
&= E\left[A(\underline{X}-E(\underline{X}))(\underline{X}-E(\underline{X}))^T A^T\right] \\
&= A\,E\left[(\underline{X}-E(\underline{X}))(\underline{X}-E(\underline{X}))^T\right]A^T \\
&= A\ V(\underline{X})\,A^T
\end{aligned}
$$

□

Proposition 2.21. *The variance-covariance matrix,* $V(\underline{X})$*, is a symmetric matrix.*

Proof.

$$
\begin{aligned}
V(\underline{X})^T &= \left[E\left[(\underline{X}-E(\underline{X}))(\underline{X}-E(\underline{X}))^T\right]\right]^T \\
&= E\left[\left((\underline{X}-E(\underline{X}))(\underline{X}-E(\underline{X}))^T\right)^T\right] \\
&= E\left[(\underline{X}-E(\underline{X}))(\underline{X}-E(\underline{X}))^T\right] \\
&= V(\underline{X})
\end{aligned}
$$

□

Proposition 2.22. *Given two random vectors*

$$\underline{X}_1 = \begin{bmatrix} X_{1,1} \\ \vdots \\ X_{1,k_1} \end{bmatrix} \quad \textit{and} \quad \underline{X}_2 = \begin{bmatrix} X_{2,1} \\ \vdots \\ X_{2,k_2} \end{bmatrix} \ ,$$

the pair $(\underline{X}_1, \underline{X}_2)$ *has cross-covariance matrix*

$$V(\underline{X}_1, \underline{X}_2) = \begin{bmatrix} \operatorname{cov}(X_{1,1} X_{2,1}) & \cdots & \operatorname{cov}((X_{1,1} X_{2,k_2})) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_{1,k_1} X_{2,1}) & \cdots & \operatorname{cov}(X_{1,k_1} X_{2,k_2}) \end{bmatrix} .$$

Proof.

$$V(\underline{X}_1, \underline{X}_2) = E\left((\underline{X}_1 - E(\underline{X}_1))\,(\underline{X}_2 - E(\underline{X}_2))^T\right) = \begin{bmatrix} \operatorname{cov}(X_{1,1} X_{2,1}) & \cdots & \operatorname{cov}((X_{1,1} X_{2,k_2})) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_{1,k_1} X_{2,1}) & \cdots & \operatorname{cov}(X_{1,k_1} X_{2,k_2}) \end{bmatrix} .$$

□

Let A_1 and A_2 be constant matrices and $\underline{b}_1$ and $\underline{b}_2$ constant vectors, we get

$$V\,(A_1 \underline{X}_1 + \underline{b}_1 \,;\, A_2 \underline{X}_2 + \underline{b}_2) = A_1 V(\underline{X}_1 \,;\, \underline{X}_2) A_2^T .$$

It is also easy to see that

$$V\,(\underline{X}_1, \underline{X}_2) = V\,(\underline{X}_2, \underline{X}_1)^T . \tag{2.3.3}$$

If $\underline{X}_1$ and $\underline{X}_2$ are independent random vectors, we put $\underline{X}_1\,(i)\,\underline{X}_2$, the covariances between the components of both vectors will be null and so, with k_1 and k_2 the number of components of $\underline{X}_1$ and $\underline{X}_2$ respectively,

$$V\,(\underline{X}_1; \underline{X}_2) = 0_{k_1 \times k_2} \ , \tag{2.3.4}$$

where $0_{k_1 \times k_2}$ is the $k_1 \times k_2$ null matrix.

Whatever the random vectors $\underline{X}_1$ and $\underline{X}_2$,

$$V\left(\begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}\right) = \begin{bmatrix} V(\underline{X}_1) & V(\underline{X}_1; \underline{X}_2) \\ V(\underline{X}_2, \underline{X}_1) & V(\underline{X}_2) \end{bmatrix} . \tag{2.3.5}$$

If $\underline{X}_1\,(i)\,\underline{X}_2$, both with k components, we have

$$\underline{X}_1+\underline{X}_2=\begin{bmatrix} I_k & I_k \end{bmatrix}\begin{bmatrix}\underline{X}_1\\ \underline{X}_2\end{bmatrix}$$

as well as

$$V(\underline{X}_1+\underline{X}_2)=\begin{bmatrix} I_k & I_k \end{bmatrix}\begin{bmatrix} V(\underline{X}_1) & 0_{k\times k}\\ 0_{k\times k} & V(\underline{X}_2)\end{bmatrix}\begin{bmatrix} I_k\\ I_k\end{bmatrix}=V(\underline{X}_1)+V(\underline{X}_2) \quad , \qquad (2.3.6)$$

where $0_{k\times k}$ is the k-order null matrix and I_k the identity matrix of order k.

To obtain the moments their generating function is highly useful.

Definition 2.27. Given a random vector $\underline{X}$, whose components are the random variables $X_1,\ldots,X_r$, his moment generating function is given by

$$\varphi_{\underline{X}}(\underline{u})=E\left(e^{\underline{u}^T\underline{X}}\right) \ .$$

When it is defined in an open set containing the origin, it is indefinitely derivable, at the origin, and

$$\left(\frac{\partial^r\varphi(\underline{u})}{\partial u_1^{r_1}\ldots\partial u_k^{r_k}}\right)_{\underline{u}=\underline{0}}=E\left(X_1^{r_1}\ldots X_k^{r_k}\right)=\mu'_{r_1,\ldots,r_k} \ ,$$

with $r=\sum_{i=1}^{k} r_i$.

Moreover if two of these function are identical, the corresponding distribution are identical.

Proposition 2.23. *Let* A *be a constant matrix and let* $\underline{b}$ *be a constant vector. The moment generating function of the random vector* $\underline{Y}=A\underline{X}+\underline{b}$ *is,*

$$\varphi_{\underline{Y}}(\underline{u})=e^{\underline{u}^T\underline{b}}\varphi_{\underline{X}}\left(A^T\underline{u}\right)$$

Proof. From Definition 2.27 and from (2.3.1) we have

$$\varphi_{\underline{Y}}(\underline{u})=\varphi_{A\underline{X}+\underline{b}}(\underline{u})=E\left(e^{\underline{u}^T(A\underline{X}+\underline{b})}\right)=e^{\underline{u}^T\underline{b}}\varphi_{\underline{X}}\left(A^T\underline{u}\right)$$

□

Besides this, see Lukacs and Laha (1964), it may be shown that $\underline{X}_1$ and $\underline{X}_2$ are independent if and only if there joint moment generating function is the product of the moment generating functions of $\underline{X}_1$ and $\underline{X}_2$.

Proposition 2.24. *If* $\underline{X}_1 \,(\mathrm{i})\, \underline{X}_2$, *both with* k *components, the moment generating function of* $\underline{X}_1 + \underline{X}_2$ *will be*

$$\varphi_{\underline{X}_1+\underline{X}_2}(\underline{u}) = \prod_{l=1}^{2} \varphi_{\underline{X}_l}(\underline{u}).$$

Proof. Since $e^{\underline{u}^T \underline{X}_1} \,(\mathrm{i})\, e^{\underline{u}^T \underline{X}_2}$, we have

$$\varphi_{\underline{X}_1+\underline{X}_2}(\underline{u}) = E\left(e^{\underline{u}^T(\underline{X}_1+\underline{X}_2)}\right) = E\left(e^{\underline{u}^T \underline{X}_1} e^{\underline{u}^T \underline{X}_2}\right) = \prod_{l=1}^{2} E\left(e^{\underline{u}^T \underline{X}_l}\right) = \prod_{l=1}^{2} \varphi_{\underline{X}_l}(\underline{u}).$$

□

If $\underline{X}_1 \,(\mathrm{i})\, \underline{X}_2$, with $\underline{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$ and $\underline{u} = \begin{bmatrix} \underline{u}_1 \\ \underline{u}_2 \end{bmatrix}$ when $\underline{X}_l$ and $\underline{u}_l$ have k_l components, we have

$$\underline{u}^T \underline{X} = \sum_{l=1}^{2} \underline{u}_l^T \underline{X}_l ,$$

with $\underline{u}_1^T \underline{X}_1 \,(\mathrm{i})\, \underline{u}_2^T \underline{X}_2$, and so

$$\varphi_{\underline{X}}(\underline{u}) = E\left(e^{\underline{u}^T(\underline{X})}\right) = E\left(e^{\underline{u}^T \underline{X}_1} e^{\underline{u}^T \underline{X}_2}\right) = E\left(\prod_{l=1}^{2} e^{\underline{u}_l^T \underline{X}_l}\right) = \prod_{l=1}^{2} E\left(e^{\underline{u}_l^T \underline{X}_l}\right) = \prod_{l=1}^{2} \varphi_{\underline{X}_l}(\underline{u}_l). \quad (2.3.7)$$

2.3.2. Linear transformations

We write $\underline{X} \sim \mathcal{N}(\,\cdot \mid \underline{\mu}, V)$ to indicate that the random vector $\underline{X}$ has normal distribution with mean vector $\underline{\mu}$ and variance-covariance matrix V.

Definition 2.28. Given $\underline{X} \sim \mathcal{N}(\,\cdot \mid \underline{\mu}, V)$, if V is invertible, $\underline{X}$ will have density

$$n(\underline{x} \mid \underline{\mu}, V) = \frac{e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T V^{-1}(\underline{x}-\underline{\mu})}}{(2\pi)^{\frac{k}{2}} \sqrt{\det(V)}} ,$$

where k is the number of components of $\underline{X}$ and $\det(V)$ is the determinant of variance-covariance matrix .

Proposition 2.25. *The moment generating function of the normal vector* $\underline{X}$ *is*

$$\varphi_{\underline{X}}(\underline{u}) = \varphi_N(\underline{u} \mid \underline{\mu}, V) = e^{\underline{\mu}^T \underline{u} + \frac{1}{2}\underline{u}^T V \underline{u}} \quad .$$

Proof. See Mexia (1995, pag37). □

Proposition 2.26. *Let* $\underline{X}$ *be a normal vector, with mean vector* $\underline{\mu}$ *and variance-covariance matrix* V. *Let* A *be a constant matrix and* $\underline{b}$ *a constant vector. The moment generating function of the normal type of* $A\underline{X}+\underline{b}$ *is*

$$\varphi_{A\underline{X}+\underline{b}}(\underline{u}) = \varphi_{\mathcal{N}}\left(\underline{u} \mid A\underline{\mu} + \underline{b}\,,\, A V A^T\right) \; .$$

Proof. Since

$$\begin{cases} E(A\underline{X}+\underline{b}) = A\underline{\mu}+\underline{b} \\ V(A\underline{X}+\underline{b}) = A\; V(\underline{X})\, A^T = A\; V\; A^T \end{cases}$$

we get, from Propositions 2.23 and 2.25,

$$\begin{aligned} \varphi_{A\underline{X}+\underline{b}}(\underline{u}) &= e^{\underline{u}^T\underline{b}}\varphi_{\underline{X}}\left(A^T\underline{u}\right) \\ &= e^{\underline{u}^T\underline{b}}\; e^{\left(A^T\underline{u}\right)^T\underline{\mu}+\frac{1}{2}\left(A^T\underline{u}\right)^T V \left(A^T\underline{u}\right)} \\ &= e^{\underline{u}^T\left(A\underline{\mu}+\underline{b}\right)+\frac{1}{2}\underline{u}^T\left(AVA^T\right)\underline{u}} \end{aligned}$$

so the moment generating function of the normal type of $A\underline{X}+\underline{b}$ is

$$\varphi_{A\underline{X}+\underline{b}}(\underline{v}) = \varphi_{\mathcal{N}}\left(\underline{v} \mid A\underline{\mu} + \underline{b}\,,\, A V A^T\right) \; .$$

□

Now $A V A^T$ may not be invertible so $A\underline{X}+\underline{b}$ will not have density. Then, in what follows, we say that $\underline{X} \sim \mathcal{N}\left(\;\cdot \mid \underline{\mu}\,,\, V\right)$ if

$$\varphi_{\underline{X}}(\underline{u}) = \varphi_{\mathcal{N}}\left(\underline{u} \mid \underline{\mu}\,,\, V\right) \quad , \qquad (2.3.8)$$

and so, the normal family of distributions will be closed for linear transformations.

If $\underline{X}_1 \,(\mathrm{i})\, \underline{X}_2$, with $\underline{X}_l \sim \mathcal{N}\left(\;\cdot \mid \underline{\mu}_l, V_l\right)$, $l = 1,2$, and

$$\underline{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix} \;;\quad \underline{u} = \begin{bmatrix} \underline{u}_1 \\ \underline{u}_2 \end{bmatrix} \;;\quad \underline{\mu} = \begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix} \;;\quad V = \begin{bmatrix} V_1 & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & V_2 \end{bmatrix}$$

we have

$$\varphi_{\underline{X}}(\underline{u}) = \prod_{l=1}^{2} \varphi_{\underline{X}_l}(\underline{u}_l) = \prod_{l=1}^{2} e^{\underline{\mu}_l^T \underline{u}_l + \frac{1}{2}\underline{u}_l^T V \underline{u}_l} = e^{\underline{\mu}^T \underline{u} + \frac{1}{2}\underline{u}^T V \underline{u}} = \varphi_{\mathcal{N}}(\underline{u} \mid \underline{\mu}, V) \tag{2.3.9}$$

so

$$\underline{X} \sim \mathcal{N}(\quad \mid \underline{\mu}, V) \ ,$$

this is, superimposing independent normal vectors gives normal vectors. Then

$$A_1 \underline{X}_1 + A_2 \underline{X}_2 = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$$

will be normal, whenever $\underline{X}_1$ and $\underline{X}_2$ are normal and independent, with

$$\begin{cases} E(A_1 \underline{X}_1 + A_2 \underline{X}_2) = A_1 \underline{\mu}_1 + A_2 \underline{\mu}_2 \\ V(A_1 \underline{X}_1 + A_2 \underline{X}_2) = A_1 V_1 A_1^T + A_2 V_2 A_2^T \end{cases} , \tag{2.3.10}$$

whenever $\underline{X}_l \sim \mathcal{N}(\quad \mid \underline{\mu}_l, V_l), l = 1, 2$.

We also have

Proposition 2.27. *If*

$$\underline{X} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}, \begin{bmatrix} V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \end{bmatrix} \right) ,$$

then $\underline{X}_1 (i) \ \underline{X}_2$ *if and only if* $V_{1,2} = 0_{k_1 \times k_2}$ *with* k_l *the number of components of* $\underline{X}_l$, $l = 1, 2$.

Proof. If $\underline{X}_1 (i) \ \underline{X}_2$ we have $V_{1,2} = 0_{k_1 \times k_2}$ since the covariance of the components of both vectors are null. Moreover if $V_{1,2} = 0_{k_1 \times k_2}$ we have $V_{2,1} = V_{1,2}^T = 0_{k_2 \times k_1}$ and, as we saw, with $\underline{u} = \begin{bmatrix} \underline{u}_1 \\ \underline{u}_2 \end{bmatrix}$, we have $\varphi_{\underline{X}}(\underline{u}) = \prod_{l=1}^{2} \varphi_{\underline{X}_l}(\underline{u}_l)$ and $\underline{X}_1 (i) \ \underline{X}_2$.

2.3.3. Associated distributions

Definition 2.29. Let $X_1, \ldots, X_n$ be n independent and identically distributed random variables with standard normal distribution, we write $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{N}(\cdot \mid 0, 1)$. The variable

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2$$

has a central chi-square distribution with n degrees of freedom.

The variable χ_n^2 has density function

$$g(x \mid n) = \begin{cases} 0 & ; \quad x \le 0 \\ \dfrac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\dfrac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} & ; \quad x > 0 \end{cases} , \qquad (2.3.11)$$

where $\Gamma(z)$ is the Gamma function,

$$\Gamma(z) = \int_0^{+\infty} e^{-t}\, t^{z-1}\, dt . \qquad (2.3.12)$$

Let us now establish

Proposition 2.28. *The moment generating function of a chi-square random variable with n degrees of freedom is*

$$\varphi_{\chi_n^2}(u) = \varphi(u \mid n) = \frac{1}{(1-2u)^{\frac{n}{2}}} ,$$

defined for any $u < \dfrac{1}{2}$.

Proof. From the definition of moment generating function of a random variable and from (2.3.11), it comes

$$\varphi(u \mid n) = \int_0^{+\infty} e^{ux} \frac{1}{2\,\Gamma\left(\frac{n}{2}\right)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}\, dx .$$

Making the transformation $v = \dfrac{1-2u}{2} x$, we get

$$\varphi(u \mid n) = \frac{1}{2\,\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{1-2u}\right)^{\frac{n}{2}-1} \left(\frac{2}{1-2u}\right) \int_0^{+\infty} e^{-v}\, v^{\frac{n}{2}-1}\, dv$$

$$= \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{1-2u}\right)^{\frac{n}{2}} \int_0^{+\infty} e^{-v}\, v^{\frac{n}{2}-1}\, dv \quad .$$

Since the last integral is the gamma function for $\dfrac{n}{2}$, $\Gamma\left(\frac{n}{2}\right)$,

$$\varphi(u \mid n) = \frac{1}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{1-2u}\right)^{\frac{n}{2}}\Gamma\left(\tfrac{n}{2}\right)$$

$$= \frac{1}{(1-2u)^{\frac{n}{2}}} .$$

□

Proposition 2.29. *If* χ_n^2 *and* χ_m^2 *are two independent chi-square random variables with* n *and* m *degrees of freedom, respectively, then* $\chi_n^2 + \chi_m^2$ *has a chi-square distribution with* $n+m$ *degrees of freedom with moment generating function*

$$\varphi_{\chi_n^2+\chi_m^2}(u) = \varphi_{\chi_{n+m}^2}(u) \quad .$$

We say that the chi-square is reproductive on the degrees of freedom.

Proof. Since $\chi_n^2 \,(\mathrm{i})\, \chi_m^2$, from Proposition 2.24, the moment generating function of a sum of independent random variables is the product of their moment generating functions, then

$$\varphi_{\chi_n^2+\chi_m^2}(u) = \varphi_{\chi_n^2}(u)\varphi_{\chi_m^2}(u) = \frac{1}{(1-2u)^{\frac{n+m}{2}}} = \varphi_{\chi_{n+m}^2}(u) ,$$

where $\varphi_{\chi_n^2}(u) = \frac{1}{(1-2u)^{\frac{n}{2}}}$ and $\varphi_{\chi_m^2}(u) = \frac{1}{(1-2u)^{\frac{m}{2}}}$ are the moment generating functions of χ_n^2 and χ_m^2, respectively.

□

From the expression of $\varphi_{\chi_n^2}(u)$ we get the first and the second derivatives given by

$$\begin{cases} \varphi'(u \mid n) = \dfrac{n}{(1-2u)^{\frac{n+2}{2}}} \\[2ex] \varphi''(u \mid n) = \dfrac{n(n+2)}{(1-2u)^{\frac{n+4}{2}}} \end{cases} ,$$

so the first two moments in order to the origin of χ_n^2 will be

$$\begin{cases} E\left(\chi_n^2\right) = n \\ E\left(\left(\chi_n^2\right)^2\right) = n(n+2) \end{cases} . \quad (2.3.13)$$

Thus the variance will be

$$V\left(\chi_n^2\right) = 2n \ . \quad (2.3.14)$$

Besides central we are also interested in non-central chi-squares.

Definition 2.30. Let $X_1,\ldots,X_n$ be n independent random variables, $X_1(i)\ldots(i)X_n$, with distributions

$$\begin{matrix} X_1 \sim \mathcal{N}\left(\cdot \mid \mu_1, 1\right) \\ \vdots \\ X_n \sim \mathcal{N}\left(\cdot \mid \mu_n, 1\right) \end{matrix} ,$$

then the distribution of

$$\chi_{n,\delta}^2 = \sum_{i=1}^{n} X_i^2$$

will be a chi-square with n degrees of freedom and non-centrality parameter $\delta = \sum_{i=1}^{n} \mu_i^2$.

The moment generating function of $\chi_{n,\delta}^2$ will be, see Mexia (1995, pg.48), given by

$$\varphi_{\chi_{n,\delta}^2}(u) = \varphi\left(u \mid n, \delta\right) = \frac{e^{\frac{\delta u}{1-2u}}}{(1-2u)^{\frac{n}{2}}} , \quad (2.3.15)$$

with $\delta = \sum_{i=1}^{n} \mu_i^2$. Since

$$\frac{\delta u}{1-2u} = -\frac{\delta}{2} + \frac{\delta}{2(1-2u)} ,$$

we have

$$\varphi\left(u \mid n, \delta\right) = e^{-\frac{\delta}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^j}{j\,!} \frac{1}{(1-2u)^{\frac{n+2j}{2}}}$$

$$= e^{-\frac{\delta}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^j}{j!} \varphi(u \mid n+2j) \qquad (2.3.16)$$

so it may be shown, see Mexia (1995, pg.49), that

$$G(x \mid n, \delta) = e^{-\frac{\delta}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^j}{j!} G(x \mid n+2j) , \qquad (2.3.17)$$

where $G(x \mid n, \delta)$ $[G(x \mid n+2j), j = 0, \ldots]$ is the distribution function of $\chi^2_{n,\delta}$ $[\chi^2_{n+2j}, j = 0, \ldots]$.

We can assume there is an indicator variable N, with Poisson distribution with parameter $\frac{\delta}{2}$, and that, when $N = j$, $\chi^2_{n,\delta}$ is distributed as a central chi-square with $n+2j$ degrees of freedom, χ^2_{n+2j}.

If $\chi^2_{n,\delta}$ (i) $\chi^2_{m,\delta'}$ it is easy to see that

$$\varphi_{\chi^2_{n,\delta}+\chi^2_{m,\delta'}}(u) = \frac{e^{\frac{(\delta+\delta')u}{1-2u}}}{(1-2u)^{\frac{n+m}{2}}} , \qquad (2.3.18)$$

so when we add independent chi-squares we add both the degrees of freedom and the non-centrality parameters. Thus there is additivity both for degrees of freedom and the non-centrality parameters.

Definition 2.31. Let χ^2_r and χ^2_s be independent central chi-squared random variables with r and s degrees of freedom, respectively. The quotient

$$\mathcal{F} = \frac{\chi^2_r / r}{\chi^2_s / s}$$

has a central F distribution with r and s degrees of freedom, we put $F(. \mid r,s)$.

Now , see Mood et al (1987, pg 246), the density of $\mathcal{F}$ is given by

$$f(z \mid r,s) = \begin{cases} 0 & ; \quad z \le 0 \\ \dfrac{\Gamma\left(\frac{r+s}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\Gamma\left(\frac{s}{2}\right)} \dfrac{\frac{r}{s}\left(\frac{r}{s}z\right)^{\frac{r}{2}-1}}{\left(1+\frac{r}{s}z\right)^{\frac{r+s}{2}}} & ; \quad z > 0 \end{cases} \qquad (2.3.19)$$

Definition 2.32. Let $\chi^2_{r,\delta}$ be a non-central chi-squared random variable with r degrees of freedom and non-centrality parameter δ and χ^2_s a central chi-squared random variable with s degrees of freedom. If $\chi^2_{r,\delta}$ (i) χ^2_s, the quotient

$$\frac{\chi^2_{r,\delta}/r}{\chi^2_s/s}$$

has an F distribution with r and s degrees of freedom and non-centrality parameters δ and zero, $F(\cdot \mid r, s, \delta, 0)$.

Instead of the F distribution we will use mainly the, more tractable, $\bar{F}$ distribution.

Definition 2.33. Let χ^2_r and χ^2_s be independent central chi-squared random variables with r and s degrees of freedom, respectively. The quotient

$$\mathcal{T} = \frac{\chi^2_r}{\chi^2_s}$$

has an $\bar{F}$ distribution, with r and s degrees of freedom, $\bar{F}(\cdot \mid r, s)$.

Now with χ^2_r (i) χ^2_s, since $\mathcal{F} = \frac{s}{r}\mathcal{T}$, the density of $\mathcal{T} = \frac{\chi^2_r}{\chi^2_s}$ is given by

$$\bar{f}(z \mid r,s) = \frac{s}{r} f\left(\frac{s}{r} z \mid r,s\right) = \begin{cases} 0 & ; \quad z \le 0 \\ \dfrac{\Gamma\left(\frac{r+s}{2}\right)}{\Gamma\left(\frac{r}{2}\right)\Gamma\left(\frac{s}{2}\right)} \dfrac{z^{\frac{r}{2}-1}}{(1+z)^{\frac{r+s}{2}}} & ; \quad z > 0 \end{cases} . \tag{2.3.20}$$

Let $\chi^2_{r,\delta}$ be a chi-squared random variable with r degrees of freedom and non-centrality parameter δ and let χ^2_s be a central chi-squared random variable with s degrees of freedom, using indicator variables it may be shown that, see Robbins (1948) and Robbins & Pitman (1949), the density of the quotient $\frac{\chi^2_{r,\delta}}{\chi^2_s}$ will be

$$\bar{f}(z \mid r,s,\delta) = e^{-\delta/2} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^j}{j!} \bar{f}(z \mid r+2j, s). \tag{2.3.21}$$

Likewise, see Mexia (1995) and Nunes (2005), with $\bar{F}(z \mid r, s, \delta)$ and $\bar{F}(z \mid r+2j, s)$, $j = 0, \ldots$, the corresponding distribution functions we have

$$\bar{F}(z \mid r, s, \delta) = e^{-\delta/2} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^j}{j!} \bar{F}(z \mid r+2j, s), \tag{2.3.22}$$

so

$$\begin{aligned} \frac{\partial \bar{F}(z \mid r, s, \delta)}{\partial \delta} &= -\frac{1}{2} e^{-\delta/2} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^j}{j!} \bar{F}(z \mid r+2j, s) + \frac{1}{2} e^{-\delta/2} \sum_{j=1}^{\infty} \frac{\left(\frac{\delta}{2}\right)^{j-1}}{(j-1)!} \bar{F}(z \mid r+2j, s) \\ &= -\frac{1}{2} \bar{F}(z \mid r, s, \delta) + \frac{e^{-\delta/2}}{2} \sum_{j'=0}^{\infty} \frac{\left(\frac{\delta}{2}\right)^{j'}}{j'!} \bar{F}(z \mid r+2+2j', s) \\ &= \frac{\bar{F}(z \mid r+2, s, \delta) - \bar{F}(z \mid r, s, \delta)}{2} \quad . \end{aligned} \tag{2.3.23}$$

Consider the independent chi-squares random variables $\chi^2_{r,\delta}$, χ^2_2 and χ^2_s. Since $\chi^2_{r,\delta} + \chi^2_2$ is a chi-square with $r+2$ degrees of freedom and non-centrality parameter δ, we have

$$\begin{cases} \dfrac{\chi^2_{r,\delta}}{\chi^2_s} \sim \bar{F}(z \mid r, s, \delta) \\[2ex] \dfrac{\chi^2_{r,\delta} + \chi^2_2}{\chi^2_s} \sim \bar{F}(z \mid r+2, s, \delta) \end{cases} \quad . \tag{2.3.24}$$

Now

$$\mathrm{pr}\left(\frac{\chi^2_{r,\delta}}{\chi^2_s} < \frac{\chi^2_{r,\delta} + \chi^2_2}{\chi^2_s} \right) = 1,$$

so we will have, for $z > 0$,

$$\bar{F}(z \mid r+2, s, \delta) < \bar{F}(z \mid r, s, \delta) \tag{2.3.25}$$

and so

$$\frac{\partial \bar{F}(z \mid r, s, \delta)}{\partial \delta} < 0 \quad . \tag{2.3.26}$$

Let now $f_{1-q,r,s}$ be the $(1-q)$-th quantile for the F distribution with r and s degrees of freedom and non-centrality parameter δ, $F(z \mid r, s, \delta)$. If $\mathcal{F} \sim F(z \mid r, s, \delta)$ we will have

$$\mathcal{T} = \frac{r}{s}\mathcal{F} \sim \bar{F}\left(\cdot \mid r,\, s,\, \delta\right) , \tag{2.3.27}$$

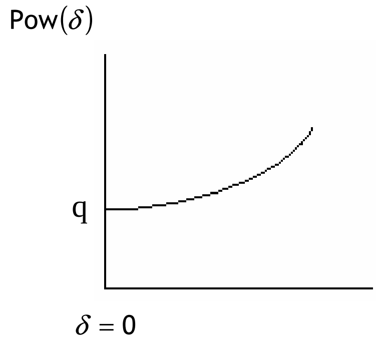
so $f_{1-q,r,s}$, the critical value for a q level test with statistic $\mathcal{F}$, will be replaced by the critical value $\frac{r}{s}f_{1-q,r,s}$ when we use statistic $\mathcal{T}$. The power of the test will depend on δ being given by

$$\text{Pow}(\delta) = 1 - \bar{F}\left(\frac{r}{s}f_{1-q,r,s} \mid r, s,\, \delta\right) \tag{2.3.28}$$

which increases with δ. Thus if we are testing an hypothesis that may be written as

$$H_0 : \delta = 0 , \tag{2.3.29}$$

we will have a strictly unbiased test, as may be seen from the following graph, since the alternatives correspond to $\delta > 0$.



2.3.4. An Application

In this application we consider a normal model, which will be studied in more detail in Chapter 4.

We write $\underline{Y} \sim \mathcal{N}\left(\underline{\mu}, V\right)$, for a normal model, $\underline{Y}$, with mean vector

$$\underline{\mu} = X_0 \underline{\beta}_0 \tag{2.3.30}$$

and variance-covariance matrix

$$V = \sum_{j=1}^{m} \gamma_j Q_j \tag{2.3.31}$$

with $Q_1, \ldots, Q_m$ pairwise orthogonal orthogonal projection matrices such that

$$\sum_{j=1}^{m} Q_j = I_n \ , \tag{2.3.32}$$

$\gamma_j > 0$, $j = 1,\ldots,m$, and

$$T = X_0 X_o^{+} = \sum_{j=1}^{z} Q_j \ \ , \tag{2.3.33}$$

with T the orthogonal projection matrix on $\Omega = R(X_0)$, the space spanned by $\underline{\mu}$.

As we will see, there may be linear restrictions on the $\gamma_1,\ldots,\gamma_m$, but in this section we will not consider them.

Let the row vectors of A_j constitute an orthogonal basis for $\nu_j = R(Q_j)$, $j = 1,\ldots,m$. Then

$$\begin{cases} A_j^T A_j = Q_j & , j = 1,\ldots m \\ A_j A_j^T = I_{g_j} & , j = 1,\ldots m \end{cases} ,$$

so

$$\begin{cases} A_j Q_j = A_j & , \ j = 1,\ldots m \\ A_j Q_{j'} = 0_{g_j \times g_{j'}} & , \ j \neq j' \end{cases} .$$

We then will have

$$\begin{cases} V^{-1} = \sum_{j=1}^{m} \gamma_j^{-1} Q_j & \\ A_j V A_j^T = \gamma_j I_{g_j} & , j = 1,\ldots,m \\ A_j V A_{j'} = 0_{g_j \times g_{j'}} & , \ j \neq j' \end{cases} , \tag{2.3.34}$$

thus taking

$$\begin{cases} \underline{\tilde{\eta}}_j = A_j \underline{Y} & , j = 1,\ldots,m \\ \underline{\eta}_j = A_j \underline{\mu} & , j = 1,\ldots,m \end{cases} , \tag{2.3.35}$$

we will have

$$\underline{\tilde{\eta}}_j \sim \mathcal{N}\left(\underline{\eta}_j, \gamma_j I_{g_j}\right), \ j = 1,\ldots,m,$$

with $\underline{\tilde{\eta}}_1 \, (\mathrm{i}) \ldots (\mathrm{i}) \, \underline{\tilde{\eta}}_m$, since they have joint normal distribution being linear transform of $\underline{Y}$ and null cross-covariance matrices.

Moreover

$$S_j = \| \underline{\tilde{\eta}}_j \|^2 \ , \ j=1,\ldots,m, \tag{2.3.36}$$

will be the product by γ_j of a chi-square with g_j degrees of freedom , $j=1,\ldots,m$, and non-centrality parameter

$$\delta_j = \frac{1}{\gamma_j} \| \underline{\eta}_j \|^2 , \ j=1,\ldots,m,$$

since the components of $\frac{1}{\sqrt{\gamma_j}} \underline{\tilde{\eta}}_j$ are normal, with mean values that are the components of $\frac{1}{\sqrt{\gamma_j}} \underline{\eta}_j$, variances equal to 1 and null covariances. We put $S_j \sim \gamma_j \chi^2_{g_j,\delta_j}$, j=1,…,m.

It is easy to see that

$$\underline{\mu} = T\,\underline{\mu} = \sum_{j=1}^{z} A_j^T \underline{\eta}_j \tag{2.3.37}$$

and that

$$\underline{\eta}_j = 0 \ , \ j = z+1,\ldots,m ,$$

so that

$$\delta_j = 0 \ , \ j = z+1,\ldots,m \ .$$

Moreover

$$\begin{cases} A_j V^{-1} A_j^T = \dfrac{1}{\gamma_j} I_{g_j} & , j=1,\ldots,m \\ A_j V^{-1} A_{j'}^T = 0_{g_j \times g_{j'}} & , j \neq j' \end{cases} \tag{2.3.38}$$

so that

$$\begin{aligned} (\underline{y}-\underline{\mu})^T V^{-1} (\underline{y}-\underline{\mu}) &= \sum_{j=1}^{m} \frac{1}{\gamma_j} (\underline{y}-\underline{\mu})^T Q_j (\underline{y}-\underline{\mu}) \\ &= \sum_{j=1}^{m} \frac{1}{\gamma_j} (\underline{y}-\underline{\mu})^T A_j^T A_j \ (\underline{y}-\underline{\mu}) \\ &= \sum_{j=1}^{m} \frac{1}{\gamma_j} \| \underline{\tilde{\eta}}_j - \underline{\eta}_j \|^2 \\ &= \sum_{j=1}^{z} \frac{1}{\gamma_j} \| \underline{\tilde{\eta}}_j - \underline{\eta}_j \|^2 + \sum_{j=z+1}^{m} \frac{1}{\gamma_j} \| \underline{\tilde{\eta}}_j \|^2 \end{aligned}$$

$$= \sum_{j=1}^{z} \frac{\left\| \underline{\tilde{\eta}}_j - \underline{\eta}_j \right\|^2}{\gamma_j} + \sum_{j=z+1}^{m} \frac{S_j}{\gamma_j} . \tag{2.3.39}$$

Now

$$\det(V) = \prod_{j=1}^{m} \gamma_j^{g_j} , \tag{2.3.40}$$

so , from (2.3.39) , (2.3.40) and from Definition 2.28, $\underline{y}$ will have the density

$$n(\underline{y}) = \frac{e^{-\frac{1}{2}\left(\sum_{j=1}^{z} \frac{\left\| \underline{\tilde{\eta}}_j - \underline{\eta}_j \right\|^2}{\gamma_j} + \sum_{j=z+1}^{m} \frac{S_j}{\gamma_j} \right)}}{(2\pi)^{n/2} \prod_{j=1}^{m} \gamma_j^{g_j/2}} . \tag{2.3.41}$$

Since, with $\underline{\tilde{\eta}}_{j,1},\ldots,\underline{\tilde{\eta}}_{j,g_j}$ $\left[\underline{\eta}_{j,1},\ldots,\underline{\eta}_{j,g_j}\right]$ the components of $\underline{\tilde{\eta}}_j$ $\left[\underline{\eta}_j\right]$, $j = 1,\ldots,z$, we have

$$\left\| \underline{\tilde{\eta}}_j - \underline{\eta}_j \right\|^2 = S_j - 2\sum_{l=1}^{g_j} \eta_{j,l}\, \tilde{\eta}_{j,l} + \left\| \underline{\eta}_j \right\|^2 , \tag{2.3.42}$$

so we can write the density as

$$n(\underline{y}) = \frac{e^{-\frac{1}{2}\sum_{j=1}^{z}\left\| \underline{\eta}_j \right\|^2}}{(2\pi)^{n/2} \prod_{j=1}^{m} \gamma_j^{g_j/2}} \cdot e^{-\frac{1}{2}\sum_{j=1}^{m} \frac{S_j}{\gamma_j} + \sum_{j=1}^{z}\sum_{l=1}^{g_j} \frac{\eta_{j,l}}{\gamma_j}\tilde{\eta}_{j,l}} . \tag{2.3.43}$$

Thus taking

$$\begin{array}{lll} \theta_i = \gamma_i & , & i = 1,\ldots,m \\ \theta_{m+l} = \eta_{1,l} & , & l = 1,\ldots,g_1 \\ \theta_{m+g_1+l} = \eta_{2,l} & , & l = 1,\ldots,g_2 \\ \vdots & & \end{array}$$

and

$$\begin{array}{lll} T_i = S_i & , & i = 1,\ldots,m \\ T_{m+l} = \tilde{\eta}_{1,l} & , & l = 1,\ldots,g_1 \\ T_{m+g_1+l} = \tilde{\eta}_{2,l} & , & l = 1,\ldots,g_2 \\ \vdots & & \end{array} ,$$

from (2.2.30), it is easy to see that $n(\underline{y})$ belongs to a full rank exponential family and that the statistics $T_1,\ldots,T_m,T_{m+1},\ldots,T_{m+g_1},\ldots$ constitute a minimal sufficient complete statistic so that we would have UMVUE for this model.

When we return to these models we will have to consider linear restrictions on the $\gamma_1,\ldots,\gamma_m$ and so the situation is not as convenient as this one.

It is interesting to point out that for the preceding model we have the UMVUE

$$\underline{\tilde{\Psi}} = \sum_{j=1}^{m} C_j \underline{\tilde{\eta}}_j \tag{2.3.44}$$

for estimable vectors and the estimators

$$\tilde{\gamma}_j = \frac{S_j}{g_j} \quad j = z+1,\ldots,m \quad , \tag{2.3.45}$$

for some of the canonical variance components. Later on we will consider special cases of error orthogonal models, for which we have unbiased estimators for the remaining canonical variance components and for the usual variance components but these estimators will not be UMVUE. Moreover for these models the $\underline{\tilde{\Psi}}$ will be BLUE but not UMVUE . Thus there is a trade-off between what we can estimate and the quality of our estimators. The reason why we may prefer the formulation that we will consider later on, despite this trade-off, is that it may better convey the action of the factors described by the models.

3. Models and operations

Mixed models play a central part in this thesis. In particular, we will focus on models with commutative orthogonal block structure (COBS), introduced by Fonseca et al. (2008), who constitute a very interesting class within the models with Orthogonal Block Structure (OBS).

VanLeeuwen et al (1998, 1999) introduced error-orthogonal models (EO) and showed that EO and COBS are identical classes of models.

In studying EO we will favour the COBS approach and we present an independent proof of the identity of EO and COBS.

3.1. COBS and related models

In order to introduce the study of mixed models let us consider a general linear model. Let

$$y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_k x_{ik} + \varepsilon_i \quad , \quad i = 1, \ldots, n . \qquad (3.1.1)$$

be the value of the response variable at the i-th of n levels. Since each of the k explanatory variables $x_1, x_2, \ldots, x_k$ has n levels, $x_{i,j}$ represents the i-th level of the j-th explanatory variable x_j, $j = 1, \ldots, k$. The β_i are unknown parameters and $\underline{\varepsilon}$ is the errors vector. Rewriting this model as

$$y_i = [x_{i0}, x_{i1}, \ldots, x_{ik}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \varepsilon_i , \quad i = 1, \ldots, n \ , \qquad (3.1.2)$$

and collecting these n equations, using the matrix notation, we have

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \qquad (3.1.3)$$

where

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad (3.1.4)$$

is the vector of responses (observations),

$$X = \begin{bmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1k} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \qquad (3.1.5)$$

is an $n \times (k+1)$ matrix, often called the model (or design) matrix, of the levels of the explanatory variables. Typically, although not always, $x_{i0} = 1$, for all i, and then β_0 is the intercept term in the model. The $k \times 1$ vector

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad (3.1.6)$$

is the vector of unknown parameters, β_j, $j = 1, \ldots, k$, that can be fixed or random variables, and

$$\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (3.1.7)$$

is the errors vector, which is assumed to have null mean vector and variance-covariance matrix $\sigma^2 V$, with V a known matrix and σ^2 unknown.

Traditionally, linear models are classified in three categories depending on the nature of the parameters $\beta_0, \beta_1, \ldots, \beta_k$. When these parameters are assumed to be constants they are called fixed effects and the model is called a fixed effects model. The practical use of this type of model is very common, however, there are many situations in which is advantageous to include more random effects besides the error term, namely when we are confronted with correlated data from, for example, repeated measurements.

If we treat all or some of the parameters $\beta_0, \beta_1, \ldots, \beta_k$ as random variables (random effects) we have two possible types of models. When all the effects in a model (except for the intercept) are considered random effects, then the model is called a random effects model. When some effects are fixed and others are random, the model is called a mixed effects model.

Definition.3.1. A linear model

$$\underline{Y} = \sum_{i=0}^{w} X_i \underline{\beta}_i$$

where $\underline{\beta}_0$ is fixed and $\underline{\beta}_1, \ldots, \underline{\beta}_w$ are independent random vectors with null mean vectors and variance-covariance matrices $\sigma_1^2 I_{c_1}, \ldots, \sigma_w^2 I_{c_w}$, where $c_i = \text{rank}(X_i)$, $i = 1, \ldots, w$, is said to be a mixed effects linear model, or simply, a mixed model.

These models will play a central part in our work.

The matrices $X_1, \cdots, X_w$ will be known and such that

$$R\left([X_1 \quad \cdots \quad X_w]\right) = IR^n .$$

From Definition 3.1 and from the properties of the operators $E(\cdot)$ and $V(\cdot)$, it is straightforward to calculate the mean vector and the variance-covariance matrix of $\underline{Y}$. The mean vector is given by

$$\begin{aligned}
\underline{\mu} &= E\left(\sum_{i=0}^{w} X_i \underline{\beta}_i\right) \\
&= E(X_0 \beta_0) + E\left(\sum_{i=1}^{w} X_i \underline{\beta}_i\right) \\
&= X_0 \underline{\beta}_0 + \sum_{i=1}^{w} X_i \, E\left(\underline{\beta}_i\right) \\
&= X_0 \underline{\beta}_0
\end{aligned} \tag{3.1.8}$$

and variance-covariance matrix is given by

$$\begin{aligned}
V &= V\left(\sum_{i=0}^{w} X_i \underline{\beta}_i\right) \\
&= 0 + V\left(\sum_{i=1}^{w} X_i \underline{\beta}_i\right) \\
&= \sum_{i=1}^{w} X_i \, V\left(\underline{\beta}_i\right) X_i^T \\
&= \sum_{i=1}^{w} X_i \left(\sigma_i^2 \, I_{C_i}\right) X_i^T \\
&= \sum_{i=1}^{w} \sigma_i^2 X_i X_i^T \\
&= \sum_{i=1}^{w} \sigma_i^2 M_i
\end{aligned} \tag{3.1.9}$$

where $M_i = X_i X_i^T$, $i = 1, \dots, w$.

We now establish

Proposition 3.1. *Let $\underline{\sigma}^2$ be the vector of variance-covariance components, with components $\sigma_1^2, \dots, \sigma_w^2$. When $\underline{\sigma}^2 > \underline{0}$ ($\sigma_i^2 > 0$, $i = 1, \dots, w$), V is positive definite.*

Proof. If $\underline{\sigma}^2 > \underline{0}$ we have, see Silvey (1975)

$$\begin{aligned}
R(V) &= R\left[(\sigma_1 X_1 \cdots \sigma_w X_w)(\sigma_1 X_1 \cdots \sigma_w X_w)^T\right] \\
&= R[(\sigma_1 X_1 \cdots \sigma_w X_w)] = R[(X_1 \cdots X_w)] = IR^n \quad ,
\end{aligned}$$

so V will be invertible. Now invertible variance-covariance matrices are definite positive, see Mexia (1995), and the thesis is established.

□

We point out that:

- we may take $\sigma_w^2 = \sigma^2$, $X_w = I_n$ and $\underline{\beta}_w = \underline{\varepsilon}$ with $\underline{\varepsilon}$ the error vector;

- we can consider both fixed effects models and random effects models as particular cases of the mixed effects model. When the model is of random effects, $X_0 = 1_n$ and $\underline{\beta}_0 = \mu$ and when the model is of fixed effects, $w = 1$ and $X_w \underline{\beta}_w = \underline{\varepsilon}$.

The space, Ω, spanned by $\underline{\mu}$ will be $R(X_0)$ so, according to (2.1.14), the orthogonal projection matrix on Ω will be

$$T = X_0 \left(X_0^T X_0\right)^+ X_0^T = X_0 X_0^+ \quad . \qquad (3.1.10)$$

Definition 3.2. A linear model has orthogonal block structure when its variance-covariance matrix, V, can be written as the linear combination

$$V = \sum_{j=1}^{m^0} \gamma_j^0 Q_j^0 ,$$

with $Q_1^0, \ldots, Q_{m^0}^0$ known POOPM such that

$$\sum_{j=1}^{m^0} Q_j^0 = I_n .$$

To lighten the writing we name the linear models with orthogonal block structure, simply, as OBS. These models, introduced by J. A. Nelder , see Nelder (1965a, 1965b), have been intensively studied, see for instance Houtman & Speed (1983) and Mejza (1992) and continue to play a central part in the theory of randomized block designs, see Calinski & Kageyama (2000, 2003).

We now establish

Proposition 3.2. *When the matrices* $M_1, \ldots, M_w$ *commute the model is OBS.*

Proof. When the matrices $M_1,\dots,M_w$ commute they generate a CJAS, $\mathcal{A}^0$, with principal basis $\underline{Q}^0 = \{Q_1^0,\dots,Q_{m^0}^0\}$ thus

$$M_i = \sum_{j=1}^{m^0} b_{i,j}^0 Q_j^0\ ,\ i = 1,\dots,w\ ,$$

and

$$V = \sum_{j=1}^{m^0} \gamma_j^0 Q_j^0\ \ ,$$

with $\gamma_j^0 = \sum_{i=1}^{w} b_{i,j}^0 \sigma_i^2$, $j = 1,\dots,m^0$. The thesis is established since $\sum_{i=1}^{w} M_i \in \mathcal{A}$ is invertible so $\mathcal{A}^0$ is complete and $\sum_{j=1}^{m^0} Q_j^0 = I_n$.

□

The special class of OBS we present next will have high relevance in our work.

Definition 3.3. A mixed model has commutative orthogonal block structure if it is OBS and moreover, with T the orthogonal projection matrix on the space spanned by the mean vector,

$$T Q_j^0 = Q_j^0 T\ ,\quad j = 1,\dots,m^0\ .$$

Models with commutative orthogonal block structure were introduced on Fonseca et al (2008) and have also been studied on Santos et al (2007a, 2007b), Nunes et al (2008) and Carvalho et al (2008). Similarly to what was previously done, to lighten the writing we name these models as COBS.

Lets establish the

Proposition 3.3. *If the matrices* $M_1,\dots,M_w$ *and* T *commute the model is COBS.*

Proof. Both T and $Q_1^0,\dots,Q_{m^0}^0$ will belong to the CJAS , $\mathcal{A}^0$, generated by the $M_1,\dots,M_w$ and T so they will commute and the model is COBS.

□

Let $\underline{Q} = \{Q_1,\dots,Q_m\}$ be the principal basis of $\mathcal{A}$, and put $T^C = I_n - T$, we now have

Proposition 3.4. $\underline{Q}$ *is constituted by non null matrices* $Q_j^0 T$ *and* $Q_j^0 T^C$, $j = 1,\dots,m^0$.

Proof. Any CJAS A' that contains $\underline{M} = \{M_1,\dots,M_w\}$ and T will contain A^0 thus containing $\underline{Q}^0$. Since A^0 is complete containing I_n, A' will be complete, so besides T and I_n it will contain T^C and the non-null matrices $Q_j^0 T$ and $Q_j^0 T^C$, $j = 1,\dots,m^0$. To complete the proof we have only to point out that these non-null matrices are POOPM then constituting the principal basis of a complete CJAS since

$$\sum_{j=1}^{m^0}\left(Q_j^0 T + Q_j^0 T^C\right) = \sum_{j=1}^{m^0} Q_j^0 = I_n \ .$$

In this expression we could include any null matrices since they did not alter the sum.

□

We now will have the matrices $M_1,\dots,M_w$ represented by

$$M_i = \sum_{j=1}^{m} b_{i,j} Q_j \ , \ i = 1,\dots,w \ , \tag{3.1.11}$$

with $B = \lfloor b_{i,j} \rfloor$ the transition matrix $\underline{M} \setminus \underline{Q}$. Then

$$V = \sum_{i=1}^{w} \sigma_i^2 M_i = \sum_{j=1}^{m} \gamma_j Q_j \ , \tag{3.1.12}$$

with the canonical variance components

$$\gamma_j = \sum_{i=1}^{w} b_{i,j} \sigma_i^2 \ , \ j = 1,\dots,m . \tag{3.1.13}$$

Let $z^0 \geq 0$ be the number of matrices of $\underline{Q}^0$ such that $Q_j^0 T = Q_j^0$ and z be the number of matrices of $\underline{Q}^0$ such that $Q_j^0 T \neq 0_{n\times n}$; $Q_j^0 T \neq Q_j^0$. We can always order the matrices in $\underline{Q}^0$ and $\underline{Q}$ to have

$$\begin{cases} Q_j = Q_j^0 & , \quad j = 1,\dots,z^0 \ \left(\text{if } z^0 > 0\right) \\ Q_j = Q_j^0 T & , \quad j = z^0+1,\dots,z \\ Q_j = Q_{j-z}^0 T^C & , \quad j = z+1,\dots,2z-z^0 \\ Q_j = Q_{j-z}^0 T^C & , \quad j = 2z-z^0+1,\dots,m \end{cases} .$$

Then we will have

$$T = \sum_{j=1}^{z} Q_j \tag{3.1.14}$$

and

$$m^0 = m - z \tag{3.1.15}$$

as well as

$$\begin{cases} Q_j^0 = Q_j & , \quad j = 1,\ldots,z^0 \left(\text{if } z^0 > 0\right) \\ Q_j^0 = Q_j + Q_{j+z} & , \quad j = z^0+1,\ldots,z \\ Q_j^0 = Q_{j+z} & , \quad j = z+1,\ldots,m^0 \end{cases} . \tag{3.1.16}$$

Since

$$V = \sum_{j=1}^{m^0} \gamma_j^0 Q_j^0 = \sum_{j=1}^{m} \gamma_j Q_j , \tag{3.1.17}$$

from (3.1.16) we get

$$\begin{cases} \gamma_j^0 = \gamma_j & , \quad j = 1,\ldots,z^0 \quad \left(\text{if } z^0 > 0\right) \\ \gamma_j^0 = \gamma_j + \gamma_{j+z} & , \quad j = z^0+1,\ldots,z \\ \gamma_j^0 = \gamma_{j+z} & , \quad j = z+1,\ldots,m^0 \end{cases} \tag{3.1.18}$$

Likewise, from

$$M_i = \sum_{j=1}^{m^0} b_{i,j}^0 Q_j^0 = \sum_{j=1}^{m} b_{i,j} Q_j \tag{3.1.19}$$

we get

$$\begin{cases} b_{i,j}^0 = b_{i,j} & , \quad j = 1,\ldots,z^0 \left(\text{if } z^0 > 0\right), \ i = 1,\ldots,w \\ b_{i,j}^0 = b_{i,j} + b_{i,j+z} & , \quad j = z^0+1,\ldots,z , \ i = 1,\ldots,w \\ b_{i,j}^0 = b_{i,j+z} & , \quad j = z+1,\ldots,m^0 , \ i = 1,\ldots,w \end{cases} . \tag{3.1.20}$$

Thus $B^0 = \left\lfloor b_{i,j}^0 \right\rfloor$ is sub-matrix of $B = \left\lfloor b_{i,j} \right\rfloor$, since every column of B^0 is column of B. Moreover the column of B with indexes j and $j+z$, $j = z^0+1 , \ldots, z$, will be identical, and every column of B is equal to a column of B^0 so

$$R\left(B^0\right) = R\left(B\right) \tag{3.1.21}$$

and

$$\text{rank}\left(B^0\right) = \text{rank}\left(B\right) . \tag{3.1.22}$$

If matrices $M_0, M_1, \ldots, M_w$ commute they will generate a CJAS, $\overline{\mathcal{A}}$, with principal basis $\underline{\overline{Q}} = \{\overline{Q}_1, \ldots, \overline{Q}_{\overline{m}}\}$ and will have the transition matrix $\overline{B} = [\overline{b}_{i,j}]$ with

$$M_i = \sum_{j=1}^{\overline{m}} \overline{b}_{i+1,j} \overline{Q}_j \quad , \quad i = 1, \ldots, w \quad . \tag{3.1.23}$$

Since T is the orthogonal projection matrix on the range space of M_0 we will have $T \in \overline{\mathcal{A}}$ and, if necessary, we can reorder the matrices in $\underline{\overline{Q}}$ to have

$$T = \sum_{j=1}^{\overline{z}} \overline{Q}_j \quad , \tag{3.1.24}$$

then T and the $M_1, \ldots, M_w$ will commute and so the model will be COBS. Moreover, also since T is the orthogonal projection matrix on $R(M_0)$ we will have

$$M_0 = \sum_{j=1}^{\overline{z}} \overline{b}_{1,j} \overline{Q}_j \ . \tag{3.1.25}$$

Definition 3.4. A mixed model in which matrices $M_0, M_1, \ldots, M_w$ commute is said to have Completely Commutative Orthogonal Block Structure. We name this model as Complete COBS, or simply, CCOBS.

3.2. Error-orthogonal models

In this section our study will focus on error-orthogonal models and we shall show that COBS are identical to the error-orthogonal models.

The notion of an ''error-orthogonal design'' were introduced by VanLeeuwen, Seely and Birkes , see Vanleeuwen et al (1998), and defined as follows.

Definition 3.5. A linear model has an error-orthogonal design if the least-squares estimator of the mean vector is a uniformly best linear unbiased estimator, UBLUE, and the covariance matrix of the vector of least-squares residuals has orthogonal block structure.

To lighten the writing, from now on, we name the error-orthogonal models, simply, as EO.

Definition 3.6. The LSE for estimable vectors in EO are uniformly BLUE, UBLUE, if they are BLUE whatever the variance components.

From our previous results it follows that COBS are EO. To show that EO are COBS, thus identifying both classes of models, we have to show that in an OBS where LSE, for estimable vectors, are UBLUE, the matrices T and $Q_1,\dots,Q_w$ commute. Now

$$\underline{\Psi} = G\underline{\beta}_0 \tag{3.2.1}$$

is estimable if and only if

$$G = UX_0, \tag{3.2.2}$$

see for instance Mexia (1990), thus if

$$\underline{\Psi} = U\underline{\mu} \ . \tag{3.2.3}$$

Let us establish

Proposition 3.5. *An OBS with mean vector* $\underline{\mu} = X_0\underline{\beta}_0$ *is EO if and only if the* $M\underline{Y}$*, with* $MT = M$*, are UBLUE.*

Proof. Comes from Definition 3.6 and from Proposition 2.17.

□

Now, if $\underline{\mu} = X_0\underline{\beta}_0$, we have

$$E(M\underline{Y}) = E(MT\underline{Y}) \tag{3.2.4}$$

and so, when the model is OBS, with variance-covariance matrix

$$V(\underline{\gamma}^0) = \sum_{j=1}^{m^0}\gamma_j^0 Q_j^0 , \tag{3.2.5}$$

the model is EO if and only if, whatever M and $\underline{\gamma}^0$, we have

$$MT\left(\sum_{j=1}^{m^0}\gamma_j^0 Q_j^0\right)TM^T \le M\left(\sum_{j=1}^{m^0}\gamma_j^0 Q_j^0\right)M^T, \tag{3.2.6}$$

which mean that $M\left(\sum_{j=1}^{m^0}\gamma_j^0 Q_j^0\right)M^T - MT\left(\sum_{j=1}^{m^0}\gamma_j^0 Q_j^0\right)TM^T$ is positive semi-definite.

We now get the

Lemma 3.1. *An OBS with mean vector* $\underline{\mu} = X_0 \underline{\beta}_0$ *and variance-covariance matrix* $V(\underline{\gamma}^0) = \sum_{j=1}^{m^0} \gamma_j^0 Q_j^0$ *is EO if and only if* $W_j = T Q_j^0 T \leq Q_j^0$, $j = 1, \ldots, m^0$.

Proof. To establish necessity we have only to take $M = I_n$ and $\underline{\gamma}^0$ with only one non-null component. Going over to sufficiency, whatever $\underline{v}$, when $W_j \leq Q_j^0$, $j = 1, \ldots, m^0$, with $V(M\underline{Y})$ and $V(M T \underline{Y})$ the variance-covariance matrices of $M\underline{Y}$ and $M T \underline{Y}$ we have

$$\underline{v}^T V(M\underline{Y})\underline{v} - \underline{v}^T V(M T \underline{Y})\underline{v} = \underline{v}^T \left[M\left(\sum_{j=1}^{m^0} \gamma_j^0 Q_j^0\right) M^T - M\left(\sum_{j=1}^{m^0} \gamma_j^0 W_j\right) M^T \right] \underline{v}$$

$$= (M^T \underline{v})^T \left(\sum_{j=1}^{m^0} \gamma_j^0 (Q_j^0 - W_j)\right)(M^T \underline{v}) = \sum_{j=1}^{m^0} \gamma_j^0 (M^T \underline{v})^T (Q_j^0 - W_j)(M^T \underline{v}) \geq 0$$

□

Lemma 3.2. *An OBS with mean vector* $\underline{\mu} = X_0 \underline{\beta}_0$ *and variance-covariance matrix* $V(\underline{\gamma}^0) = \sum_{j=1}^{m^0} \gamma_j^0 Q_j^0$ *is EO if and only if* $R(W_j) \leq R(Q_j)$, $j = 1, \ldots, m^0$.

Proof. According to Lemma 3.1 an OBS is EO if and only if $W_j \leq Q_j^0$, this is, whatever $\underline{v}$, $0 \leq \underline{v}^T W_j \underline{v} \leq \underline{v}^T Q_j^0 \underline{v}$, thus if $N(Q_j^0) \subseteq N(W_j)$, since the vectors that render null the quadratic form of a positive semi-definite matrix K belong to $N(K)$, $j = 1, \ldots, m^0$. To complete the proof we have only to point out that $N(Q_j^0) = R(Q_j^0)^{\perp}$ and $N(W_j) = R(W_j)^{\perp}$, this is, $N(Q_j^0)$ and $N(W_j)$ are the orthogonal complement of $R(Q_j^0)$ and $R(W_j)$, since these matrices are symmetric, $j = 1, \ldots, m^0$.

□

Lemma 3.3. *In an OBS, with mean vector* $\underline{\mu} = X_0 \underline{\beta}_0$ *and variance-covariance matrix* $V(\underline{\gamma}^0) = \sum_{j=1}^{m^0} \gamma_j^0 Q_j^0$, $W_j Q_j^0 = Q_j^0 W_j$, *this is, matrices* W_j *and* Q_j^0 *commute,* $j = 1, \ldots, m^0$.

Proof. Since the matrices are symmetric and $R(W_j) \subseteq R(Q_j^0)$, we have

$$Q_j^0 W_j = W_j = W_j^T = \left(Q_j^0 W_j\right)^T = W_j^T Q_j^{0\,T} = W_j Q_j^0,\ j = 1,\ldots,m^0.$$

□

Let

$$\underline{Q}_j = \{Q_{j,1},\ldots,Q_{j,m_j}\},\ j = 1,\ldots,m^0 \qquad (3.2.7)$$

be the principal basis of CJAS $\mathcal{A}_j$ generated by W_j and Q_j^0, $j = 1,\ldots,m^0$. Then

$$W_j = \sum_{l=1}^{m_j} b_{j,l} Q_{j,l},\ j = 1,\ldots,m^0, \text{ with } b_{j,l} \geq 0,\ l = 1,\ldots,m_j,\ j = 1,\ldots,m^0. \qquad (3.2.8)$$

Moreover we have

Lemma 3.4. $Q_j^0 = \sum_{l=1}^{m_j} Q_{j,l}$, $j = 1,\ldots,m^0$

Proof. Since Q_j^0 is an OPM belonging to $\mathcal{A}_j$ it will be the sum of matrices in $\underline{Q}_j$, $j = 1,\ldots,m^0$. We can reorder the matrices of $\underline{Q}_j$ to have $Q_j^0 = \sum_{l=1}^{z_j} Q_{j,l}$, $j = 1,\ldots,m^0$ and, since $R(W_j) \subseteq R(Q_j^0)$, we also will have $W_j = \sum_{l=1}^{z_j} b_{j,l} Q_{j,l}$, $j = 1,\ldots,m^0$, thus $\{Q_{j,1},\ldots,Q_{j,z_j}\}$ is the principal basis of a CJAS containing Q_j^0 and W_j, $j = 1,\ldots,m^0$.
Since $\{Q_{j,1},\ldots,Q_{j,z_j}\}$ is also the principal basis of the CJAS generated by Q_j^0 and W_j, we must have $z_j = m_j$, since otherwise we would have a , "smaller" than A_j , CJAS containing Q_j^0 and W_j, $j = 1,\ldots,m^0$, which is impossible.

□

With $\omega_j = R(W_j)$ and $\nu_j = R(Q_j^0)$, $j = 1,\ldots,m^0$, and $\Omega = R(T)$, if the model is EO, we have

$$\omega_j = \nu_j \cap \Omega,\quad j = 1,\ldots,m^0,$$

since $\omega_j \subseteq \nu_j$ and $\omega_j \subseteq \Omega$, $j=1,\dots,m^0$. Moreover the $Q_{j,l}$, $l=1,\dots,m_j$, $j=1,\dots,m^0$ will be POOPM, with $I_n = \sum_{j=1}^{m^0}\sum_{l=1}^{m_j} Q_{j,l}$, so they constitute $pb(\mathcal{A}^{00})$ with $\mathcal{A}^{00}$ a complete CJAS that we say corresponds to the EO. We now establish

Theorem 3.1. *(Identity of classes) COBS and EO constitute the same class of models.*

Proof. As we saw, COBS are EO. Inversely, given an EO with the corresponding CJAS $\mathcal{A}^{00}$, we will have $Q_1^0,\dots,Q_{m^o}^0 \in \mathcal{A}^{00}$ as well as

$$T = TT = T I_n T = T\left(\sum_{j=1}^{m^0} Q_j^0\right)T = \sum_{j=1}^{m^0} W_j = \sum_{j=1}^{m^0}\sum_{l=1}^{m_j} b_{j,l} Q_{j,l} \in \mathcal{A}^{00},$$

so T and the $Q_1^0,\dots,Q_{m^o}^0$ will commute and the model is COBS.

□

Stated the identity between EO and COBS, from now on we name the models as EO for priority sake. If the models were CCOBS, now they will be designated as Complete Error-orthogonal models, CEO. This does not deny the fact that we use a COBS approach in what follows.

As a parting remark we point out that in EO we may consider the decomposition $\underline{Y} = \underline{Y}_\Omega + \underline{Y}_{\Omega^\perp}$ of the observations vector in its orthogonal projection on $\Omega = R(X_0)$ and its orthogonal complement $\Omega^\perp$. The LSE of estimable vectors will depend only on $\underline{Y}_\Omega$ and the variance components will depend only on $\underline{Y}_{\Omega^\perp}$, since

$$\begin{cases} \underline{\tilde{\beta}}_0 = (X^TX)^+ X^T \underline{Y} = (X^TX)^+ X^T T \underline{Y} = (X^TX)^+ \underline{Y}_\Omega \\ S_j = \left\| Q_j \underline{Y} \right\|^2 = \left\| Q_j T^C \underline{Y} \right\|^2 = \left\| Q_j \underline{Y}_{\Omega^\perp} \right\|^2, j = z+1,\dots,m \end{cases}, \quad (3.2.9)$$

where $T^C = I_n - T$, and the only variance components we can directly estimate are the $\gamma_{z+1},\dots,\gamma_m$ for which we have the unbiased estimator

$$\tilde{\gamma}_j = \frac{S_j}{g_j}, \quad j = z+1,\dots,m \quad (3.2.10)$$

with $g_j = \text{rank}(Q_j)$, $j = 1,\dots,m$.

3.3. Segregation and matching

We now consider two structures of interest for the estimation of variance components, either usual or canonical.

Let us put

$$\underline{\gamma}_1 = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_z \end{bmatrix} \quad ; \quad \underline{\gamma}_2 = \begin{bmatrix} \gamma_{z+1} \\ \vdots \\ \gamma_m \end{bmatrix} \quad ; \quad \underline{\sigma}^2 = \begin{bmatrix} \sigma_1^2 \\ \vdots \\ \sigma_w^2 \end{bmatrix} \quad ,$$

then, with

$$B = \left[B_1 \;\; B_2 \right]$$

where B_1 has z columns, we have

$$\underline{\gamma}_l = B_l^T \underline{\sigma}^2 \,, \; l = 1,2 \quad . \tag{3.3.1}$$

When B_2 is horizontally free, HF, having linearly independent row vectors, the column vectors of B_2^T will be linearly independents and we obtain, see Mexia (1995)

$$\underline{\sigma}^2 = \left(B_2^T\right)^+ \underline{\gamma}_2 \tag{3.3.2}$$

and

$$\underline{\gamma}_1 = B_1^T \left(B_2^T\right)^+ \underline{\gamma}_2 \tag{3.3.3}$$

so that we may estimate $\underline{\sigma}^2$ and $\underline{\gamma}_1$ through $\underline{\gamma}_2$. Then the relevant parameters for the random effects part of the model, $\underline{\gamma}_2$ and $\underline{\sigma}^2$, determine each other. Thus this part of the model segregates as a sub-model and we say that there is segregation, see Ferreira (2006).

The EO with segregation will be indicated as SEO.

Definition 3.7. The EO

$$\underline{Y} = \sum_{i=1}^{w} X_i \underline{\beta}_i$$

is an expanding EO, EEO, if $R(X_i) \subset R(X_{i+1})$, $i = 1, \dots, w-1$.

Let

$$\underline{Y} = \sum_{i=1}^{w} X_i \underline{\beta}_i \tag{3.3.4}$$

be a SEO then we must have

$$R(X_w) = R([\ X_1 \ldots X_w\ ]) = IR^n$$

and with $\underline{Q} = \{Q_1, \ldots, Q_m\}$, the principal basis of the CJAS corresponding to this EO, we will have

$$M_i = \sum_{j \in C_i} b_{i,j} Q_j \ , \ i = 1, \ldots, w \ , \tag{3.3.5}$$

where $C_i = \{ j : b_{i,j} \neq 0 \}$, with $C_i \subset C_{i+1}$, $i = 1, \ldots, w-1$. We also will have $\{1, \cdots, z\} \subset C_i$, $i = 1, \ldots, w$ and $C_w = \{1, \cdots, m\}$.

We now establish

Proposition 3.6. *EEO are SEO*

Proof. Since $\{1, \cdots, z\} \subset C_i \subset C_{i+1}$, $i = 1, \ldots, w-1$, the set $C_i \setminus \{1, \cdots, z\}$ of indexes of non-null elements of the i-th row of B_2 will be strictly contained in the corresponding line for the next row so B_2 will be HF.

□

When B_2 is HF we have

$$\underline{\sigma}^2 = L \underline{\gamma}_2 \ , \tag{3.3.6}$$

where L is any left inverse of B_2^T, and so we also have

$$\underline{\gamma}_1 = B_1^T L \underline{\gamma}_2 \ , \tag{3.3.7}$$

thus there are different possible estimators for $\underline{\sigma}^2$ and $\underline{\gamma}_1$ when the model is SEO. We now are going to show why we choose $L = \left(B_2^T\right)^+$. For this, we start by establishing

Lemma 3.5. *If* W *is* $k \times k$ *positive semi-definite we have* $W = W^{1/2} W^{1/2}$ *with* $W^{1/2}$ *positive semi-definite and, whatever the matrix* C *for which* CWC^T *is defined*

$$\left\| CWC^T \right\| \leq \left\| W^{1/2} \right\|^2 \left\| C^t C \right\| \ ,$$

where $\|\cdot\|$ *denote the Euclidean norm of a matrix.*

Proof. When W is positive semi-definite we have $W = P^T D(r_1,\ldots r_k)P$, where P is an orthogonal matrix and $D(r_1,\ldots r_k)$ is the diagonal matrix whose principal elements are the non negative eigenvalues, $r_1,\ldots r_k$, of W. Thus $W = W^{1/2} W^{1/2}$ with

$$W^{1/2} = P^T D\left(r_1^{1/2},\ldots r_k^{1/2}\right)P$$

positive semi-definite since it has the eigenvalues $r_j^{1/2} \geq 0$, $j = 1,\ldots,k$.

Now $\| W \|^2$ will be the sum of the squares of the Euclidean norms of column vectors or row vectors of matrix W, thus

$$\| W \|^2 = \left\| P^T D(r_1,\ldots r_k)P \right\|^2 = \left\| P P^T D(r_1,\ldots r_k) P P^T \right\|^2 = \| D(r_1,\ldots r_k) \|^2 = \sum_{j=1}^{k} r_j^2$$

since pre or post multiplying a matrix by an orthogonal matrix, in this case P and P^T, does not alter the Euclidean norm of its column vectors or its row vectors. Moreover MM^T and $M^T M$ are, see Mexia (1995), positive semi-definite with the same non null eigenvalues thus

$$\left\| MM^T \right\| = \left\| M^T M \right\| ,$$

so

$$\left\| CWC^T \right\| = \left\| C W^{1/2} W^{1/2} C^T \right\| = \left\| W^{1/2} C^T C W^{1/2} \right\|.$$

The Euclidean norm defined in (2.1.1) is a matrix norm, see Schott (1997), so with A and B matrices of order k, the inequality $\| AB \| \leq \| A \| \| B \|$ holds.

Thus

$$\left\| W^{1/2} \left(C^T C\right) W^{1/2} \right\| \leq \left\| W^{1/2} \right\| \left\| C^T C \right\| \left\| W^{1/2} \right\|$$

□

We will use Lemma 3.5 to obtain an upper bound for the Euclidean norm of the variance-covariance matrix of the estimator of $\underline{\sigma}^2$ and show that using $\left(B_2^T\right)^+$ we minimize that bound.

Since B_2 is HF the single value decomposition of B_2^T gives, see Schott (1997, pg.131),

$$B_2^T = P^T \begin{bmatrix} \Delta \\ 0 \end{bmatrix} Q \qquad (3.3.8)$$

where P and Q are orthogonal matrices, Δ is a diagonal matrix and 0 is the null matrix. It is now easy to see that, if L is a left inverse of B_2^T, this is, if

$$L\,B_2^T = I_c \tag{3.3.9}$$

with $c = \text{rank}\left(B_2\right)$, we have, with an arbitrary sub-matrix U

$$L = Q^T \left[\ \Delta^{-1} \quad U\ \right] P \tag{3.3.10}$$

while $L^+ = \left(B_2^T\right)^+$ will be given by

$$L^+ = Q^T \left[\ \Delta^{-1} \quad 0\ \right] P\,. \tag{3.3.11}$$

Let $V\left(\underline{\tilde{\gamma}}(2)\right)$ be the variance-covariance matrix of

$$\underline{\tilde{\gamma}}_2 = \begin{bmatrix} \tilde{\gamma}_{z+1} \\ \vdots \\ \tilde{\gamma}_m \end{bmatrix} \tag{3.3.12}$$

then

$$\begin{cases} V\left(L\,\underline{\tilde{\gamma}}_2\right) = L\,V\left(\underline{\tilde{\gamma}}_2\right)L^T \\ V\left(L^+\,\underline{\tilde{\gamma}}_2\right) = L^+\,V\left(\underline{\tilde{\gamma}}_2\right){L^+}^T \end{cases} \tag{3.3.13}$$

and, according to Lemma 3.5

$$\begin{cases} \left\| V\left(L\,\underline{\tilde{\gamma}}_2\right)\right\| \le \left\| V\left(\underline{\tilde{\gamma}}_2\right)^{1/2}\right\|^2 \left\| L^T L\right\| \\ \left\| V\left(L^+\,\underline{\tilde{\gamma}}_2\right)\right\| \le \left\| V\left(\underline{\tilde{\gamma}}_2\right)^{1/2}\right\|^2 \left\| {L^+}^T L^+\right\| \end{cases}\,. \tag{3.3.14}$$

We now have the

Proposition 3.7. *Using* $L^+ = \left(B_2^T\right)^+$ *we minimize the upper bound for the Euclidean norm of the variance-covariance matrix of the estimator of* $\underline{\sigma}^2$.

Proof. According to (3.3.11) to (3.3.14) we have only to point out that

$$\left\| {L^+}^T L^+ \right\| = \left\| \begin{bmatrix} \Delta^{-2} & 0 \\ 0 & 0 \end{bmatrix} \right\| \le \left\| \begin{bmatrix} \Delta^{-2} & \Delta^{-1} U \\ U^T \Delta^{-1} & U^T U \end{bmatrix} \right\| \le \left\| L^T L \right\|$$

to establish the thesis.

□

We now consider the matching. This case corresponds to having $z^0 = 0$ and so

$$\begin{cases} \gamma_j = \gamma_{j+z} & , \ j = 1,\dots,z \\ b_{i,j} = b_{i,j+z} & , \ i = 1,\dots,w \ , \ j = 1,\dots z \end{cases} \tag{3.3.15}$$

so that the columns of B_1 are identical to the first z columns of B_2.

Clearly, while estimating the $\gamma_{z+1},\dots,\gamma_m$ we are also, in this case, estimating the $\gamma_1,\dots,\gamma_z$.

We considered in Section 2.2.2 the partition of the observation vectors $\underline{Y}$ in sub-vectors $\underline{Y}_1,\dots,\underline{Y}_{n^0}$ with mean vectors $\mu_1 \underline{1}_{r_1},\dots,\mu_{n^0} \underline{1}_{r_{n^0}}$, so that $X_0 = D\left(\underline{1}_{r_1} \cdots \underline{1}_{r_{n^0}}\right)$, the block diagonal matrix with principal blocks $\underline{1}_{r_1} \cdots \underline{1}_{r_{n^0}}$, and $\underline{\beta}_0$ has the components $\beta_{0,i} = \mu_i$, $i = 1,\dots,n^0$. Thus

$$T = D\left(\frac{1}{r_1} J_{r_1} \cdots \frac{1}{r_{n^0}} J_{r_{n^0}}\right). \tag{3.3.16}$$

If the variance-covariance matrix is defined by

$$V = D\left(\gamma_1 I_{r_1} \cdots \gamma_{n^0} I_{r_{n^0}}\right) \tag{3.3.17}$$

we can consider

$$\underline{Q} = \left\{Q_1,\cdots,Q_{2n^0}\right\}, \tag{3.3.18}$$

with

$$Q_j = D\left(Q_{j,1},\dots,Q_{j,n^0}\right), \ j = 1,\dots,m = 2n^0 \tag{3.3.19}$$

where

$$\begin{cases} Q_{j,j} = \frac{1}{r_j} J_{r_j} & ; \quad j = 1,\dots,n^0 \\ Q_{j+n^0,j} = K_{r_j} & ; \quad j = 1,\dots,n^0 \\ Q_{j,l} = Q_{j+n^0,l} = 0_{r_l \times r_l} & ; \quad l \neq j \quad ; \quad j = 1,\dots,n^0 \end{cases} , \tag{3.3.20}$$

with $k_{r_j} = I_{r_j} - \frac{1}{r_j} J_{r_j}$, as principal basis of the corresponding CJAS. Then

$$\begin{cases} z^0 = 0 \\ z = n^0 \end{cases}, \tag{3.3.21}$$

and

$$V=\sum_{j=1}^{n^0}\gamma_j\left(Q_j+Q_{j+n^0}\right) \tag{3.3.22}$$

Since $z^0=0$, this is, there is matching, we say that these models will be MEO.

For instance if

$$\underline{Y}=\sum_{j=1}^{n^0}X_i\,\underline{\beta}_i \tag{3.3.23}$$

with $\underline{\beta}_0$ fixed and $\underline{\beta}_1,\dots,\underline{\beta}_{n^0}$ independents, with null mean vector and variance-covariance matrices $\sigma_1^2 I_{c_1},\dots,\sigma_{n^0}^2 I_{c_{n^0}}$, with

$$c_i=\sum_{l=1}^{i}r_l+n^0-i\,,\ i=1,\dots,n^0\ ,$$

while

$$\begin{cases} X_0=D\left(1_{r_1}\ \cdots\ 1_{r_{n^0}}\right) \\ X_i=D\left(X_{i,1},\dots,X_{i,n^0}\right)\ ,\ i=1,\dots,n^0 \end{cases}, \tag{3.3.24}$$

with

$$\begin{cases} X_{i,l}=I_{r_l} & ,\ l\le i\ ,\ i=1,\dots,n^0 \\ X_{i,l}=\underline{0}_{r_l} & ,\ l>i\ ,\ i=1,\dots,n^0 \end{cases}, \tag{3.3.25}$$

we have a model of the type described above.

Since

$$M_i=D\left(M_{i,1},\dots,M_{i,n^0}\right),\ i=1,\dots,n^0 \tag{3.3.26}$$

with

$$\begin{cases} M_{i,l}=I_{r_l}=\frac{1}{r_l}J_{r_l}+k_{r_l} & ,\ l=1,\dots,i\ ,\ i=1,\dots,n^0 \\ M_{i,l}=0_{r_l\times r_l} & ,\ l>i\ ,\ i=1,\dots,n^0 \end{cases} \tag{3.3.27}$$

it is easy to see that we have, as expected, a MEO with

$$B_1=B_2=\begin{bmatrix}1&0&\cdots&0\\1&1&\cdots&0\\\vdots&\vdots&&\vdots\\1&1&\cdots&1\end{bmatrix}. \tag{3.3.28}$$

If the model is CEO we can also consider the transition matrix $\overline{B}$ for which we will have a partition

$$\overline{B} = \begin{bmatrix} \overline{b}^T & \underline{0}^T \\ \overline{B}_1 & \overline{B}_2 \end{bmatrix}, \tag{3.3.29}$$

where $\overline{b}$ $\left[\overline{B}_1\right]$ has $\overline{z}$ components [columns]. Reasoning as before, we will have

$$V = \sum_{j=1}^{\overline{m}} \overline{\gamma}_j \overline{Q}_j \tag{3.3.30}$$

with

$$\overline{\gamma}_j = \sum_{i=1}^{w} \overline{b}_{i,j}\sigma_i^2 \quad , \quad j = 1,\dots,\overline{m}.$$

Then with

$$\underline{\overline{\gamma}}_1 = \begin{bmatrix} \overline{\gamma}_1 \\ \vdots \\ \overline{\gamma}_z \end{bmatrix} \quad , \quad \underline{\overline{\gamma}}_2 = \begin{bmatrix} \overline{\gamma}_{z+1} \\ \vdots \\ \overline{\gamma}_{\overline{m}} \end{bmatrix}$$

we will have

$$\underline{\overline{\gamma}}_l = \overline{B}_l^T \underline{\sigma}^2 , \; l = 1,2 \tag{3.3.31}$$

as well as

$$\underline{\sigma}^2 = \left(\overline{B}_2^T \right)^+ \underline{\overline{\gamma}}_2 \ , \tag{3.3.32}$$

when the row vectors of $\overline{B}_2$ are linearly independent. Then the CEO model will be segregated, SCEO. Then segregation for CEO follows the same pattern as in the general case. The same is true for matching. Thereby CEO with matching will be MCEO.

3.4. Model Crossing

Now we are interested in studying model crossing. This technique enables us to obtain complex models from simple ones.

The study of model crossing using CJA was introduced by Fonseca et al (2006).

Let us consider first the case where a single model is studied.

We say that two factors of a model are crossed when every level of one factor occurs with every level of the other factor. When there is crossing between the factors, the treatments are formed as the combinations of all levels of the factors.

Suppose a model has u factors with $a_1,\dots,a_u$ levels. When each one of the a_1 levels of the first factor is combined with the a_2 levels of the second factor, these a_2 levels are combined with the a_3 levels of the third factor and so on until the a_u levels of the u-th factor, we are dealing with crossed factors and we obtain

$$c = \prod_{i=1}^{n} a_i$$

treatments.

The next figure is a schematic representation of the case where we have crossing of three factors, u=3, with $a_1 = 2$, $a_2 = 2$ and $a_3 = 4$ levels, respectively.

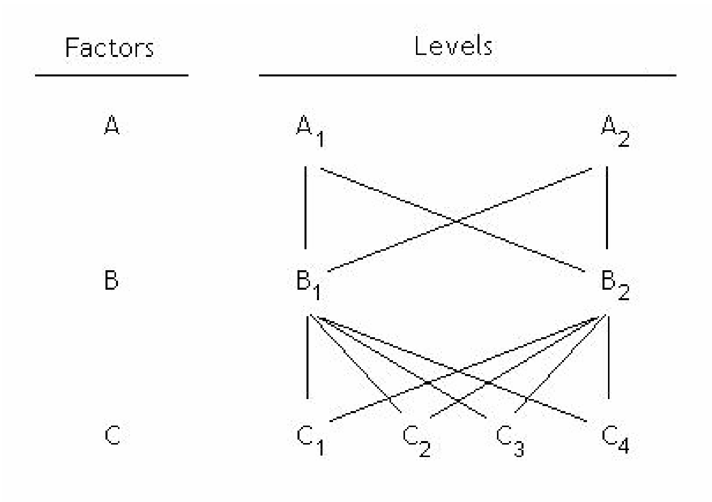


Figure 3.1: Factors crossing

In this case, the total number of treatments will be c = 16.

Now consider u models, each one with only one factor with $a_1,\dots,a_u$ levels. Crossing these models we obtain the same combination of levels we had above, thus the same number of treatments.

The next figure is a schematic representation of the case where three models are crossed, each one with only one factor with $a_1 = 2$, $a_2 = 2$ and $a_3 = 4$ levels respectively.

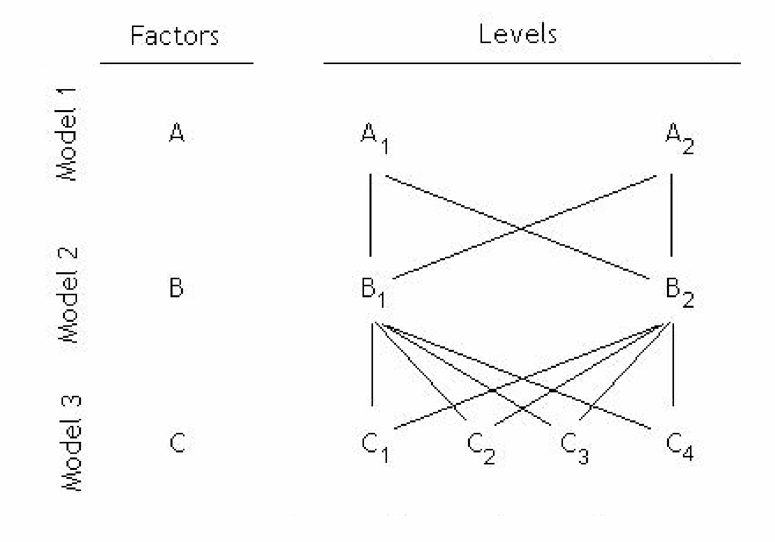


Figure 3.2: Model Crossing

Generalizing this concept, we can cross several models, each one of them with more than one factor.

Crossing models we obtain a model where treatments are all the possible combinations of the treatments of the initial models.

Let

$$\underline{Y}(l) = \sum_{i=0}^{w(l)} X_i(l)\underline{\beta}_i(l) \quad , \; l=1,2 \qquad (3.4.1)$$

be mixed models where $\beta_0(1)$ and $\beta_0(2)$ are fixed while the remaining vectors are independent with null mean vectors and variance-covariance matrices $\sigma^2_{i'}(1) I_{c_{i'}(1)}$, $i' = 1,\dots,w(1)$ and $\sigma^2_{i''}(2) I_{c_{i''}(2)}$, $i'' = 1,\dots,w(2)$. Crossing these models we get, see Fonseca et al (2006), the model

$$\underline{Y} = \sum_{i'=0}^{w(1)}\sum_{i''=0}^{w(2)} (X_{i'}(1) \otimes X_{i''}(2))\beta_{i',i''} \qquad (3.4.2)$$

where $\underline{\beta}_{0,0}$ will be fixed and the remaining are independent with null mean vectors and variance-covariance matrices $\sigma^2_{i',i''} I_{c_{i',i''}}$, $i' + i'' > 0$, with $c_{i',i''} = c_{i'}(1)\, c_{i''}(2)$, $i' + i'' > 0$.

Let us assume the initial models, in (3.4.1), to be CEO with transition matrices $\bar{B}(l) = [\bar{b}_{i,j}(l)]$, $l = 1,2$, and principal basis

$$\underline{Q}(l) = \{\bar{Q}_1(l),\dots,\bar{Q}_{\bar{m}(l)}(l)\} \; , \; l=1,2 \qquad (3.4.3)$$

for the corresponding CJAS $\overline{A}(l)$, $l=1,2$. Then the orthogonal projection matrices, on the $R(X_0(l))$, $l=1,2$, will be

$$T(l)=X_0(l)X_0(l)^{+}=\sum_{j=1}^{\bar{z}(l)}\overline{Q}_j(l)\ ,\ l=1,2\ , \tag{3.4.4}$$

where $\overline{z}(l)<\overline{m}(l)$, and, for the transition matrices, we will have the partitions

$$\overline{B}(l)=\begin{bmatrix} \overline{\underline{b}}(l)^{T} & \underline{0}^{T} \\ \overline{B}(l,1) & \overline{B}(l,2) \end{bmatrix}\ ,\ l=1,2\ , \tag{3.4.5}$$

where $\overline{\underline{b}}(l)$ [$\overline{B}(l,1)$] has $\overline{z}(l)$ components [columns], $l=1,2$.

The model obtained through crossing, defined in (3.4.2), will have mean vector

$$\underline{\mu}=X_0\,\underline{\beta}_0 \tag{3.4.6}$$

with

$$X_0=X_0(1)\otimes X_0(2) \tag{3.4.7}$$

and

$$\underline{\beta}_0=\underline{\beta}_{0,0}\,. \tag{3.4.8}$$

Then the orthogonal projection matrix on the space spanned by $\underline{\mu}$ will be

$$\begin{aligned} T=X_0\,X_0^{+}&=(X_0(1)\otimes X_0(2))(X_0(1)\otimes X_0(2))^{+}=(X_0(1)\otimes X_0(2))\left(X_0(1)^{+}\otimes X_0(2)^{+}\right)\\ &=\left(X_0(1)\,X_0(1)^{+}\right)\otimes\left(X_0(2)\,X_0(2)^{+}\right)=T(1)\otimes T(2)\ , \end{aligned}$$

thus

$$T=\left(\sum_{j'=1}^{\bar{z}(1)}\overline{Q}_{j'}(1)\right)\otimes\left(\sum_{j''=1}^{\bar{z}(2)}\overline{Q}_{j''}(2)\right)=\sum_{j'=1}^{\bar{z}(1)}\sum_{j''=1}^{\bar{z}(2)}\overline{Q}_{j'}(1)\otimes\overline{Q}_{j''}(2)=\sum_{j'=1}^{\bar{z}(1)}\sum_{j''=1}^{\bar{z}(2)}\overline{Q}_{j',j''} \tag{3.4.9}$$

with $\overline{Q}_{j',j''}=\overline{Q}_{j'}(1)\otimes\overline{Q}_{j''}(2)$; $j'=1,\ldots,\overline{m}(1)$; $j''=1,\ldots,\overline{m}(2)$.

Moreover

$$\begin{aligned} M_{i',i''}&=(X_{i'}(1)\otimes X_{i''}(2))(X_{i'}(1)\otimes X_{i''}(2))^{T}=(X_{i'}(1)\otimes X_{i''}(2))\left(X_{i'}(1)^{T}\otimes X_{i''}(2)^{T}\right)\\ &=\left(X_{i'}(1)X_{i'}(1)^{T}\right)\otimes\left(X_{i''}(2)X_{i''}(2)^{T}\right)=M_{i'}(1)\otimes M_{i''}(2)\ , \end{aligned}$$

$i' = 0,\ldots,w(1)$, $i'' = 0,\ldots,w(2)$, and so, with

$$M_{i'}(1) = \sum_{j'=1}^{\overline{m}(1)} \overline{b}_{i'+1,j'}(1)\overline{Q}_{j'}(1)$$

and

$$M_{i''}(2) = \sum_{j''=1}^{\overline{m}(2)} \overline{b}_{i''+1,j''}(2)\overline{Q}_{j''}(2)$$

we have

$$M_{i',i''} = \sum_{j'=1}^{\overline{m}(1)} \sum_{j''=1}^{\overline{m}(2)} \overline{b}_{i'+1,j'}(1)\overline{b}_{i''+1,j''}(2)\overline{Q}_{j',j''} \quad , \tag{3.4.10}$$

$i' = 0,\ldots,w(1)$, $i'' = 0,\ldots,w(2)$. Thus these matrices commute and so the model defined in (3.4.2) is CEO.

The coefficients in the right-hand member of (3.4.10) are the elements of $\overline{B}(1)\otimes\overline{B}(2)$. Reordering the rows and columns of this matrix we get the transition matrix

$$\overline{B} = \begin{bmatrix} \underline{\overline{b}}^{T} & \underline{0}^{T} \\ \overline{B}(1) & \overline{B}(2) \end{bmatrix} \tag{3.4.11}$$

with

$$\underline{\overline{b}} = \underline{\overline{b}}(1) \otimes \underline{\overline{b}}(2) \ , \tag{3.4.12}$$

$$\overline{B}(1) = \begin{bmatrix} \underline{\overline{b}}(1)^{T} \otimes \overline{B}(2,1) \\ \overline{B}(1,1) \otimes \underline{\overline{b}}(2)^{T} \\ \overline{B}(1,1) \otimes \overline{B}(2,1) \end{bmatrix} \tag{3.4.13}$$

and

$$\overline{B}(2) = \begin{bmatrix} \underline{\overline{b}}(1)^{T} \otimes \overline{B}(2,2) & \underline{0} & \underline{0} \\ 0 & \overline{B}(1,2) \otimes \underline{\overline{b}}(2)^{T} & 0 \\ \overline{B}(1,1) \otimes \overline{B}(2,2) & \overline{B}(1,2) \otimes \overline{B}(2,1) & \overline{B}(1,2) \otimes \overline{B}(2,2) \end{bmatrix} . \tag{3.4.14}$$

We now have

Proposition 3.8. *Crossing SCEO gives SCEO.*

Proof. To establish the thesis it is sufficient to point out that, if the matrices $\overline{B}(l,2)$ are HF, $l = 1,2$, so is $\overline{B}(2)$.

□

For the model obtained through crossing not to enjoy matching, does not being MCEO, there must be a pair (j', j'') such that

$$\overline{Q}_{j',j''}T = \overline{Q}_{j',j''}. \tag{3.4.15}$$

Now

$$\overline{Q}_{j',j''}T = \left(\overline{Q}_{j'}(1) \otimes \overline{Q}_{j''}(2)\right)\left(T(1) \otimes T(2)\right) = \left(\overline{Q}_{j'}(1)T(1)\right) \otimes \left(\overline{Q}_{j''}(2)T(2)\right)$$

so, for having

$$\overline{Q}_{j',j''}T = \overline{Q}_{j',j''} = \overline{Q}_{j'}(1) \otimes \overline{Q}_{j''}(2)$$

we must have

$$\begin{cases} \overline{Q}_{j'}(1)T(1) = \overline{Q}_{j'}(1) \\ \overline{Q}_{j''}(2)T(2) = \overline{Q}_{j''}(2) \end{cases},$$

so that neither of the usual model would be MCEO. Thus, if one of the initial models is MCEO so is the model obtained through crossing.

3.5. Model Nesting

In order to introduce the study of model nesting, let us consider first the case when there is a single model whose factors are nested. We say that one factor is nested within another when any given level of the nested factor appears at only one level of the nesting factor, this is, when the levels of the nested factor are divided among the levels of the nesting factor.

Suppose a model has u factors with $a_1,\ldots,a_u$ levels, respectively. We have balanced nesting when the a_i, $i = 2,\ldots,u$, levels of a factor are divided evenly for the a_{i-1}, $i = 2,\ldots,u$, levels of the preceding factor. Thus we would have $\prod_{i=1}^{u} a_i$ treatments, this is, the number of treatments of balanced nested models is the product of the number of the levels of the factors, which imposes strong restrictions on these numbers.

The next figure is a schematic representation of an example of balanced nesting of three factors, u=3, with $a_1 = 2$, $a_2 = 6$ and $a_3 = 12$ levels respectively.

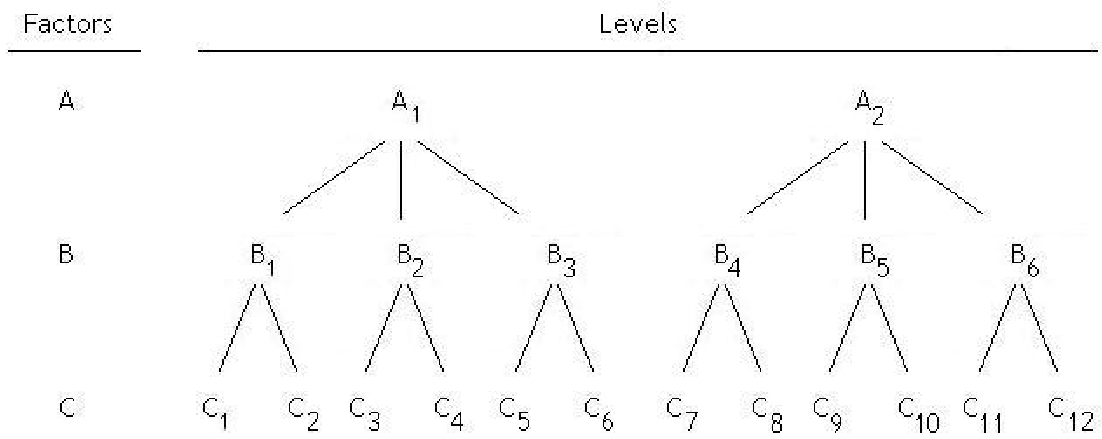


Figure 3.3: Factors balanced nesting

When there is an uneven distribution of the levels of a given factor by the levels of the preceding factor, the nesting is called unbalanced.

The next figure is a schematic representation of an example of unbalanced nested of three factors, u=3, with $a_1 = 2$, $a_2 = 5$ and $a_3 = 12$ levels respectively.

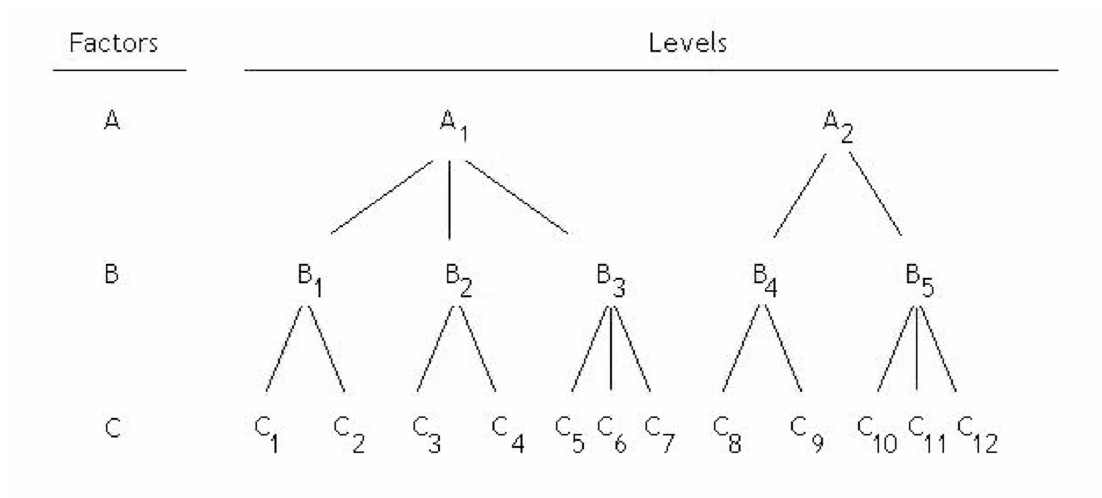


Figure 3.4: Factors unbalanced nesting

Although balanced nesting is the most usual form of nesting, it has a great disadvantage: the number of treatments may be too large. To overcome this disadvantage we can use unbalanced nested designs, as the one we present in Section 3.7.

Analogously to what was seen previously, in the case of crossing, nesting u factors of a model is equivalent to nesting u models each one with only one factor. Generalizing this concept, we can nest several models, each one of them with more than one factor. This is what we call model nesting.

In model nesting each treatment of a model nests all the treatments of another model.

The study of model nesting using CJA was introduced by Fonseca et al (2006).

Since random effect factors do not nest fixed effect factors, we have to consider two cases. In the first case the nesting model is mixed and in the second case the nesting model has fixed effects.

3.5.1. First case

Let the nesting model be a mixed model given by

$$\underline{Y}(1) = \sum_{i=0}^{w(1)} X_i(1)\underline{\beta}_i(1) \ , \tag{3.5.1}$$

with $\underline{\beta}_0(1)$ fixed and the $\underline{\beta}_1(1) \dots \underline{\beta}_{w(1)}(1)$ independent with null mean vector and variance-covariance matrices $\sigma_i^2(1)\, I_{c_i(1)}$, $i = 1,\dots,w(1)$, and the nested model, a random effects model given by

$$\underline{Y}(2) = \sum_{i=0}^{w(2)} X_i(2)\underline{\beta}_i(2) \ , \tag{3.5.2}$$

with $X_0(2) = 1_{n(2)}$ and the $\underline{\beta}_1(2) \dots \underline{\beta}_{w(2)}(2)$ independent with null mean vector and variance-covariance matrices $\sigma_i^2(2)\, I_{c_i(2)}$, $i = 1,\dots,w(2)$. The main difference between both models is on the fixed term that in the second models reduces to $1_{n(2)}\, \underline{\mu}(2)$.

In what follows we assume both models to be EO.

Through nesting we get the model

$$\underline{Y} = \sum_{i=0}^{w(1)} \left(X_i(1) \otimes 1_{n(2)}\right)\underline{\beta}_i + \sum_{i=w(1)+1}^{w} \left(I_{n(1)} \otimes X_{i-w(1)}(2)\right)\underline{\beta}_i \tag{3.5.3}$$

where $n(1)$ [$n(2)$] is the number of observations for the first (nesting) model [second(nested) model], $w = w(1) + w(2)$, $\underline{\beta}_0$ is fixed and the $\underline{\beta}_1 \dots \underline{\beta}_w$ are independent with null mean vector and variance-covariance matrices $\sigma_i^2\ I_{c_i}$, $i = 1,\dots,w$.

The mean vector of the model in (3.5.3) will be

$$\underline{\mu} = \left(X_0(1) \otimes 1_{n(2)}\right)\underline{\beta}_0 , \tag{3.5.4}$$

so the orthogonal projection matrix on Ω , the space spanned by $\underline{\mu}$, will be

$$\begin{aligned} T &= \left(X_0(1)\otimes 1_{n(2)}\right)\left(X_0(1)\otimes 1_{n(2)}\right)^{+} = \left(X_0(1)\otimes 1_{n(2)}\right)\left(X_0(1)^{+}\otimes\frac{1}{n(2)}\ 1_{n(2)}^{T}\right) \\ &= \left(X_0(1)\,X_0(1)^{+}\right)\otimes\left(\frac{1}{n(2)}\ 1_{n(2)}1_{n(2)}^{T}\right) = T(1)\otimes\left(\frac{1}{n(2)}J_{n(2)}\right) \end{aligned} \quad , \qquad (3.5.5)$$

with $T(1)$ the orthogonal projection matrix on $\Omega(1)$, the space spanned by the mean vector $X(1)\underline{\beta}_0(1)$ of the nesting model.

Let, for the nesting model,

$$\underline{Q}(1) = \{Q_1(1),\ldots,Q_{m(1)}(1)\} = pb(\mathcal{A}(1)), \qquad (3.5.6)$$

with $\mathcal{A}(1)$ the corresponding CJAS, thus

$$T(1) = \sum_{j=1}^{z(1)} Q_j(1). \qquad (3.5.7)$$

For the second initial model, the nested model, with mean vector $1_{n(2)}\,\mu(2)$, we have

$$T(2) = \frac{1}{n(2)}J_{n(2)} \qquad (3.5.8)$$

and

$$\underline{Q}(2) = \{Q_1(2),\ldots,Q_{m(2)}(2)\} = pb(\mathcal{A}(2)), \qquad (3.5.9)$$

with $\mathcal{A}(2)$ the corresponding CJAS. This must be regular since we will have

$$\frac{1}{n(2)}J_{n(2)} = T(2) = Q_1(2) \qquad (3.5.10)$$

and so $z(2)=1$.

Since both initial models are EO, we will have

$$\sum_{j=1}^{m(l)} Q_j(l) = I_{n(l)} \quad l = 1,2 \ . \qquad (3.5.11)$$

For the model obtained through nesting we have, see Fonseca et al (2006), a CJAS that is the restricted Kronecker product between the CJAS corresponding to the initial models,

$$\mathcal{A} = \mathcal{A}(1) * \mathcal{A}(2),$$

with, according to Proposition 2.15, the principal basis

$$\underline{Q} = \left\{ Q_j(1) \otimes \frac{1}{n(2)} J_{n(2)};\ j = 1,\ldots,m(1) \right\} \cup \left\{ I_{n(1)} \otimes Q_j(2);\ j = 2,\ldots,m(2) \right\} \\ = \left\{ Q_j;\ j = 1,\ldots,m \right\} \quad , \tag{3.5.12}$$

where $m = m(1) + m(2) - 1$.

For the model obtained through nesting, we have the matrices

$$\begin{cases} M_i = M_i(1) \otimes J_{n(2)} & i = 0,\ldots,w(1) \\ M_i = I_{n(1)} \otimes M_{i-w(1)}(2) & i = w(1)+1,\ldots,w \end{cases}. \tag{3.5.13}$$

We now establish

Proposition 3.9. *The model derived through nesting is EO.*

Proof. Matrices M_i, $i = 0.\ldots,w$ belong to A, so the variance-covariance matrix of the model will also belong to A. Moreover, since

$$\sum_{j=1}^{m(1)} Q_j = \left(\sum_{j=1}^{m(1)} Q_j(1) \right) \otimes \frac{1}{n(2)} J_{n(2)} = I_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)}$$

we have

$$\sum_{j=1}^{m} Q_j = I_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)} + \sum_{j=m(1)+1}^{m} Q_j = I_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)} + I_{n(1)} \otimes \sum_{j=2}^{m(2)} Q_j(2) =$$

$$= I_{n(1)} \otimes \sum_{j=1}^{m(2)} Q_j(2) = I_{n(1)} \otimes I_{n(2)} = I_n \quad ,$$

with $n = n(1)\,n(2)$ the number of observations for the model obtained through nesting. To complete the proof we have only to point out that $TQ_j = Q_j T$, $j = 1,\ldots,m$ so the model is EO.

□

We had, for the initial models, transition matrices $B(l) = \lfloor b_{i,j}(l) \rfloor$, $l = 1,2$, so

$$M_i(l) = \sum_{j=1}^{m(l)} b_{i,j}(l)\, Q_j(l), \quad i = 1,\ldots,m(l), \quad l = 1,2 \tag{3.5.14}$$

thus

$$M_i = M_i(1) \otimes J_{n(2)} = n(2) \sum_{j=1}^{m(1)} b_{i,j}(1) Q_j(1) \otimes \frac{1}{n(2)} J_{n(2)} = \sum_{j=1}^{m(1)} b_{i,j} Q_j \quad , \ i = 1,\ldots,w(l), \tag{3.5.15}$$

with

$$b_{i,j} = n(2)\, b_{i,j}(1)\ ,\ \ j = 1,\dots,m(1),\ \ i = 1,\dots,w(1)\ . \qquad (3.5.16)$$

Moreover, since

$$Q_1(2) = \frac{1}{n(2)} J_{n(2)},$$

and

$$\sum_{j=1}^{m(1)} Q_j = I_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)} = I_{n(1)} \otimes Q_1(2),$$

we have

$$\begin{aligned} M_i &= I_{n(1)} \otimes M_{i-w(1)}(2) = I_{n(1)} \otimes \left(\sum_{j=1}^{m(2)} b_{i-w(1),j}(2) Q_j(2) \right) \\ &= b_{i-w(1),1}(2) \left(I_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)} \right) + \sum_{j=2}^{m(2)} b_{i-w(1),j}(2) \left(I_{n(1)} \otimes Q_j(2) \right) \\ &= b_{i-w(1),1}(2) \sum_{j=1}^{m(1)} Q_j + \sum_{j=m(1)+1}^{m} b_{i-w(1),j-m(1)+1}(2)\, Q_j \\ &= \sum_{j=1}^{m} b_{i,j} Q_j \end{aligned} \qquad (3.5.17)$$

with

$$\begin{cases} b_{i,j} = b_{i-w(1),1}(2) & ,\ j = 1,\dots,m(1) & ,\ i = w(1)+1,\dots,w \\ b_{i,j} = b_{i-w(1),j-m(1)+1}(2) & ,\ j = m(1)+1,\dots,m & ,\ i = w(1)+1,\dots,w \end{cases} .$$

Thus, with

$$\begin{cases} B(1) = [\, B_1(1) \quad B_2(1)\,] \\ B(2) = [\, \underline{b}_1(2) \quad B_2(2)\,] \end{cases} , \qquad (3.5.18)$$

where $B_1(2) = \underline{b}_1(2)$ is reduced to a single column since $z(2) = 1$ due to the nested model only having random effects factors, we will have for the model obtained through nesting

$$B = [\, B_1 \quad B_2\,] \qquad (3.5.19)$$

with

$$B_1 = \begin{bmatrix} n(2) & B_1(1) \\ \underline{b}_1(2) & 1_{z(1)}^T \end{bmatrix}$$

and

$$B_2 = \begin{bmatrix} n(2) & B_1(2) & 0 \\ \underline{b}_1(2) & 1^T_{m(1)-z(1)} & B_2(2) \end{bmatrix}.$$

We now establish

Proposition 3.10. *Nesting SEO gives SEO and if the nesting model is MEO so is the model obtained through nesting.*

Proof. We have $\text{rank}(B_1(2)) + \text{rank}(B_2(2)) = \text{rank}(B_2) \le w_1 + w_2$ since B_2 has $w_1 + w_2$ rows. Thus if the initial models are SEO and $\text{rank}(B_2(l)) = w_l$, $l = 1,2$, we have $\text{rank}(B_2) = w_1 + w_2$ and the first part of the thesis is established.
Moreover

$$T = T(1) \otimes \frac{1}{n(2)} J_{n(2)}$$

so

$$\begin{cases} TQ_j = \left(T(1)Q_j(1)\right) \otimes \frac{1}{n(2)} J_{n(2)} & ; \ j = 1, \cdots, m(1) \\ TQ_j = \left(T(1)\, I_{n(1)}\right) \otimes \frac{1}{n(2)} J_{n(2)}\, Q_{j-m(1)+1}(2) & ; \ j = m(1)+1, \cdots, m \end{cases}$$

and so $TQ_j = Q_j$ is only possible with $j \le z(1)$ and when $T(1)Q_j(1) = Q_j(1)$.Thus if the nesting model is MEO, and so $TQ_j \ne Q_j$, $j \le z(1)$ the model obtained through nesting will also be MEO.

□

3.5.2. Second case

We now consider that the nesting model has fixed effects and the nested model may be mixed. Since the first (nesting) model only have fixed effects factors we may replace it by

$$\underline{Y}(1) = I_{n(1)} \underline{\beta}_o(1) \quad , \tag{3.5.20}$$

thus lightning the writing. Now $n(1)$ will be the number of treatments in the nesting model and the components of $\underline{\beta}_o(1)$ are their mean values. Thus we will have

$$w(1) = m(1) = z(1) = 1$$

and

$$\underline{Q}(1)=\{I_{n(1)}\}. \tag{3.5.21}$$

We assume the basis

$$\underline{Q}(2)=\{Q_1(2),\ldots,Q_{m(2)}(2)\} \quad , \tag{3.5.22}$$

of the CJAS $\mathcal{A}(2)$ corresponding to the nested model, to have

$$\begin{cases} Q_1(2)=\dfrac{1}{n(2)}J_{n(2)} \\ \sum_{j=1}^{m(2)} Q_j(2)=I_{n(2)} \end{cases} \quad , \tag{3.5.23}$$

with $n(2)$ the number of observations for the nested model. Thus the CJAS $\mathcal{A}(2)$ will be regular and complete.

With $\mathcal{A}(1)$ the CJAS corresponding to the nesting model we will have

$$\mathcal{A}=\mathcal{A}(1)*\mathcal{A}(2)=\mathcal{A}(1)\otimes\mathcal{A}(2).$$

Once the principal basis of $\mathcal{A}$ is , with $m=m(2)$,

$$\begin{aligned}\underline{Q}&=\left\{\ I_{n(1)}\otimes\frac{1}{n(2)}J_{n(2)}\ ,\ I_{n(1)}\otimes Q_2(2)\ ,\ \ldots\ ,\ I_{n(1)}\otimes Q_{m(2)}(2)\right\}\\ &=\{\,Q_1\,,\ldots,Q_m\,\}\end{aligned} \tag{3.5.24}$$

with

$$Q_j=I_{n(1)}\otimes Q_j(2)\,,\,j=1,\ldots,m=m(2).$$

Given the nested model

$$\underline{Y}(2)=\sum_{i=0}^{w(2)}X_i(2)\underline{\beta}_i(2) \tag{3.5.25}$$

let us assume that

$$X_i(2)=\left[\underline{1}_{n(2)}\quad X_0^o(2)\ \right] \tag{3.5.26}$$

with

$$\underline{1}_{n(2)}^T\,X_0^o(2)=\underline{0}_{c_0(2)}^T \quad , \tag{3.5.27}$$

thus $\underline{1}_{n(2)}$ is assumed to be orthogonal to the column vectors of $X_0^o(2)$ whose sums of components will be null. Thus

$$\begin{cases} T(2) = X_0(2)X_0(2)^+ = \sum_{j=1}^{z(2)} Q_j(2) \\ T^o(2) = X_0^o(2)X_0^o(2)^+ = \sum_{j=2}^{z(2)} Q_j(2) \end{cases} . \quad (3.5.28)$$

Through nesting we get the model

$$\underline{Y} = \sum_{i=0}^{w} X_i \underline{\beta}_i \quad , \quad (3.5.29)$$

with $w = w(2)$ and

$$\begin{cases} X_0 = \left[I_{n(1)} \otimes 1_{n(2)} \quad I_{n(1)} \otimes X_0^o(2) \right] \\ X_i = I_{n(1)} \otimes X_i(2) \quad , \quad i = 2, \ldots, w \end{cases} . \quad (3.5.30)$$

Now

$$\begin{cases} \left(I_{n(1)} \otimes 1_{n(2)}\right)\left(I_{n(1)} \otimes 1_{n(2)}\right)^+ = \left(I_{n(1)} \otimes 1_{n(2)}\right)\left(I_{n(1)} \otimes \frac{1}{n(2)} 1_{n(2)}^T \right)^+ = I_{n(1)} \otimes \frac{1}{n(2)} J_{n(2)} = Q_1 \\ \left(I_{n(1)} \otimes X_0^o(2)\right)\left(I_{n(1)} \otimes X_0^o(2)\right)^+ = \left(I_{n(1)} \otimes X_0^o(2)\right)\left(I_{n(1)} \otimes X_0^o(2)^+\right) = I_{n(1)} \otimes \left(\sum_{j=2}^{z(2)} Q_j(2) \right) = \sum_{j=2}^{z(2)} Q_j \end{cases} ,$$

(3.5.31)

thus

$$T = \sum_{j=2}^{z} Q_j \, ,$$

with $z = z(2)$.

Moreover

$$M_i = I_{n(1)} \otimes M_i(2) = I_{n(1)} \otimes \left(\sum_{j=1}^{m(2)} b_{i,j}(2) Q_j(2) \right) = \sum_{j=1}^{m(2)} b_{i,j}(2) \left(I_{n(2)} \otimes Q_j(2) \right) = \sum_{j=1}^{m(2)} b_{i,j} Q_j \quad , \quad (3.5.32)$$

$i = 1, \ldots, w = w(2)$, with $b_{i,j} = b_{i,j}(2)$, $i = 1, \ldots, w$, $j = 1, \ldots, n$, thus

$$B = B(2)$$

and

$$B_i = B_i(2) \quad , \quad i = 1, 2 .$$

This gives us the

Proposition 3.11. *When the nesting model has fixed effects and the nested model is mixed the model obtained through nesting is EO being SEO and/or MEO if and only if the nested model is SEO and/or MEO.*

Proof. First T and the $M_i = \sum_{j=1}^{m} b_{i,j} Q_j$, $i = 1,\dots,w$, commute so , according to Proposition 3.3, the model is COBS thus EO. The rest of the proof follows from $B_i = B_i(2)$, $i = 1,2$.

□

3.6. Model Joining

The operation we now study, model joining, is another possible method to building up complex models from simple ones.

Consider the mixed models

$$\underline{Y}(l) = \sum_{i=0}^{w(l)} X_i(l)\underline{\beta}_i(l) \quad , \quad l = 1,\dots,h \ , \tag{3.6.1}$$

defined as in Definition 3.1, with observations vectors

$$\underline{Y}(1) = \begin{bmatrix} Y_1(1) \\ \vdots \\ Y_{w(1)}(1) \end{bmatrix} \ , \quad \underline{Y}(2) = \begin{bmatrix} Y_1(2) \\ \vdots \\ Y_{w(2)}(2) \end{bmatrix} \ , \ \dots \ , \ \underline{Y}(1) = \begin{bmatrix} Y_1(h) \\ \vdots \\ Y_{w(h)}(h) \end{bmatrix} . \tag{3.6.2}$$

When we join these models we obtain a model with observations vector

$$\underline{Y} = \begin{bmatrix} Y_1(1) \\ \vdots \\ Y_{w(1)}(1) \\ Y_1(2) \\ \vdots \\ Y_{w(2)}(2) \\ \vdots \\ Y_1(h) \\ \vdots \\ Y_{w(h)}(h) \end{bmatrix} = \left[\underline{Y}(1)^T \ \dots \ \underline{Y}(h)^T\right]^T \tag{3.6.3}$$

for which we have the expression

$$\underline{Y} = \sum_{i=0}^{w} X_i \ \underline{\beta}_i \ . \tag{3.6.4}$$

Let

$$\underline{\mu}(l) = X_0(l)\underline{\beta}_0(l) \quad l = 1,\dots,h \tag{3.6.5}$$

be the mean vectors for the initial models , then the mean vector for the final model will be

$$\underline{\mu} = \left[\underline{\mu}(1)^T \dots \underline{\mu}(h)^T\right]^T = X_0\,\underline{\beta}_0 \ , \tag{3.6.6}$$

with

$$\begin{cases} X_0 = D(X_0(1)\dots X_0(h)) \\ \underline{\beta}_0 = \left[\underline{\beta}_0(1)^T \dots \underline{\beta}_0(h)^T\right]^T \end{cases} . \tag{3.6.7}$$

Now if we assume the $\underline{\beta}_i(l)$, $i = 1,\dots,w(l)$, $l = 1,\dots,h$ with null mean vector and variance-covariance matrices $\sigma_i^2(l)\, I_{c_i}(l)$, $i = 1,\dots,w(l)$, $l = 1,\dots,h$ it will be convenient to keep separated the terms originating from different initial models. With

$$\begin{cases} \overline{w}(0) = 0 \\ \overline{w}(l) = \sum_{k=1}^{l} w(k), \ l = 1,\dots,h \end{cases} \tag{3.6.8}$$

when

$$\overline{w}(l-1) < i \le \overline{w}(l)$$

we may take

$$X_i = \left[X_{i,1}^T, \dots, X_{i,h}^T\right]^T$$

where

$$\begin{cases} X_{i,i'} = 0_{n_{i'} \times c_{i-\overline{w}(l-1)}(l)} \\ X_{i,i} = X_{i-\overline{w}(l-1)}(l) \end{cases} ,$$

so

$$M_i = X_i X_i^T = D\left(M_{i,1} \dots M_{i,h}\right) \ , \tag{3.6.9}$$

with

$$\begin{cases} M_{i,i'} = 0_{n_{i'} \times n_{i'}} \\ M_{i,i} = M_{i-\overline{w}(l-1)}(l) \end{cases} .$$

Let now the initial models be EO with corresponding CJAS $\mathcal{A}(l)$ with principal basis

$$\underline{Q}(l) = \{Q_1(l), \dots, Q_{m(l)}(l)\} \tag{3.6.10}$$

and transition matrices $B(l)=\lfloor b_{i,j}(l)\rfloor$. Then, with $T(l)$ the orthogonal projection matrix on the space spanned by $\underline{\mu}(l)$, $l=1,\dots,h$, we will have

$$T(l)=\sum_{j=1}^{z(l)}Q_j(l) \quad , \; l=1,\dots,h \; . \tag{3.6.11}$$

We now consider the Cartesian product of the CJAS $\mathcal{A}(1),\dots,\mathcal{A}(h)$, the CJAS

$$\mathcal{A}=\mathop{\mathrm{X}}_{l=1}^{h}\mathcal{A}(l), \tag{3.6.12}$$

with principal basis

$$\underline{Q}=\{Q_1,\dots,Q_m\}, \tag{3.6.13}$$

where $m=\sum_{l=1}^{h}m(l)$. Putting

$$\begin{cases} \overline{m}(0)=0 \\ \overline{m}(l)=\sum_{k=1}^{l}m(k), l=1,\dots,h \end{cases},$$

when

$$\overline{m}(l-1)<j\leq\overline{m}(l) \; ,$$

we can take

$$Q_j=D(Q_{j,1},\dots,Q_{j,h}) \tag{3.6.14}$$

with

$$\begin{cases} Q_{j,l'}=0_{n(l')\times n(l')} \\ Q_{j,l}=Q_{j-\overline{m}(l-1)}(l) \end{cases},$$

where $n(l')$ is the number of observations for the l'- th initial model.

Now

$$M_{i'}(l)=\sum_{j=1}^{m(l)}b_{i',j}(l)Q_j(l) \quad , \; i'=1,\dots,w(l) \; , \; l=1,\dots,h \tag{3.6.15}$$

so, when $\overline{w}(l-1)<i'\leq\overline{w}(l)$

$$M_{i'} = \sum_{j=1}^{m(l)} b_{i'-\overline{w}(l-1),j}(l)Q_j(l) \ , \quad \overline{w}(l-1) < i' \leq \overline{w}(l) \ , \ l = 1,\dots,h \qquad (3.6.16)$$

and so for the joint model we will have the transition matrix

$$B = D\left(B(1) \dots B(h) \right) . \qquad (3.6.17)$$

Since for the $B(l), \ l = 1,\dots,h$, we have the matrix partitions

$$B(l) = [\ B_1(l) \quad B_2(l)\ ] \ , \qquad (3.6.18)$$

for B we will have

$$B_v = D\left(B_v(1) \dots B_v(h) \right) \ , \quad v = 1,2 \qquad (3.6.19)$$

It being straight forward to establish

Proposition 3.12. *If the initial models are SEO [MEO] so is the joined model.*

3.7. Step nesting

Now we will present another class of models, those of step nesting, introduced in Cox & Solomon (2003). As we will see, this is a useful alternative to balanced nesting since it leads to great economy in the number of observations.

In this study of step nesting models, see Fernandes (2009) and Fernandes et al (2010), the Cartesian product of CJAS is relevant.

Consider a model with u factors, with a_i, $i = 1,\dots,u$, active levels, respectively. In step nesting we will have u steps. In the first of these steps we will have a_1 levels of the first factor, each of which nests a level of each of the remaining factors. In the second step we have a unique level of the first factor, distinct from these of the first step, which nests a_2 levels of the second factor. Each of these a_2 levels nests an unique level of each of the remaining factors, and so on. At the end, the i-th factor will have

$$c_i = a_i + u - i \quad , \ i = 1,\dots,u$$

levels, a_i of these corresponding to the branching at the i-th step and $u - i$ construction levels.

According to the above defined, in step nesting, the total number of treatments will be

$$n = \sum_{i=1}^{u} a_i \ . \qquad (3.7.1)$$

The next figure is a schematic representation of the structure of step nesting with 3 factors, $u = 3$, where the factors have $a_1 = 3$, $a_2 = 2$ and $a_3 = 4$ active levels.

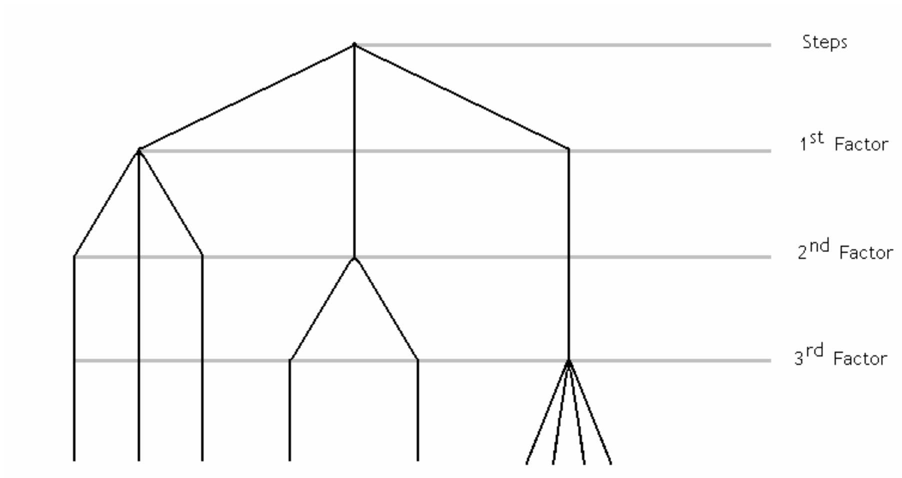


Figure 3.5: Step nesting, with $u = 3$, $a_1 = 3$, $a_2 = 2$ and $a_3 = 4$

In this case, the number of treatments will be $n = \sum_{i=1}^{3} a_i = 3 + 2 + 4 = 9$.

For the corresponding balanced nesting we would have the structure represented in the next figure.

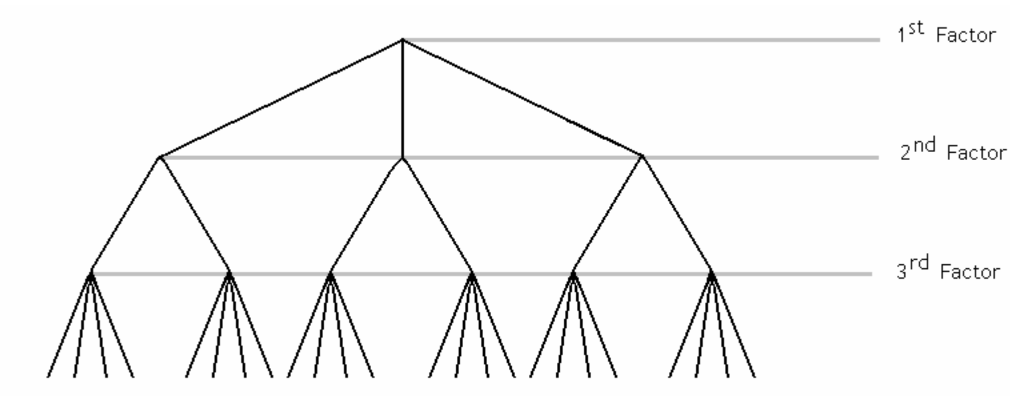


Figure 3.6: Balanced nesting, with $u = 3$, $a_1 = 3$, $a_2 = 2$ and $a_3 = 4$

As we saw in Section 3.5, the number of treatments of balanced nested models is the product of the number of the levels of the factors, so, in this case, we will have $\prod_{i=1}^{3} a_i \; 3 \times 2 \times 4 = 24$ treatments.

In general, the step nested model may be written as

$$Y = \sum_{i=0}^{u} X_i \underline{\beta}_i \tag{3.7.2}$$

where, as before, $\underline{\beta}_0$ is fixed and the $\underline{\beta}_1, \dots, \underline{\beta}_u$ are independent with null mean vectors and variance-covariance matrices $\sigma_i^2 \, I_{c_i}$, $i = 1,\dots,u$ and

$$X_i = D(X_{i,1},\dots,X_{i,u}),\ i = 0,\dots,u \tag{3.7.3}$$

where

$$\begin{cases} X_{i,l} = I_{a_l} & ; \quad l \le i \ ; \ i = 0,\dots,u \\ X_{i,l} = 1_{a_l} & ; \quad l > i \ ; \ i = 0,\dots,u \end{cases}$$

so

$$M_i = X_i X_i^T = D\left(M_{i,1} \dots M_{i,u} \right),\ i = 1,\dots,u \tag{3.7.4}$$

with

$$\begin{cases} M_{i,l} = I_{a_l} = \frac{1}{a_l} J_{a_l} + k_{a_l} & ; \quad l \le i \ ; \ i = 1,\dots,u \\ M_{i,l} = J_{a_l} = a_l \frac{1}{a_l} J_{a_l} & ; \quad l > i \ ; \ i = 1,\dots,u \end{cases}. \tag{3.7.5}$$

Let $\mathcal{A}(a_l)$, $l = 1,\dots,u$, be the CJAS with principal basis constituted by $\frac{1}{a_l} J_{a_l}$ and $k_{a_l} = I_{a_l} - \frac{1}{a_l} J_{a_l}$, $l = 1,\dots,u$. The Cartesian product of these CJAS,

$$\mathcal{A} = \mathop{\mathrm{X}}_{l=1}^{u} \mathcal{A}(a_l),$$

will have the principal basis constituted by the

$$Q_j = D\left(Q_{j,1} \dots Q_{j,u} \right),\ j = 1,\dots,2u \tag{3.7.6}$$

with

$$\begin{cases} Q_{j,l} = Q_{j+u,l} = 0_{a_l \times a_l} & ; \ l \ne j \ ; \ j = 1, \dots, u \\ Q_{j,l} = \frac{1}{a_j} J_{a_j} & ; \ j = 1, \dots, u \\ Q_{j+u,l} = k_{a_j} & ; \ j = 1, \dots, u \end{cases}. \tag{3.7.7}$$

Since

$$X_0 = D\left(\underline{1}_{a_1}, \dots, \underline{1}_{a_u}\right) \tag{3.7.8}$$

we have

$$X_o^+ = D\left(\frac{1}{a_1}\underline{1}_{a_1}, \dots, \frac{1}{a_u}\underline{1}_{a_u}\right) \tag{3.7.9}$$

as well as

$$T = D\left(\frac{1}{a_1}J_{a_1}, \dots, \frac{1}{a_u}J_{a_u}\right) = \sum_{j=1}^{u} Q_j \tag{3.7.10}$$

so

$$z = u \quad .$$

Moreover

$$M_i = \sum_{j=1}^{i}\left(Q_j + Q_{j+u}\right) + \sum_{j=i+1}^{u} a_j Q_j \quad , \ i = 1,\dots,u \tag{3.7.11}$$

so

$$B = \begin{bmatrix} B_1 & B_2 \end{bmatrix} \tag{3.7.12}$$

with

$$B_1 = \begin{bmatrix} 1 & a_2 & \cdots & a_u \\ 1 & 1 & \cdots & a_u \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad ; \quad B_2 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{3.7.13}$$

We now have

Proposition 3.13. *Step nesting originates EO that are SEO but not MEO.*

Proof. **S**ince matrices $T, M_1, \dots, M_u \in \mathcal{A} = \overset{u}{\underset{l=1}{\mathrm{X}}} \mathcal{A}(a_l)$ they commute and so, according to Proposition 3.3, step nesting originates COBS thus EO. The rest of the proof follows directly from (3.7.13). From the expression of B_2 we get

$$\begin{cases} \gamma_{u+1} = \sigma_1^2 + \dots + \sigma_u^2 \\ \quad \vdots \\ \gamma_{2u} = \qquad\qquad \sigma_u^2 \end{cases}$$

so we will have

$$\begin{cases} \sigma_u^2 = \gamma_{2u} \\ \sigma_{u-1}^2 = \gamma_{2u-1} - \gamma_{2u} \\ \vdots \\ \sigma_1^2 = \gamma_{u+1} - \gamma_{u+2} \end{cases} .$$

□

In figure 3.5 we have a level for steps that does not appear in figure 3.6. This occurs since we can consider, in step nesting, an initial fixed effects factor, with u levels, while in balanced nesting we have only u random effects factor. This is an additional advantage of step nesting which is even more relevant since LSE of estimable vectors will be UBLUE since step nesting originates EO.

We point out that we have

$$A_j = \lfloor A_{j,1}, \dots, A_{j,n} \rfloor \quad , \quad j = 1,\dots,u \tag{3.7.14}$$

with

$$\begin{cases} A_{j,l} = \underline{0}_{a_l}^T & ; \quad l \neq j \quad ; \quad j = 1, \dots, u \\ A_{j,j} = \dfrac{1}{a_j} 1_{a_j}^T & ; \quad j = 1, \dots, u \end{cases} , \tag{3.7.15}$$

since

$$\begin{cases} Q_j = A_j^T A_j & ; \quad j = 1,\dots,u \\ I_1 = A_j A_j^T & ; \quad j = 1,\dots,u \end{cases} , \tag{3.7.16}$$

and $\text{rank}(Q_j) = 1 \quad , \quad j = 1,\dots,u$.

If $\underline{\beta}_0$ has components $\beta_{0,j}$, $j = 1,\dots,u$, the canonical estimable vectors will be

$$\eta_j = \beta_{0,j} \quad , \ j = 1,\dots,u \tag{3.7.17}$$

and their LSE will be the means of the observations for the u step.

Matrices C_j will, for the models, be reduced to vectors $\underline{c}_j$, $j = 1,\dots,u$, and the expression of estimable vectors as generalized linear combinations will be

$$\underline{\Psi} = \sum_{j=1}^{u} \underline{c}_j \underline{\eta}_j = \sum_{j=1}^{u} \underline{c}_j \beta_{j,0} \quad . \tag{3.7.18}$$

The corresponding expression for LSE will be

$$\underline{\tilde{\Psi}} = \sum_{j=1}^{u} \underline{c}_j \underline{\tilde{\eta}}_j = \sum_{j=1}^{u} \underline{c}_j \tilde{\beta}_{j,0} = \sum_{j=1}^{u} \underline{c}_j y_{j,\cdot} \tag{3.7.19}$$

with $y_{j,\cdot}$ the mean of the observations for the j-th step.

We now will try to generate step nesting. We start by replacing the $A(a_l)$, $l = 1,\ldots,u$, by complete and regular CJAS , $A(l)$, $l = 1,\ldots,u$, with principal basis

$$\underline{Q}(l) = \{Q_1(l), \ldots, Q_{m(l)}(l)\} \ , \ l = 1,\ldots,u .$$

There will be $m = \sum_{l=1}^{u} m(l)$ matrices in

$$\underline{Q} = \text{pb}(A)$$

with

$$A = \underset{l=1}{\overset{u}{\mathsf{X}}} A(l) .$$

These matrices will correspond to the vectors of

$$\Gamma_0 = \bigcup_{l=1}^{u} \{ v\underline{\delta}_l \} \tag{3.7.20}$$

where $\underline{\delta}_l$ has u components, all null except the l-th which will be equal to 1 , $l = 1,\ldots,u$.

Then, with

$$Q_0(l) = 0_{a_l \times a_l} \ , \ l = 1,\ldots,L \tag{3.7.21}$$

and

$$Q(\underline{v}) = D\left(Q_{v_1}(1), \ldots, Q_{v_u}(u)\right) \ , \tag{3.7.22}$$

the matrices in $\underline{Q}$ will be the $Q(\underline{v})$ with $\underline{v} \in \Gamma_0$. Since the orthogonal projection matrices in a CJAS are sums of matrices on its principal basis, the orthogonal projection matrices in A will be the $Q(\underline{v})$ with

$$\underline{v} \in \Gamma = \{ \underline{v} : v_l = 0, \ldots, m(l), l = 1,\ldots,u \} \ , \tag{3.7.23}$$

if we include the null matrix $Q(\underline{0})$ in the family of orthogonal projection matrices.

Given the EO

$$\underline{Y}(l) = \sum_{i=0}^{w(l)} X_i(l) \underline{\beta}_i(l) \ , \ l = 1,\ldots,u \ , \tag{3.7.24}$$

we can combine them into

$$\underline{Y} = \sum_{i=0}^{w} X_i \, \underline{\beta}_i \quad , \tag{3.7.25}$$

where $\underline{\beta}_0$ is fixed and the $\underline{\beta}_1 , \dots , \underline{\beta}_u$ are independent, with null mean vectors and variance-covariance matrices $\sigma_i^2 \; I_{c_i} , \dots , \sigma_w^2 \; I_{c_w}$. Moreover we will assume that

$$\begin{cases} X_0 = D \left(\underline{1}_{a_1} , \dots , \underline{1}_{a_u} \right) \\ X_i = D \left(X_{v_1(i)}(1), \dots , X_{v_u(i)}(u) \right) \quad ; \quad i = 1 , \dots , w \end{cases} . \tag{3.7.26}$$

Then we will have

$$X_0^+ = D \left(\frac{1}{a_1} \underline{1}_{a_1}^T \; \dots \; \frac{1}{a_u} \underline{1}_{a_u}^T \right) \tag{3.7.27}$$

as well as

$$T = X_0 X_0^+ = D \left(\frac{1}{a_1} J_{a_1}^T \; \dots \; \frac{1}{a_u} J_{a_u}^T \right) = \sum_{l=1}^{u} Q \left(\underline{\delta}_l \right) \in \mathcal{A} \tag{3.7.28}$$

and as

$$\begin{aligned} M_i &= D\left(M_{v_1(i)}(1), \dots , M_{v_u(i)}(u)\right) = \sum_{l=1}^{u} \sum_{j=1}^{m(l)} b_{v_l(i),j}(l) Q_j(l) \\ &= \sum_{l=1}^{u} \sum_{j=1}^{m(l)} Q\left(b_{v_l(i),j}(l)\, \underline{\delta}_l \right) \in \mathcal{A} \quad , \quad i = 1 , \dots , w \quad . \end{aligned} \tag{3.7.29}$$

Thus T and the $M_1 , \dots , M_w$ commute so, according to Proposition 3.3, these models are COBS thus EO.

For the new model the i-th row vector of the transition matrix will be

$$\underline{b}_i = \begin{bmatrix} \underline{b}_{v_i(1)}(1) \\ \vdots \\ \underline{b}_{v_i(u)}(u) \end{bmatrix} , \; i = 1 , \dots , u , \tag{3.7.30}$$

with $\underline{b}_v(l)$ the v-th row vector of $B(l)$, $l = 1, \dots , u$. If we want to have the partition $B = [\, B_1 \quad B_2 \,]$, the i-th row vector of B will be constituted by the first components of the $\underline{b}_{v_i(1)}(1) \dots \underline{b}_{v_i(u)}(u)$, the remaining components of these vectors will constitute the i-th row vector of B_2.

3.8. L Extensions

We now consider a possibility of extending our results to a wider class of models. Following Moreira et al (2009), we have

Definition 3.8. The model

$$\underline{Y} = L\underline{Y}^0 + \underline{\varepsilon} \ ,$$

is an L extension of the EO

$$\underline{Y}^0 = \sum_{i=0}^{\omega} X_i \underline{\beta}_i \ ,$$

whose observation vector, $\underline{Y}^0$, has n^0 components and whose principal basis of the corresponding CJAS $\mathcal{A}^0$, is $pb(\mathcal{A}^0) = \{Q_1, \ldots, Q_w\}$, when L is a matrix with linearly independent column vectors and $\underline{\varepsilon}$ is an error vector, with null mean vector and variance-covariance matrix $\sigma^2 I_n$, independent from $\underline{Y}^0$.

The introduction of matrix L can be seen as a generalization of the approach of Khuri & Ghosh (1990) to unbalanced models that are unbalanced only with respect to the last stage.

If we have n^0 treatments and $r_1, \ldots, r_{n^0}$ observations per treatment, we may take a block diagonal matrix with the principal blocks $\underline{1}_{r_1}, \ldots, \underline{1}_{r_{n^0}}$,

$$L = D\left(\underline{1}_{r_1}, \ \ldots \ , \underline{1}_{r_{n^0}} \right) \ . \tag{3.8.1}$$

The OPM on $\overline{\Omega} = R(L)$ will be

$$\overline{T} = LL^+ \tag{3.8.2}$$

and, since the column vectors of L are linearly independent,

$$L^+ L = I_{n^0}$$

so

$$\underline{Y}^{00} = L^+ \underline{Y} = \underline{Y}^0 + \underline{\varepsilon}^0 \ , \tag{3.8.3}$$

with $\underline{\varepsilon}^0 = L^+ \underline{\varepsilon}$.

The mean vector of $\underline{\varepsilon}^0$ will be null and it's variance-covariance matrix will be $\sigma^2 L^+ L^{+T}$. This vector is independent from $\underline{Y}^0$, thus $\underline{Y}^{00}$ will have mean vector

$$\underline{\mu}^{00} = X_0 \, \underline{\beta}_0 \tag{3.8.4}$$

and variance-covariance matrix

$$V_{\underline{Y}^{00}} = \sum_{i=1}^{w} \sigma_i^2 M_i + \sigma^2 L^+ {L^+}^T = \sum_{j=1}^{m} \gamma_j Q_j + \sigma^2 L^+ {L^+}^T . \qquad (3.8.5)$$

Besides this $\underline{Y}^{\perp} = \left(I_n - \overline{T}\right)\underline{Y}$ will have null mean vector and variance-covariance matrix $\sigma^2\left(I_n - \overline{T}\right)$. Moreover the cross-covariance matrix for $\underline{Y}^{00}$ and $\underline{Y}^{\perp}$ will be

$$V_{\underline{Y}^{00},\underline{Y}^{\perp}} = \sigma^2 L^+\left(I_n - \overline{T}\right) = \sigma^2\left(L^+ - L^+\right) = 0$$

since $L^+ T = L^+ L L^+ = L^+$. Thus we have the unbiased estimator

$$\tilde{\sigma}^2 = \frac{S}{n - n^0} \qquad (3.8.6)$$

where

$$S = \left\| \underline{Y}^{\perp} \right\|^2 .$$

Let the row vectors of A_j constitute an orthogonal basis for $R\left(Q_j\right)$, $j = 1,\dots,m$, then

$$\begin{cases} Q_j = A_j^T A_j & , \quad j = 1,\dots,m \\ I_{g_j} = A_j A_j^T & , \quad j = 1,\dots,m \end{cases} \qquad (3.8.7)$$

with

$$g_j = \text{rank}\left(Q_j\right) = \text{rank}\left(A_j\right) \quad , \quad j = 1,\dots,m .$$

The random vector

$$\tilde{\underline{\eta}}_j = A_j \underline{Y}^{00} = A_j \underline{Y}^0 + A_j \underline{\varepsilon}^0 \quad , \quad j = 1,\dots,m \qquad (3.8.8)$$

will have mean vector

$$\underline{\eta}_j = A_j \underline{\mu}^{00} \quad , \quad j = 1,\dots,m \ , \qquad (3.8.9)$$

with $\underline{\eta}_j = \underline{0}$, $j = z+1,\dots,m$, whenever the OPM on the space spanned by $\underline{\mu}^{00}$ will be

$$T = \sum_{j=1}^{z} Q_j . \qquad (3.8.10)$$

Since $\underline{Y}^0$ and $\underline{\varepsilon}^0$ are independent, the $\tilde{\underline{\eta}}_j$, $j = 1,\dots,m$, will have variance-covariance matrices

$$V_{\underline{\tilde{\eta}}_j} = \gamma_j I_{g_j} + \sigma^2 A_j L^+ L^{+T} A_j^T \quad , \ j = 1,\ldots,m \quad , \tag{3.8.11}$$

then taking

$$S_j = \left\| \underline{\tilde{\eta}}_j \right\|^2 \qquad j = 1,\ldots,m \tag{3.8.12}$$

we have the mean value

$$E(S_j) = g_j \gamma_j + t_j \sigma^2 , \ j = z+1,\ldots,m , \tag{3.8.13}$$

with $t_j = tr\left(A_j L^+ L^{+T} A_j^T \right)$, $j = z+1,\ldots,m$. Then we will have the unbiased estimators

$$\underline{\tilde{\gamma}}_j = \frac{S_j}{g_j} - \frac{t_j}{g_j} \tilde{\sigma}^2 \quad , \ j = z+1,\ldots,m \ , \tag{3.8.14}$$

which will be the components of $\underline{\tilde{\gamma}}$.

If there is segregation or matching for $\underline{Y}^0$ we can use our previous results on estimation of variance components.

Besides this, the $\underline{\tilde{\eta}}_j$, $j = 1,\ldots,z$, will be unbiased estimators of the $\underline{\eta}_j$, $j = 1,\ldots,z$, thus

$$\underline{\tilde{\Psi}} = \sum_{j=1}^{z} C_j \underline{\tilde{\eta}}_j \tag{3.8.15}$$

will be an unbiased estimator of

$$\underline{\Psi} = \sum_{j=1}^{z} C_j \underline{\eta}_j . \tag{3.8.16}$$

When the OPM on $R(LX_0)$ does not commute with $V_{\underline{Y}}$, the $\underline{\tilde{\Psi}}$ may not be BLUE for the $\underline{\Psi}$. But when the column vectors of L are pairwise orthogonal with norm 1, we have

$$L^+ = L^T$$

as well as

$$\begin{cases} L^T L = I_n \\ L L^T = \overline{T} \end{cases}$$

and the OPM on $R(LX_0)$ will be

$$LX_0 \left(X_0^T L^T L X_0 \right)^+ X_0^T L^T = LX_0 \left(X_0^T X_0 \right)^+ X_0^T L^T = LTL^T \tag{3.8.17}$$

which commutes with

$$V_{\underline{Y}} = L\left(\sum_{j=1}^{m} \gamma_j Q_j\right)L^T + \sigma^2 I_n \qquad (3.8.18)$$

and the $\underline{\tilde{\Psi}}$ will be BLUE for the $\underline{\Psi}$. These will be orthogonal L extensions.

4. Normal Models

We now will focus on a particular case of mixed models, where the random effects parameters are normal. The assumption of normality turns out to be of great importance since it allows our previous treatments to lead directly to sufficient statistics, see Nunes et al (2008). As for completeness a very specific problem arises when we consider mixed models since linear restrictions on the $\underline{\tilde{\eta}}_1,\dots,\underline{\tilde{\eta}}_z$ or the canonical variance components may arise and we will only have sufficient but not complete statistics.

Based on the normality of the observation vectors we include some results on inference.

A particular case of L Extensions in which the column vectors of matrix L are pairwise orthogonal with norm 1 will be studied.

4.1. Densities and statistics

Definition 4.1. The mixed model

$$\underline{Y}=\sum_{i=0}^{\omega} X_i\underline{\beta}_i$$

is called a normal mixed model if the random effects parameters, $\underline{\beta}_1,\dots,\underline{\beta}_w$, are normally distributed.

The mean vector of $\underline{Y}$ will be

$$\underline{\mu}=X_0\underline{\beta}_0 \quad (4.1.1)$$

and , if the model is EO, it will have the regular variance-covariance matrix

$$V=\sum_{j=1}^{m}\gamma_j Q_j \; . \quad (4.1.2)$$

So that

$$V^{-1}=\sum_{j=1}^{m}\gamma_j^{-1} Q_j \quad (4.1.3)$$

and, since $\gamma_1,\dots,\gamma_m$ are the eigenvalues of V with multiplicities

$$g_j=\text{rank}\left(Q_j\right),\; j=1,\dots,m$$

we also have

$$\det(V) = \prod_{j=1}^{m} \gamma_j^{g_j} \ . \tag{4.1.4}$$

Moreover

$$Q_j = A_j^T A_j \quad , \ j = 1,\dots,m \tag{4.1.5}$$

and so

$$\left(\underline{Y}-\underline{\mu}\right)^T V^{-1}\left(\underline{Y}-\underline{\mu}\right) = \sum_{j=1}^{m} \gamma_j^{-1}\left(\underline{Y}-\underline{\mu}\right)^T A_j^T A_j\left(\underline{Y}-\underline{\mu}\right) = \sum_{j=1}^{m} \frac{\left\| \underline{\tilde{\eta}}_j - \underline{\eta}_j \right\|^2}{\gamma_j} \tag{4.1.6}$$

where , as before, $\underline{\eta}_j = A_j \underline{\mu}$ and $\underline{\tilde{\eta}}_j = A_j \underline{Y}$, $j = 1,\dots,m$. Now $\underline{\eta}_j = \underline{0}$, $j = z+1,\dots,m$, thus, with

$$S_j = \left\| \underline{\tilde{\eta}}_j \right\|^2 \quad , \ j = 1,\dots,m \quad , \tag{4.1.7}$$

we have

$$\left(\underline{Y}-\underline{\mu}\right)^T V^{-1}\left(\underline{Y}-\underline{\mu}\right) = \sum_{j=1}^{z} \frac{\left\| \underline{\tilde{\eta}}_j - \underline{\eta}_j \right\|^2}{\gamma_j} + \sum_{j=z+1}^{m} \frac{S_j}{\gamma_j} \quad . \tag{4.1.8}$$

Thus $\underline{Y}$ will have the density, according to Definition 2.28,

$$n(\underline{Y}) = \frac{e^{-\frac{1}{2}\sum_{j=1}^{z}\frac{\left\| \underline{\tilde{\eta}}_j - \underline{\eta}_j \right\|^2}{\gamma_j} - \frac{1}{2}\sum_{j=z+1}^{m}\frac{S_j}{\gamma_j}}}{(2\pi)^{n/2}\prod_{j=1}^{m}\gamma_j^{g_j/2}} \ ,$$

which belongs to the exponential family and, according to the factorization theorem, has the sufficient statistics $\underline{\tilde{\eta}}_1,\dots,\underline{\tilde{\eta}}_z, S_{z+1},\dots,S_m$.

Putting

$$\begin{cases} U(1) = \left[A_1^T \ \dots \ A_z^T\right]^T \\ U(2) = \left[A_{z+1}^T \ \dots \ A_m^T\right]^T \end{cases} \tag{4.1.9}$$

and taking

$$\begin{cases} \underline{\mu}(1) = U(1)\underline{\mu} \\ \underline{\mu}(2) = U(2)\underline{\mu} = \underline{0} \end{cases} \tag{4.1.10}$$

and

$$\begin{cases} \underline{Z}(1) = U(1)\underline{Y} \\ \underline{Z}(2) = U(2)\underline{Y} \end{cases} \qquad (4.1.11)$$

we have the pair $(\underline{Z}(1), \underline{Z}(2))$ of vectors with joint normal distribution, with mean vectors $\underline{\mu}(1)$ and $\underline{\mu}(2)$, variance-covariance matrices

$$\begin{cases} V(1) = U(1)\,V\,U(1)^T = D\left(\gamma_1 I_{g_1}, \ldots, \gamma_z I_{g_z}\right) \\ V(2) = U(2)\,V\,U(2)^T = D\left(\gamma_{z+1} I_{g_{z+1}}, \ldots, \gamma_m I_{g_m}\right) \end{cases} \qquad (4.1.12)$$

and null cross covariance matrices, so $\underline{Z}(1)$ and $\underline{Z}(2)$ are independent.

Putting

$$\begin{cases} g(1) = \sum_{j=1}^{z} g_j \\ g(2) = \sum_{j=z+1}^{m} g_j \end{cases}, \qquad (4.1.13)$$

these random vectors will have the densities

$$\begin{cases} n_1(\underline{Y}) = \dfrac{e^{-\frac{1}{2}\sum_{j=1}^{z}\frac{\left\|\tilde{\underline{\eta}}_j - \underline{\eta}_j\right\|^2}{\gamma_j}}}{(2\pi)^{g(1)/2}\prod_{j=1}^{z}\gamma_j^{g_j/2}} \\ n_2(\underline{Y}) = \dfrac{e^{-\frac{1}{2}\sum_{j=z+1}^{m}\frac{S_j}{\gamma_j}}}{(2\pi)^{g(2)/2}\prod_{j=z+1}^{m}\gamma_j^{g_j/2}} \end{cases}. \qquad (4.1.14)$$

The first of these has sufficient statistics $\tilde{\underline{\eta}}_j$, $j = 1,\ldots,z$, while the sufficient statistics for the second one will be $S_{z+1},\ldots,S_m$.

As we saw, the LSE for estimable vectors in these models are BLUE. We now look for optimal properties for estimators of variance components.

We have $\frac{1}{\gamma_j}\underline{\eta}_j \in R\left(A_j X_0\right)$, $j = 1,\ldots,m$, so there will be no linear restrictions on the g_j components of $\underline{\eta}_j$ if and only if

$$\operatorname{rank}\left(A_j X_0\right) = g_j, \quad j = 1,\ldots,z. \qquad (4.1.15)$$

Thus we have

Proposition 4.1. *The* $\underline{\tilde{\eta}}_1,\dots,\underline{\tilde{\eta}}_z$ *are sufficient and complete if and only if*

$$\mathrm{rank}\left(A_j X_0\right) = g_j,\quad j = 1,\dots,z.$$

Proof. Density $n_1\left(\underline{Y}\right)$ belongs to the exponential family and has parameters $\frac{1}{\gamma_j}\underline{\eta}_j$, $j = 1,\dots,m$. Using the factorization criterion it is straightforward to see that the $\underline{\tilde{\eta}}_1,\dots,\underline{\tilde{\eta}}_z$ are sufficient. Now, see Lukacs & Laha (1964, pgs 37 to 42), these statistics will be sufficient when there is no linear restrictions on the components of the $\frac{1}{\gamma_j}\underline{\eta}_j$, $j = 1,\dots,z$.

□

Using Proposition 4.1. and the Blackwell-Lehman-Scheffé theorem we see that when (4.1.15) holds the $\underline{\tilde{\eta}}_j$, $j = 1,\dots,z$, are UMVUE for the $\underline{\eta}_j$, $j = 1,\dots,z$, and the $\underline{\tilde{\Psi}} = \sum_{j=1}^{z} U_j \underline{\tilde{\eta}}_j$ for the $\underline{\Psi} = \sum_{j=1}^{z} U_j \underline{\eta}_j$.

Density $n_2\left(\cdot\right)$ belongs to the exponential family and, see Lukacs & Laha (1964, pgs 37 to 42), statistics $S_{z+1},\dots,S_m$ will be sufficient and complete whenever the parameters space of $n_2\left(\cdot\right)$ contains the Cartesian product of $m - z$ non degenerate intervals, since we can assume the $\gamma_{z+1},\dots,\gamma_m$ to be linearly independent. We now have the

Proposition 4.2. *The statistics* $S_{z+1},\dots,S_m$ *are sufficient and complete if and only if matrix* B_2 *is invertible.*

Proof. From (3.3.1), $\underline{\gamma}(2) \in R\left(B_2^T\right)$ for the parameter space of $n_2\left(\cdot\right)$ containing the required Cartesian product of non degenerated intervals we must have

$$\mathrm{rank}\left(B_2^T\right) = m - z.$$

Since B_2^T is a $w \times (m - z)$ matrix it is invertible if and only if

$$w = m - z = \mathrm{rank}\left(B_2^T\right).$$

□

Corollary 4.1. *When* B_2 *is invertible we have segregation and the* $\underline{\tilde{\gamma}}_2$, $\underline{\tilde{\sigma}}^2$ *and* $\underline{\tilde{\gamma}}_1$ *will be* UMVUE($\underline{Z}(2)$)*, this is, they will be UMVUE in the family of estimators derived from* $\underline{Z}(2)$.

Proof. If B_2 is invertible its row vectors will be linearly independent and we will have segregation. The rest of the proof follows from the Blackwell-Lehman-Scheffé theorem.

□

4.2. Inference

Starting with the directly estimable canonic variance components, since $\underline{\tilde{\eta}}_j$, $j = z+1,\dots,m$, are normal with null mean vectors and variance-covariance matrices $\gamma_j I_{g_j}$, $j = z+1,\dots,m$, and, by construction, $\frac{S_j}{\gamma_j}$ has a central chi-square distribution with g_j degrees of freedom , S_j , $j = z+1,\dots,m$ will be the products by the γ_j , $j = z+1,\dots,m$ of central chi-squares with g_j degrees of freedom, $S_j \sim \gamma_j \chi^2_{g_j}$, $j = z+1,\dots,m$.

We then have for γ_j , $j = z+1,\dots,m$, the $1-q$ level confidence intervals

$$\left[\frac{S_j}{x_{g_j,1-q/2}} ; \frac{S_j}{x_{g_j,q/2}} \right] \quad , \quad \left[0 ; \frac{S_j}{x_{g_j,q}} \right] \quad \text{and} \quad \left[\frac{S_j}{x_{g_j,1-q}} ; +\infty \right[,$$

with $x_{g,p}$ the p-th quantile for χ^2_g. We can use these intervals to, through duality, test the hypothesis

$$H_{0,j} : \gamma_j = \gamma_{0,j} \ , \ j = z+1,\dots,m \ . \tag{4.2.1}$$

Then the q level bilateral [right unilateral; left unilateral] test reject $H_{0,j}$ if $\gamma_{0,j}$ does not belong to the first [second ; third] $1-q$ level confidence interval, $j = z+1,\dots,m$.

Now the $\underline{\tilde{\eta}}_j$, $j = 1,\dots,m$, have joint normal distribution and null cross covariance matrices so they will be independent. Then the S_j , $j = 1,\dots,m$, will be independent, and

$$\mathcal{F}_{j,j'} = \frac{g_{j'}}{g_j} \frac{S_j}{S_{j'}} \ ; \ \ j,j' = z+1,\dots,m \tag{4.2.2}$$

will be the product by

$$\rho_{j,j'} = \frac{\gamma_j}{\gamma_{j'}} \ , \ \ j,j' = z+1,\dots,m \tag{4.2.3}$$

of random variables with central F distribution, $F(\ \mid g_j, g_{j'})$, with g_j and $g_{j'}$ degrees of freedom, $j, j' = z+1,\dots,m$. Thus, with $f_{r,s,p}$ the p-th quantile of $F(\ \mid r, s)$ we get, for $\rho_{j,j'}$, $j, j' = z+1,\dots,m$, the $1-q$ level confidence intervals

$$\left[\frac{F_{j,j'}}{f_{g_j,g_{j'},1-q/2}};\frac{F_{j,j'}}{f_{g_j,g_{j'},q/2}}\right] \quad , \quad \left[0;\frac{F_{j,j'}}{f_{g_j,g_{j'},q}}\right] \quad \text{and} \quad \left[\frac{F_{j,j'}}{f_{g_j,g_{j'},1-q}};+\infty\right[\ .$$

As before, we can use these confidence intervals to test, through duality,

$$H_{0,j,j'}: \rho_{j,j'} = \rho_{0,j,j'} \ , \ j, j' = z+1,\dots,m \ .$$

Moreover, if $g_{j'} > 2$, the mean value of $F(\ \mid g_j, g_{j'})$ will be $\frac{g_{j'}}{g_{j'}-2}$ so the mean value of $F_{j,j'}$ will be $\frac{g_{j'}}{g_{j'}-2}\rho_{j,j'}$ and we have the unbiased estimators

$$\tilde{\rho}_{j,j'} = \frac{g_{j'}-2}{g_{j'}} F_{j,j'} = \frac{g_{j'}-2}{g_j}\frac{S_j}{S_{j'}} \ , \ j, j' = z+1,\dots,m \ . \qquad (4.2.4)$$

If the model is MEO these results apply to all canonical variance components.

Whenever $j < z < j'$, if $\gamma_j = \gamma_{j'}$

$$\underline{\tilde{\Psi}} = G_j \underline{\tilde{\eta}}_j \qquad (4.2.5)$$

will be normal, with mean vector

$$\underline{\Psi} = G_j \underline{\eta}_j \qquad (4.2.6)$$

and variance covariance matrix $\gamma_j G_j G_j^T$, independent from $S_{j'}$. Thus

$$\left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_j\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_j\right) \sim \gamma_j \ \chi^2_{r_j} \ , \qquad (4.2.7)$$

with

$$r_j = \text{rank}\left(G_j G_j^T\right) = \text{rank}\left(G_j\right),$$

see Mexia (1990), are independent of $S_j \sim \gamma_j \chi^2_{g_j}$, so

$$F'_{j,j'} = \frac{g_{j'}}{r_j} \frac{\left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_j\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_j\right)}{S_{j'}} \ , \ j,j' = z+1,\dots,m, \qquad (4.2.8)$$

will have distribution $F(\cdot \mid r_j, g_{j'})$. We then get, for $\underline{\Psi}_j$, the $1-q$ level confidence ellipsoid given by

$$\left(\underline{\Psi}_j - \underline{\tilde{\Psi}}_j\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\Psi}_j - \underline{\tilde{\Psi}}_j\right) \leq r_j f_{r_j, g_{j'}, 1-q} \frac{S_{j'}}{g_{j'}} .$$

We may use this ellipsoid to test, through duality, the hypothesis

$$H'_{0,j} : \underline{\Psi}_j = \underline{\Psi}_{0,j} . \tag{4.2.9}$$

The q level test rejecting this hypothesis when $\underline{\Psi}_{0,j}$ does not belongs to the $1-q$ level confidence ellipsoid.

Besides this, an extension of the Scheffé theorem gives us, see Scheffé (1959) and Mexia (1989), the simultaneous confidence interval

$$\bigwedge_{\underline{d}} \left(\left| \underline{d}^T \underline{\Psi}_j - \underline{d}^T \underline{\tilde{\Psi}}_j \right| \leq \sqrt{r_j f_{r_j, g_{j'}, 1-q} \, \underline{d}^T \left(G_j G_j^T\right) \underline{d}} \right) , \tag{4.2.10}$$

for the $\underline{d}^T \underline{\Psi}_j$. In this expression, $\bigwedge_{\underline{d}}$ indicates that all possible eligible vectors $\underline{d}$ are used. The joint confidence level for all these intervals is $1-q$.

Moreover the statistics

$$F_{0,j,j'} = \frac{g_{j'}}{r_j} \frac{\left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_{0,j}\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_{0,j}\right)}{S_{j'}} \tag{4.2.11}$$

will have the F distribution with r_j and $g_{j'}$ degrees of freedom and non-centrality parameter

$$\delta_j = \frac{1}{\gamma_j} \left(\underline{\Psi}_j - \underline{\Psi}_{0,j}\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\Psi}_j - \underline{\Psi}_{0,j}\right) , \tag{4.2.12}$$

$F(\,. \mid r_j, g_{j'}, \delta_j\,)$. So we may use the statistic defined in (4.2.11) to test

$$H_{0,j} : \underline{\Psi}_j = \underline{\Psi}_{0,j} . \tag{4.2.13}$$

When $H'_{0,j}$ hold the statistic $F_{0,j,j'}$ can be rewritten as $F_{j,j'}$, which has distribution $F(\cdot \mid r_j, g_{j'})$. This is easy to see that this test enjoys duality since, when it has level q, $H_{0,j}$ is rejected when and only when the $1-q$ level confidence ellipsoid does not contain $\underline{\Psi}_{0,j}$.

If the row vectors of G_j are linearly independent

$$\left(G_j G_j^T\right)^+ = \left(G_j G_j^T\right)^{-1} \tag{4.2.14}$$

and $H_{0,j}$ may be written as

$$H_{0,j} : \delta_j = 0 \tag{4.2.15}$$

and, see Mexia (1995) and Nunes (2005), this test is strictly unbiased.

When the model is MEO these results apply to all pairs $(j, j+z)$ $j = 1, \dots, z$.

4.3. Orthogonal L Extensions

We now consider orthogonal L extensions of normal EO.
Thus we will have

$$L^{+} = L^{T} \tag{4.3.1}$$

and so

$$L^{+} L = I_n . \tag{4.3.2}$$

So $\underline{Y}$ will have mean vector

$$\underline{\mu} = L\, X_0\, \underline{\beta}_0 \ , \tag{4.3.3}$$

and variance -covariance matrix

$$V = \sum_{j=1}^{m} \gamma_j\, L\, Q_j\, L^T + \sigma^2 I_n . \tag{4.3.4}$$

Matrices

$$\overline{Q}_j = L\, Q_j\, L^T , \ j = 1, \dots, m \tag{4.3.5}$$

will be POOPM, since they are symmetrical and idempotent and $\overline{Q}_j \overline{Q}_{j'} = 0_{n \times n}$, $j \neq j'$.

As before, we take

$$Q_j = A_j^T A_j \ , \ j = 1, \dots, m \ , \tag{4.3.6}$$

moreover

$$\sum_{j=1}^{m} \overline{Q}_j = L\, L^T = \overline{T} , \tag{4.3.7}$$

where $\overline{T}$ is the OPM on $R(L)$ and, with

$$\overline{T}^C = I_n - \overline{T} , \tag{4.3.8}$$

we will have

$$\overline{T}^{C}\,\overline{Q}_j = \overline{Q}_j\,\overline{T}^{C} = 0_{n\times n}\,,\ \ j=1,\ldots,m \tag{4.3.9}$$

thus the $\overline{Q}_1,\ldots,\overline{Q}_m$ and $\overline{T}^{C}$ are POOPM. We now can write

$$V = \sum_{j=1}^{m}\left(\gamma_j+\sigma^2\right)\overline{Q}_j + \sigma^2\,\overline{T}^{C} \tag{4.3.10}$$

so we obtain

$$\begin{cases} V^{-1} = \sum_{j=1}^{m}\left(\gamma_j+\sigma^2\right)^{-1}\overline{Q}_j + \dfrac{1}{\sigma^2}\overline{T}^{C} \\ \det(V) = \prod_{j=1}^{m}\left(\gamma_j+\sigma^2\right)^{g_j/2} + \left(\sigma^2\right)^{n-n^{o}} \end{cases} \tag{4.3.11}$$

with, see once again Silvey (1975),

$$\operatorname{rank}\left(\overline{Q}_j\right) = \operatorname{rank}\left(L\,A_j^{T}\right) = \operatorname{rank}\left(A_j^{T}\right) = \operatorname{rank}\left(Q_j\right),\ \ j=1,\ldots,m\,, \tag{4.3.12}$$

where $\operatorname{rank}\left(L\,A_j^{T}\right)=\operatorname{rank}\left(A_j^{T}\right)$, $j=1,\ldots,m$, since the column vectors of L are linearly independent and so $N\left(L\,A_j^{T}\right)=N\left(A_j^{T}\right)$, $j=1,\ldots,m$, as well as

$$\operatorname{rank}\left(L\,A_j^{T}\right) = n^{o} - \dim\left(N\left(L\,A_j^{T}\right)\right) = n^{o} - \dim\left(N\left(A_j^{T}\right)\right) = \operatorname{rank}\left(A_j^{T}\right)\ ,\ \ j=1,\ldots,m\ . \tag{4.3.13}$$

Since, from (4.3.5) and (4.3.6), we have

$$\overline{Q}_j = L\,A_j^{T}A_j\,L^{T},\ \ j=1,\ldots,m\ , \tag{4.3.14}$$

we will take

$$\begin{cases} \underline{\tilde{\eta}}_j = A_j\,L^{T}\underline{Y} & j=1,\ldots,m \\ \underline{\eta}_j = A_j\,L^{T}\underline{\mu} & j=1,\ldots,m \end{cases}\ , \tag{4.3.15}$$

so

$$\left(\underline{Y}-\underline{\mu}\right)^{T}\overline{Q}_j\left(\underline{Y}-\underline{\mu}\right) = \left(\underline{Y}-\underline{\mu}\right)^{T}L\,A_j^{T}A_j\,L^{T}\left(\underline{Y}-\underline{\mu}\right) = \left\|\underline{\tilde{\eta}}_j-\underline{\eta}_j\right\|^2\ ,\ \ j=1,\ldots,m \tag{4.3.16}$$

as well as

$$\begin{cases} \underline{\tilde{\eta}}_j = A_j\,L^{T}\left[L\underline{Y}^{0}+\underline{\varepsilon}\right] = A_j\,\underline{Y}^{0}+A_j\,\underline{\varepsilon} & j=1,\ldots,m \\ \underline{\eta}_j = A_j\,L^{T}L\,\underline{\mu}^{0} = A_j\,\underline{\mu}^{0} & j=1,\ldots,m \end{cases}\ , \tag{4.3.17}$$

with $\underline{\mu}^0 = E(\underline{Y}^0)$.

Since $S = \left\| \underline{Y}^{\perp} \right\|^2 \sim \sigma^2 \chi^2_{n-n^0}$, reasoning as above we can establish

Proposition 4.3. Statistics $\underline{\tilde{\eta}}_j$, $j = 1,\dots,z$, $S_j = \left\| \underline{\tilde{\eta}}_j \right\|^2$, $j = z+1,\dots,m$ and S are sufficient.

Thus we will prefer using estimators given by functions of these statistics.

Moreover for σ^2 we have the 1-q level confidence intervals

$$\left[\frac{S}{\chi_{n-n^o,\,1-\frac{q}{2}}} ; \frac{S}{\chi_{n-n^o,\,\frac{q}{2}}} \right] \quad , \quad \left[0 \ ; \frac{S}{\chi_{n-n^o,\,q}} \right] \quad \text{and} \quad \left[\frac{S}{\chi_{n-n^o,\,1-q}} ;\ +\infty \right[,$$

and for $\bar{\gamma}_j = \gamma_j + \sigma^2$, $j = 1,\dots,m$ we have the 1-q level confidence intervals

$$\left[\frac{S_j}{\chi_{g_j,\,1-\frac{q}{2}}} ; \frac{S_j}{\chi_{g_j,\,\frac{q}{2}}} \right] \quad , \quad \left[0 \ ; \frac{S_j}{\chi_{g_j,\,q}} \right] \quad \text{and} \quad \left[\frac{S_j}{\chi_{g_j,\,1-q}} ;\ +\infty \right[,$$

$j = 1,\dots,m$, since $S \sim \sigma^2 \chi^2_{n-n^o}$ and $S_j \sim \bar{\gamma}_j\, \chi^2_{g_j}$, $j = 1,\dots,m$. These confidence intervals can be used to test the hypothesis

$$H_0 : \sigma^2 = \sigma_0^2 \tag{4.3.18}$$

and

$$H_{0,j} : \gamma_j = \gamma_{j,0} \quad , \ j = 1,\dots,m \tag{4.3.19}$$

thus getting, through duality, q level two-sided [right one-sided ; left one-sided] tests.

Moreover, taking

$$\bar{\gamma}_{m+1} = \sigma^2 \tag{4.3.20}$$

and

$$\bar{\rho}_{j,j'} = \frac{\bar{\gamma}_j}{\bar{\gamma}_{j'}} \quad j, j' = z+1,\dots,m \tag{4.3.21}$$

the statistics

$$F_{j,j'} = \frac{g_{j'}}{g_j} \frac{S_{j'}}{S_j} \qquad j, j' = z+1,\dots,m+1 \tag{4.3.22}$$

will be the product by $\overline{\rho}_{j,j'}$ of a variable with distribution $F(\cdot \mid g_j, g_{j'})$ where $g_{m+1} = n - n^o$.

Reasoning as in Section 4.2, we get, for the $\overline{\rho}_{j,j'}$, the confidence intervals

$$\left[\frac{F_{j,j'}}{f_{g_j,g_{j'},1-q/2}} ; \frac{F_{j,j'}}{f_{g_j,g_{j'},q/2}} \right] \quad , \quad \left[0 ; \frac{F_{j,j'}}{f_{g_j,g_{j'},q}} \right] \quad \text{and} \quad \left[\frac{F_{j,j'}}{f_{g_j,g_{j'},1-q}} ; +\infty \right[, \quad j, j' = z+1, \dots, m+1 .$$

As before, we can use the confidence intervals to obtain the q level two-sided [right one-sided; left one-sided] tests for the hypothesis

$$H_{0,j,j'} : \overline{\rho}_{j,j'} = \overline{\rho}_{0,j,j'} \qquad j, j' = z+1, \dots, m+1 \ . \tag{4.3.23}$$

We point out that the hypothesis

$$H_{0,j,m+1} : \overline{\rho}_{j,m+1} = 1 \qquad j = z+1, \dots, m \tag{4.3.24}$$

can be rewritten as

$$H_{0,j} : \gamma_j = 0 \qquad j = z+1, \dots, m \ . \tag{4.3.25}$$

Besides this, if $g_{j'} > 2$, $F_{j,j'}$ will have mean value

$$\frac{g_j}{g_{j'} - 2} \overline{\rho}_{j,j'} \qquad j, j' = z+1, \dots, m+1 \tag{4.3.26}$$

so we have the UMVUE

$$\tilde{\overline{\rho}}_{j,j'} = \frac{g_{j'} - 2}{g_j} F_{j,j'} \quad , \ j, j' = z+1, \dots, m+1 . \tag{4.3.27}$$

Let us now assume that, with $j < z < j'$, $\gamma_j = \gamma_{j'}$ so that $\overline{\gamma}_j = \overline{\gamma}_{j'}$. Thus

$$\underline{\tilde{\Psi}}_j = G_j \underline{\tilde{\eta}}_j \tag{4.3.28}$$

will be normal with mean vector $\underline{\Psi}_j = G_j \eta_j$ and variance-covariance matrix $\overline{\gamma}_j G_j G_j^T$, thus, see once again Mexia (1990),

$$\left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_j\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_j\right) \sim \overline{\gamma}_j \, \chi^2_{r_j} \ , \tag{4.3.29}$$

with $r_j = \text{rank}\left(G_j G_j^T\right)^+ = \text{rank}\left(G_j\right)$, independent from $S_{j'} \sim \overline{\gamma}_{j'} \, \chi^2_{g_{j'}}$, so

$$F_{j'} = \frac{g_{j'}}{r_j} \frac{\left(\underline{\Psi}_j - \underline{\tilde{\Psi}}_j\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\Psi}_j - \underline{\tilde{\Psi}}_j\right)}{S_{j'}} \tag{4.3.30}$$

will have distribution $F\left(\cdot \mid r_j, g_{j'}\right)$, so we have the $1-q$ level confidence ellipsoid

$$\left(\underline{\Psi}_j - \underline{\tilde{\Psi}}_j\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\Psi}_j - \underline{\tilde{\Psi}}_j\right) \le r_j f_{r_j, g_{j'}, 1-q} \frac{S_{j'}}{g_{j'}} \qquad (4.3.31)$$

for $\underline{\Psi}_j$, as well as the simultaneous confidence intervals, with joint confidence level $1-q$,

$$\underset{\underline{d}}{\wedge}\left(\left| d^T \underline{\Psi}_j - d^T \underline{\tilde{\Psi}}_j \right| \le \sqrt{r_j f_{r_j, g_{j'}, 1-q} \, d^T \left(G_j G_j^T\right) d \frac{S_{j'}}{g_{j'}}} \right). \qquad (4.3.32)$$

The confidence ellipsoid can be used to test the hypothesis

$$H_{0,j} : \underline{\Psi}_j = \underline{\Psi}_{j,0} \qquad (4.3.33)$$

through duality. The corresponding F test will have statistics

$$\mathcal{F}_j = \frac{g_{j'}}{r_j} \frac{\left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_{0,j}\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\tilde{\Psi}}_j - \underline{\Psi}_{0,j}\right)}{S_{j'}} \qquad (4.3.34)$$

with distribution $F\left(\cdot \mid r_j, g_{j'}, \delta_j\right)$, where

$$\delta_j = \frac{1}{\bar{\gamma}_j}\left(\underline{\Psi}_j - \underline{\Psi}_{0,j}\right)^T \left(G_j G_j^T\right)^+ \left(\underline{\Psi}_j - \underline{\Psi}_{0,j}\right). \qquad (4.3.35)$$

This F test enjoys duality and is unbiased being strictly unbiased when the row vectors of G_j are linearly independent.

5. Final comments and future work

In this thesis we presented the theory of Error-orthogonal models, EO, basing ourselves on their algebraic structure thus on COBS approach.

This enabled

- The estimation of variance components under general conditions;
- The introduction for EO of operations leading to complex models built combining simple ones and, in the case of step nesting, leading to a great economy on the number of observations;
- The study of conditions for having UMVUE for relevant parameters.

We restricted ourselves to isolated models. In continuation of our study, as a natural development, we intend to study structured families of EO. In these families we have a model for each one of the treatments of a base design and study the action of the factors of that design, namely on the fixed effects parameters of the models in the family.

BIBLIOGRAPHY

[1] Calinski, T. & Kageyama S. (2000). *Block Designs: A Randomization Approach. Vol. I: Analysis*. Lecture Note in Statistics, 150, New York :Springer-Verlag.

[2] Calinski, T. & Kageyama S. (2003). *Block Designs: A Randomization Approach. Vol. II: Design*. Lecture Note in Statistics, 170, New York :Springer-Verlag.

[3] Carvalho, F., Mexia, J. T. & Oliveira, M. M.(2008). *Canonic inference and commutative orthogonal block structure*. Discussiones Mathematicae – Prob. & Stat., 28(2), 171-181.

[4] Cox, D.R. & Solomon, P.J. (2003). *Components of Variance*. New York: Chapman & Hall.

[5] Fernandes, C. (2009). *Models with cross step nesting*. PhD Thesis, UBI – Covilhã.

[6] Fernandes, C., Ramos, P., & Mexia, J.T. (2010). *Algebraic structure of step nesting designs*. Discussiones Mathematicae. Prob. & Stat. 30(2), 221-235.

[7] Ferreira, S. S. (2006). *Inferência para modelos ortogonais com segregação*. PhD Thesis, UBI – Covilhã.

[8] Ferreira, S. S. ; Ferreira, D. ; Fernandes, C. & Mexia, J.T. (2007) *Orthogonal Mixed Models and Perfect Families of Symmetric Matrices*. In proceedings of 56th session of the International Statistical Institute. Lisboa.

[9] Ferreira, S.S., Ferreira, D., Moreira, E. & Mexia, J.T. (2009). *Inference for L Orthogonal Models*. Journal of Interdisciplinary Mathematics, 12 (6), 815—824.

[10] Ferreira, S.S., Ferreira, D., Nunes, C. & Mexia, J.T. (2010). *Nesting segregated mixed models*. Journal of Statistical Theory and Practice, 4 (2), 233-242.

[11] Fonseca, M., Mexia, J.T. & Zmyslony, R. (2003). *Estimators and tests for variance components in cross nested orthogonal designs*. Discussiones Mathematicae – Prob & Stat., 23(2), 175-201.

[12] Fonseca, M., Mexia, J.T. & Zmyslony, R. (2006). *Binary Operations on Jordan algebras*. Linear Algebra Appl., 117, 75-86.

[13] Fonseca, M. (2007). *Estimation and hypothesis testing in mixed linear models*. PhD Thesis. FCT – Universidade Nova de Lisboa.

[14] Fonseca, M., Mexia, J.T. , & Zmyslony, R. (2007). *Jordan algebras, generating pivot variables and orthogonal normal models*. Journal of Interdisciplinary Mathematics, 1, 305–327.

[15] Fonseca, M., Mexia, J.T. & Zmyslony, R. (2008). *Inference in normal models with commutative orthogonal block structure*. Acta et Commentationes Universitatis Tartuensis de Mathematica, 12, 3–16.

[16] Fonseca, M., Jesus, V., Mexia, J.T. & Zmyslony R. (2009). *Binary Operations and canonical forms for factorial and related models*. Linear Algebra and its Applications, 430, 2781-2797.

[17] Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. New York: John Wiley & Sons.

[18] Graham, A. (1981). *Kronecker Products and Matrix Calculus: With Applications*. Ellis Horwood Ltd.

[19] Houtman, A.M. & Speed, T.P. (1983). *Balance in designed experiments with orthogonal block structure*. Ann. Statist. 11(4), 1069-1983.

[20] Jacobson, N. (1953). *Lectures in Abstract Algebra*. Volume II-Linear Algebra. New York: D. Van Nostrand.

[21] Jesus, V., Rodrigues, P.C. & Mexia, J.T. (2007). *Inference for random effects in prime basis factorials using commutative Jordan algebras*. Discussiones Mathematicae - Probability and Statistics, 27, 15–25.

[22] Jesus, V., Mexia, J.T. & Oliveira M.M. (2009a). *Binary operations on prime basis factorials*. Biometrical Letters, 46(1), 1–14.

[23] Jesus, V., Fonseca, M., Mexia, J.T. & Zmyslony, R. (2009b). *Binary operations and canonical forms for factorial and related models*. Linear Algebra and its Applications, 43, 2781–2797.

[24] Jordan, P.; Von Neumann, J. & Wigner, E. (1934). *On an algebraic generalization of the quantum mechanical formulation*. Annals of Mathematical. 35 (1).

[25] Khuri, A.I. & Ghosh, M. (1990). *Minimal Sufficient Statistics for the Unbalanced Two-Fold Nested Model*. Statistics and Probability Letters, 10, 351-353.

[26] Khuri, A.I., Mathew, T., & Sinha, B.K. (1998). *Statistical Tests for Mixed Linear Models*. New York :Wiley.

[27] Lehmann, E.L. & Casella, G. (1998). *Theory of point estimation*. 2nd ed. Springer, Texts in Statistics. New York: Springer.

[28] Lukacs, E. & Laha, R.C. (1964). *Applications of characteristic functions* (Griffin's Statistical Monographs & Courses, No. 14). New York: T. Hafner Publishing Co.

[29] Malley, J.D. (1994). *Statistical Applications of Jordan Algebras*, Lecture Notes in Statistics, vol. 91. New York: Springer-Verlag,.

[30] Mejza, S. (1992). *On some aspects of general balance in designed experiments*. Statistica, anno LII, 2, 263-278.

[31] Mexia, J.T. (1989). *Controlled Heteroscedasticity, Quotient Vector. Spaces and F Tests for Hypothesis on Mean Vectors*. Trabalhos de Investigação, Nº1. FCT/UNL.

[32] Mexia, J.T. (1990). *Best linear unbiased estimates, duality of F tests and Sheffé multiple comparison method in presence of controlled heteroscedasticity*. Computational statistics and data analisys, 10(3): 271–281.

[33] Mexia, J.T. (1995). *Introdução à Inferência Estatística Linear*. Centro de Estudos de Matemática Aplicada. Edições Lusofonas.

[34] Michalski, A. & Zmyslony, R. (1996). *Testing hypothesis for variance components in mixed linear models*. Statistics, 27, 297-310.

[35] Michalski, A. & Zmyslony, R. (1999). *Testing hypothesis for linear functions of parameters in mixed linear models*. Tatra Mountain Mathematical Publications ,17, 103-110.

[36] Mood, A.M., Graybill, F.A., Boes D.C. (1974). *Introduction to the Theory of Statistics* . New York: McGraw-Hill.

[37] Moreira, E., Mexia J.T., Fonseca, M. & Zmyślony, R. (2009). *L models and multiple regressions designs*. Statistical Papers, Springer, 50(4), 869-885

[38] Nelder, J.A. (1965a). *The analysis of randomized experiments with orthogonal block structure I. Block structure and the null analysis of variance*. Proceedings of the Royal Society, Series A, 283, 147–162.

[39] Nelder, J.A. (1965b). *The analysis of randomized experiments with orthogonal block structure II. Treatment structure and the general analysis of variance*. Proceedings of the Royal Society, Series A, 283, 163–178.

[40] Nunes, C. (2005). *Testes F e relacionados em modelos mistos com cross-nesting ortogonal*. PhD Thesis, UBI – Covilhã.

[41] Nunes, C., Santos, C. & Mexia, J.T. (2008). *Relevant statistics for models with commutative orthogonal block structure and unbiased estimator for variance components*. Journal of Interdisciplinary Mathematics. 11(4), 553-564.

[42] Robbins, H. (1948). *Mixture of distribution*. The Annals of Mathematical Statistics. Vol. 19, 360-369.

[43] Robbins, H. & Pitman, E.J.G. (1949). *Application of the method of mixtures to quadratic forms in normal variates*. The Annals of Mathematical Statistics. 20, 552-560.

[44] Rodrigues, P.C. & Mexia, J.T. (2006). *ANOVA using commutative Jordan algebras. Discussiones Mathematicae* - Probability and Statistics, 26, 179–191.

[45] Santos, C., Nunes, C. & Mexia, J.T. (2007a). *Modelos com estrutura comutativa ortogonal por blocos e estatísticas relevantes*. Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística, 709-716

[46] Santos, C., Nunes, C. & Mexia, J.T. (2007b). *OBS, COBS and Mixed Models associated to commutative Jordan Algebra*. In proceedings of 56th session of the International Statistical Institute. Lisboa.

[47] Scheffé, H. (1959). *The analysis of variance*. New York : Jonh Wiley & Sons.

[48] Schott, J.R. (1997). *Matrix Analysis for Statistics*. New York: Jonh Wiley & Sons.

[49] Seely, J. (1970a). *Linear spaces and unbiased estimation*. Ann. Math. Stat., (41), 1725-1734.

[50] Seely, J. (1970b). Linear *spaces and unbiased estimation. An aplication to the mixed linear model*. Ann. Math. Stat., 41, 1735-1748.

[51] Seely, J. (1971). *Quadratic subspaces and completeness*. Ann. Math. Stat., 42, 710-721.

[52] Seely, J. (1977). *Minimal sufficient statistics and completeness for multivariate normal families*. Sankhya 39, 170-185.

[53] Seely, J. & Zyskind, G.(1971). *Linear spaces and minimum variance estimators*. The Annals of Mathematical Statistics, 42(2).

[54] Silvey, S.D. (1975). *Statistical Inference*. Monographs on Applied Probability and Statistics. London: Chapman & Hall.

[55] Steeb, W.H. (1991). *Kronecker Product of Matrices and Applications*. Mannheim, Germany: Bibliographisches Institut.

[56] Steeb, W.H. & Hardy, Y. (2011) *Matrix calculus and Kronecker product: A practical approach to linear and multilinear algebra*. (Second Edition) World Scientific Publishing Company.

[57] VanLeeuwen, D.M., Seely, J.F. & Birkes, D.S. (1998). *Sufficient conditions for orthogonal designs in mixed linear models*. Journal of Statistical Planning and Inference, 73, 373-389.

[58] VanLeeuwen, D.M., Birkes, D.S. and Seely, J.F. (1999). *Balance and orthogonality in designs for mixed classification models*. The Annals of Statistics, 27 (6), 1927-1947.

[59] Zmyslony, R. (1978). *A characterization of best linear unbiased estimators in the general linear model*. Lecture Notes in Statistics, 2, 365-373.