ACKNOWLEDGEMENTS

First of all I would like to thank my Supervisors Prof. Dr. João Tiago Mexia and Prof. Dr. Hab. Roman Zmyślony for their friendship, support and trust in me.

To my Friends and Colleagues in the Centre of Mathematics and Applications of the Nova University of Lisbon. Their opinions, suggestions and friendship were crucial for this work.

Last, but certainly not least, I thank my Family and Friends for their indefatigable support. Without them nothing in my life would be possible.

ABSTRACT

This thesis focuses on inference and hypothesis testing for parameters in orthogonal mixed linear models. A canonical form for such models is obtained using the matrices in principal basis of the commutative Jordan algebra associated to the model. UMVUE are derived and used for hypothesis tests. When usual F tests are not possible to use, generalized F tests arise, and the distribution for the statistic of such tests is obtained under some mild conditions. application of these results is made to cross-nested models.

SINOPSE

Esta tese centra-se na inferência e teste de hipóteses para parâmetros em modelos lineares mistos ortogonais. É obtida uma forma canónica para estes modelos usando as matrizes da base principal da álgebra de Jordan comutativa associada ao modelo. São derivados UMVUE que são usados em testes de hipóteses. Quando não é possível usar os testes F usuais, surgem os testes F generalizados, e é obtida a distribuição da estatística destes testes sob algumas condições. Estes resultados são depois aplicados a modelos com cruzamento-encaixe.

NOTATION INDEX

- $\mathbf{X}'$: transpose of matrix $\mathbf{X}$
- $\mathbf{I}_n$: identity matrix of size n
- $\mathbf{0}_{n\times m}$: null matrix of size $n \times m$
- $\mathrm{R}(\mathbf{X})$: linear space spanned by the column vectors of matrix $\mathbf{X}$
- $\mathrm{N}(\mathbf{X})$: kernel of matrix $\mathbf{X}$
- $\perp$: orthogonal
- $\oplus$: direct sum of subspaces
- $\boxplus$: direct sum orthogonal of subspaces
- $\mathbb{P}[\cdot]$: probability
- $\mathbb{E}[\mathbf{X}]$: expectation of the random vector $\mathbf{X}$
- $\mathbb{V}[\mathbf{X}]$: covariance matrix of the random vector $\mathbf{X}$
- $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$: normal random vector with mean vector $\mathbf{x}$ and covariance matrix $\mathbf{V}$
- $\chi^2_{(n)}$: chi square random variable with n degrees of freedom
- $\ell(\mathbf{t}, \boldsymbol{\theta})$: loss function of estimate $\mathbf{t}$ for $\boldsymbol{\theta}$

- $\mathcal{R}_{\boldsymbol{\theta}}(\mathbf{T})$: risk function of estimator $\mathbf{T}$ for $\boldsymbol{\theta}$
- $\beta_{\phi}(\boldsymbol{\theta})$: power function of test ϕ on $\boldsymbol{\theta}$

CONTENTS

LIST OF FIGURES

1. INTRODUCTION

This work is an attempt to solve some unanswered questions in hypothesis tests for linear mixed models. It is a very well known fact that in a three way ANOVA random model the nullity of variance components associated to the main effects can not be tested with an F test. Several alternatives have been suggested (see [7]) like Bartlett-Scheffé tests and the Satterthwaite's approximation, but they either loose information or are too unprecise when the sample is not large enough. So, under some mild conditions, an exact distribution for such hypothesis is derived. A characterization of the algebraic structure of linear models is established using commutative Jordan algebras, like in [22], [25] and [34]. This allows the expression for the model in a canonical form, which allows the immediate derivation of sufficient statistics (with normality) and F tests (also assuming normality).

Chapter 2 covers many topics in linear algebra and matrix theory essential in the analysis and inference for linear models. Special attention is directed to Jordan algebras, introduced in [6], and commutative Jordan algebras. Regarding the latter, very important results obtained in [22] and more where presented in [18], [15] and [11].

Chapter 3 deals with estimation. Firstly, general definitions of estimators and their properties are given as well as the properties of the exponential family of distributions, mainly for completeness sake. Definitions of linear models and their special types are given, like Gauss-Markov models and COBS models, which provide the foundations to necessary and sufficient conditions for the existence of sufficient

and complete sufficient statistics, both for estimable functions of fixed effects and variance components (see [25] and [34]). A canonical form for orthogonal models based on the principal basis of the commutative Jordan algebra is then presented (see [4]). From a model expressed in this form it is easy to derive complete sufficient statistics for the parameters in the model.

Chapter 4 covers hypothesis tests. Like in the previous chapter, for completeness sake, definitions of hypothesis and some general results on optimality of tests are given. Nextly, hypothesis testing in linear models is analyzed, in particular F tests, under normality. Necessary and sufficient conditions for the existence of usual F tests in models with commutative orthogonal block structure are obtained. When it is not possible to test an hypothesis in these models with an usual F test, generalized F tests arise. Under some evenness conditions for the degrees of freedom, the distribution for the quotient of convex combinations of chi-squares divided by their degrees of freedom is given (see [1]).

Finally, in Chapter 5 the results obtained in the previous chapters are applied to cross-nested random models with estimation and hypothesis testing (see [3]). In fact, exact formulas for estimators and test statistics are derived and given. Sufficiency conditions for the evenness conditions to hold are obtained. This characterization is given in terms of the number of levels in the factors.

2. ALGEBRAIC RESULTS

In this chapter, some elementary and not so elementary results on Matrix Theory will be presented. Most of these results are proven in readily available literature. The first part will be a brief discussion of orthogonal projection matrices, followed by a section on generalized inverses. Lastly, a section on Jordan algebras (or quadratic subspaces) will introduce some important results that are used in linear models.

2.1 Projection Matrices

Let $\mathcal{E}$ be a linear space, and $\mathcal{S} \subset \mathcal{E}$ a linear subspace of $\mathcal{E}$. Let also $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ be an orthonormal base for $\mathcal{E}$ and $\{\mathbf{e}_1, ..., \mathbf{e}_r\}$, $r < n$, be an orthonormal base for $\mathcal{S}$. With $\mathbf{x} \in \mathcal{E}$,

$$\mathbf{x} = \sum_{i=1}^{r} \alpha_i \mathbf{e}_i + \sum_{i=r+1}^{n} \alpha_i \mathbf{e}_i.$$

Consider now the matrix

$$\mathbf{E} = [\mathbf{E}_1 \ \mathbf{E}_2], \tag{2.1}$$

with

$$\mathbf{E}_1 = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_r] \text{ and } \mathbf{E}_2 = [\mathbf{e}_{r+1} \ \cdots \ \mathbf{e}_n]. \tag{2.2}$$

It is then clear that, with $\boldsymbol{\alpha} = [\alpha_1 \ \cdots \ \alpha_n]'$, $\boldsymbol{\alpha}_1 = [\alpha_1 \ \cdots \ \alpha_r]'$ and $\boldsymbol{\alpha}_2 = [\alpha_{r+1} \ \cdots \ \alpha_n]'$,

$$\mathbf{x} = \mathbf{E}\boldsymbol{\alpha} = \mathbf{E}_1\boldsymbol{\alpha}_1 + \mathbf{E}_2\boldsymbol{\alpha}_2. \tag{2.3}$$

Due to orthonormality,

$$\begin{cases} \mathbf{E}_1'\mathbf{E}_1 = \mathbf{I}_r \\ \mathbf{E}_2'\mathbf{E}_2 = \mathbf{I}_{n-r} \\ \mathbf{E}_1'\mathbf{E}_2 = \mathbf{0}_{r,n-r} \\ \mathbf{E}_2'\mathbf{E}_1 = \mathbf{0}_{n-r,r} \end{cases}, \tag{2.4}$$

so that, with $\mathbf{E}\boldsymbol{\alpha}_1 = \mathbf{u}$,

$$\begin{aligned} \mathbf{E}_1\mathbf{E}_1'\mathbf{x} &= \mathbf{E}_1\mathbf{E}_1'\mathbf{E}\boldsymbol{\alpha} \\ &= \mathbf{E}_1\mathbf{E}_1'[\mathbf{E}_1\ \mathbf{E}_2]\begin{bmatrix}\boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2\end{bmatrix} \\ &= [\mathbf{E}_1\ \mathbf{0}_{n-r}]\begin{bmatrix}\boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2\end{bmatrix} \\ &= \mathbf{E}_1\boldsymbol{\alpha}_1 = \mathbf{u} \end{aligned} \tag{2.5}$$

So, the projection of a vector on a subspace can be defined:

Definition 1. *The vector* $\mathbf{u}$ *described above is called the orthogonal projection of* $\mathbf{x}$ *in* $\mathcal{S}$*.*

Nextly comes

Theorem 1. *Suppose the columns of matrix* $\mathbf{E}$ *form an orthonormal basis for linear subspace* $\mathcal{S}$*. Then*

$$\mathbf{P}_{\mathcal{S}}\mathbf{x} = \mathbf{E}\mathbf{E}'\mathbf{x} = \mathbf{u}$$

is the orthogonal projection of $\mathbf{x}$ *on* $\mathcal{S}$*.*

Proof. See [18], pg. 53. □

Thus, follows the natural definition of orthogonal projection matrix.

Definition 2. *The matrix* $\mathbf{P}_{\mathcal{S}}$ *is called the orthogonal projection matrix on* $\mathcal{S}$*.*

Obviously,

$$\begin{cases} \mathrm{R}(\mathbf{P}_{\mathcal{S}}) = \mathcal{S} \\ \mathrm{N}(\mathbf{P}_{\mathcal{S}}) = \mathcal{S}^{\perp} \end{cases}. \tag{2.6}$$

Obtaining the orthogonal projection matrix for the orthogonal complement of $\mathcal{S}$ is immediate, since, using previous notation,

$$(\mathbf{I}_n - \mathbf{E}_1\mathbf{E}_1')\mathbf{x} = \mathbf{x} - \mathbf{E}_1\boldsymbol{\alpha}_1 = \mathbf{E}_2\boldsymbol{\alpha}_2 = \mathbf{v}. \tag{2.7}$$

The next theorem shows that the choice of the orthonormal basis for a subspace is independent of the orthogonal projection matrix, rendering it unique.

Theorem 2. *Suppose the columns of each matrix,* $\mathbf{E}$ *and* $\mathbf{D}$*, form an orthonormal basis for linear subspace* $\mathcal{S}$*. Then*

$$\mathbf{E}\mathbf{E}' = \mathbf{D}\mathbf{D}'.$$

Proof. See [18], pg. 53. □

It is easy to see that $\mathbf{P}_{\mathcal{S}} = \mathbf{E}_1\mathbf{E}_1'$ is symmetric,

$$\mathbf{P}_{\mathcal{S}}' = (\mathbf{E}_1\mathbf{E}_1')' = \mathbf{E}_1\mathbf{E}_1' = \mathbf{P}_{\mathcal{S}},$$

and idempotent,

$$\mathbf{P}_{\mathcal{S}}\mathbf{P}_{\mathcal{S}} = \mathbf{E}_1\mathbf{E}_1'\mathbf{E}_1\mathbf{E}_1' = \mathbf{E}_1\mathbf{I}_r\mathbf{E}_1' = \mathbf{E}_1\mathbf{E}_1' = \mathbf{P}_{\mathcal{S}}.$$

The next theorem shows the converse.

Theorem 3. *Every symmetric and idempotent matrix* $\mathbf{P}$ *is an orthogonal projection matrix.*

Proof. See [18], pg. 59. □

The following lemma enumerates equivalent statements about orthogonal projection matrices and associated subspaces.

Lemma 1. *The following propositions are equivalent:*

1. $\mathbf{P}$ *is an orthogonal projection matrix;*
2. $\mathbf{I}_n - \mathbf{P}$ *is an orthogonal projection matrix;*
3. $\mathrm{R}(\mathbf{P}) = \mathrm{N}(\mathbf{I}_n - \mathbf{P})$*;*
4. $\mathrm{R}(\mathbf{I}_n - \mathbf{P}) = \mathrm{N}(\mathbf{P})$*;*
5. $\mathrm{R}(\mathbf{P}) \perp \mathrm{R}(\mathbf{I}_n - \mathbf{P})$*;*
6. $\mathrm{N}(\mathbf{P}) \perp \mathrm{N}(\mathbf{I}_n - \mathbf{P})$.

The direct sum of linear subspaces can also be represented by the range space of orthogonal projection matrices. In this case, the direct sum of orthogonal subspaces is represented by $\boxplus$. The two following theorems show this.

Theorem 4. *Let* $\mathbf{P}_1, ..., \mathbf{P}_k$ *be orthogonal projection matrices such that* $\mathbf{P}_i\mathbf{P}_j = \mathbf{0}_n$ *when* $i \neq j$*. Then:*

1. $\mathbf{P} = \sum_{i=1}^{k} \mathbf{P}_i$ *is an orthogonal projection matrix;*
2. $\mathrm{R}(\mathbf{P}_i) \cap \mathrm{R}(\mathbf{P}_j) = \mathbf{0}$ *when* $i \neq j$*;*
3. $\mathrm{R}(\mathbf{P}) = \bigoplus_{i=1}^{k} \mathrm{R}(\mathbf{P}_i)$.

Proof. See [15], pgs. 241–242. □

Theorem 5. *Let* $\mathbf{P}$ *be an orthogonal projection matrix associated with a linear subspace* $\mathcal{S}$*. Suppose that* $\mathcal{S}$ *is a direct sum of subspaces,* i.e.*,* $\mathcal{S} = \bigoplus_{i=1}^{k} \mathcal{S}_i$*. Then there exist unique projectors* $\mathbf{P}_1, ..., \mathbf{P}_k$ *such that* $\mathbf{P} = \sum_{i=1}^{k} \mathbf{P}_i$ *and* $\mathbf{P}_i\mathbf{P}_j = \mathbf{0}_{n\times n}$ *when* $i \neq j$*.*

Proof. See [15], pg. 242. □

It is interesting to observe that the last two theorems are the reverse of each other

Another simple way to obtain the orthogonal projection matrix of the column space, R($\mathbf{X}$), of a full rank matrix is (see [15], pg. 243) to consider:

$$\mathbf{P}_{\mathrm{R}(\mathbf{X})} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{2.8}$$

2.2 Generalized Inverses

Generalized inverses are a very versatile tool, both in linear algebra, as well as in statistics. Generalized inverses can be derived from two different contexts: equation systems and linear applications.

From the first point of view, let

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

be a system of linear equations. If $\mathbf{A}$ is invertible, the solution will be $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. If $\mathbf{A}$ is not invertible, there may still be a matrix $\mathbf{G}$ such that $\mathbf{x} = \mathbf{G}\mathbf{y}$ is a solution for the system, for every $\mathbf{y}$, whenever the system is consistent. Such a matrix is called a generalized inverse of $\mathbf{A}$.

The other point of view, which will be followed throughout the dissertation, takes matrices as linear applications. The adopted definition of generalized inverse will be based on this.

Definition 3. *Let $\mathbf{A} \in \mathcal{M}_{n\times m}$. A matrix $\mathbf{G} \in \mathcal{M}_{m\times n}$ such that*

$$\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}$$

*is called a generalized inverse (*g*-inverse) of $\mathbf{A}$, and it is expressed as $\mathbf{A}^{-}$.*

The problem that arises naturally is the question of existence: is there a g-inverse for every matrix? The answer is given by the next theorem.

Theorem 6. *For every* $\mathbf{A} \in \mathcal{M}_{n\times m}$, $\mathbf{A}^- \in \mathcal{M}_{m\times n}$ *exists.*

Proof. See [15], pg. 266. □

Introducing some useful properties on generalized inverses:

Theorem 7. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$. *Then:*

1. $(\mathbf{A}^-)'$ *is a* g*-inverse of* $\mathbf{A}'$;
2. *with* $\alpha \neq 0$, $\alpha^{-1}\mathbf{A}^-$ *is a* g*-inverse of* $\alpha\mathbf{A}$;
3. *if* $\mathbf{A}$ *is square and non-singular,* $\mathbf{A}^- = \mathbf{A}^{-1}$ *uniquely;*
4. $\mathbf{B}$ *and* $\mathbf{C}$ *are non-singular,* $\mathbf{C}^{-1}\mathbf{A}^-\mathbf{B}^{-1}$ *is a* g*-inverse of* $\mathbf{BAC}$;
5. $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{AA}^-) = \operatorname{rank}(\mathbf{A}^-\mathbf{A}) \leq \operatorname{rank}(\mathbf{A}^-)$;
6. $\operatorname{rank}(\mathbf{A}) = n \Leftrightarrow \mathbf{AA}^- = \mathbf{I}_n$;
7. $\operatorname{rank}(\mathbf{A}) = m \Leftrightarrow \mathbf{A}^-\mathbf{A} = \mathbf{I}_m$;
8. $\mathbf{AA}^-$ *is idempotent.*

Proof. Properties 1 through 7 are proved in [18], pg. 194. As for property 8,

$$\mathbf{AA}^-\mathbf{AA}^- = (\mathbf{AA}^-\mathbf{A})\mathbf{A}^- = \mathbf{A}.$$

□

For a given matrix, as shown before, there exists at least one g-inverse, but uniqueness doesn't certainly hold. Thus, a definition of g-inverse class is required.

Definition 4. *The set of all* g*-inverses of* $\mathbf{A}$ *is called the class of* g*-inverses of* $\mathbf{A}$, *and it is denoted by* $\{\mathbf{A}^-\}$.

The next theorem characterizes the elements in $\{\mathbf{A}^-\}$.

Theorem 8. *Let* $\mathbf{G}$ *be a* g*-inverse of* $\mathbf{A} \in \mathcal{M}_{n\times m}$*. Then, any* g*-inverse of* $\mathbf{A}$ *has one of the following forms:*

1. $\mathbf{A}^- = \mathbf{G} + \mathbf{U} - \mathbf{GAUAG}$*, for some* $\mathbf{U} \in \mathcal{M}_{m\times n}$*;*

2. $\mathbf{A}^- = \mathbf{G}+\mathbf{V}(\mathbf{I}_n-\mathbf{AG})-(\mathbf{I}_m-\mathbf{GA})\mathbf{W}$*, for some* $\mathbf{V} \in \mathcal{M}_{m\times n}$ *and* $\mathbf{W} \in \mathcal{M}_{m\times n}$*.*

Proof. See [15], pg. 277. □

By this result, the whole class of g-inverses of $\mathbf{A}$, $\{\mathbf{A}^-\}$, is obtained from one of it's members. Conversely, the class of g-inverses identifies the original matrix, since

Theorem 9. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be two matrices such that* $\{\mathbf{A}^-\} = \{\mathbf{B}^-\}$*. Then* $\mathbf{A} = \mathbf{B}$*.*

Proof. See [15], pgs. 277–278. □

But is there any sort of invariance when computing with such matrices? The important results that follow show that there is.

Theorem 10. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$*,* $\mathbf{B} \in \mathcal{M}_{p\times n}$ *and* $\mathbf{C} \in \mathcal{M}_{m\times q}$ *be matrices such that* $\mathrm{R}(\mathbf{B}') \subseteq \mathrm{R}(\mathbf{A}')$ *and* $\mathrm{R}(\mathbf{C}) \subseteq \mathrm{R}(\mathbf{A})$*. Then* $\mathbf{BG}_1\mathbf{C} = \mathbf{BG}_2\mathbf{C}$ *for all* $\mathbf{G}_1, \mathbf{G}_2 \in \{\mathbf{A}^-\}$*.*

Proof. See [15], pg. 268. □

As a consequence, the following corollary arises.

Corollary 1. $\mathbf{A}(\mathbf{A}'\mathbf{A})^-(\mathbf{A}'\mathbf{A}) = \mathbf{A}$ *and* $(\mathbf{A}'\mathbf{A})^-(\mathbf{A}'\mathbf{A})\mathbf{A}' = \mathbf{A}'$*.*

Proof. In [15], pg. 268, this corollary is proved for matrices with complex elements which includes matrices with real elements. □

The following results connect g-inverses with orthogonal projection matrices.

Theorem 11. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$*. Then* $\mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ *is the orthogonal projection matrix of* $\mathrm{R}(\mathbf{A})$*.*

Proof. If $\mathbf{A}$ is a real matrix then (see [15], in pg. 269) $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is symmetric and idempotent. Therefore, $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is an orthogonal projection matrix. It is obvious that

$$\mathrm{R}\big(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\big) \subseteq \mathrm{R}(\mathbf{A})\,.$$

Since $\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A}$,

$$\mathrm{R}(\mathbf{A}) = \mathrm{R}\big(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A}\big) \subseteq \mathrm{R}\big(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\big)\,,$$

and hence, $\mathrm{R}(\mathbf{A}) = \mathrm{R}(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}')$, which proves that $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is the orthogonal projection matrix of $\mathrm{R}(\mathbf{A})$. □

The g-inverse is the less restrictive case of the generalized inverse. Particularly, $\mathbf{A}^{-}$ being a g-inverse of $\mathbf{A}$ does not imply that $\mathbf{A}$ is a g-inverse of $\mathbf{A}^{-}$. When this happens, $\mathbf{A}^{-}$ is called a *reflexive* g-inverse.

Definition 5. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$ *and* $\mathbf{G} \in \mathcal{M}_{m\times n}$. *When* $\mathbf{AGA} = \mathbf{A}$ *and* $\mathbf{GAG} = \mathbf{G}$, $\mathbf{G}$ *is called a reflexive* g*-inverse. Such a matrix is denoted by* $\mathbf{A}_r^{-}$.

The real interesting fact that characterizes reflexive g-inverses is that, while in a general g-inverse $\mathrm{rank}(\mathbf{A}) \leq \mathrm{rank}(\mathbf{A}^{-})$, for a reflexive g-inverse $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A}_r^{-})$, as stated in

Theorem 12. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$. $\mathbf{G} \in \mathcal{M}_{m\times n}$ *is a reflexive* g*-inverse of* $\mathbf{A}$ *if and only if* $\mathbf{G}$ *is a* g*-inverse of* $\mathbf{A}$ *and* $\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{G})$.

Proof. See [15], pg. 279. □

Existence of g-inverses is also guaranteed.

Theorem 13. *For each* $\mathbf{A} \in \mathcal{M}_{n\times m}$, *a reflexive* g*-inverse exists.*

Proof. See [15], pg. 279. □

It is interesting to characterize the subspaces associated to a matrix $\mathbf{A}$ and it's g-inverses. The following theorem enlightens this connection.

Theorem 14. *Let $\mathbf{A} \in \mathcal{M}_{n\times m}$ and $\mathbf{A}^- \in \mathcal{M}_{m\times n}$ one of it's* g*-inverses. Let*

$$\begin{aligned}\mathbf{N} &= \mathbf{A}^- - \mathbf{A}^-\mathbf{A}\mathbf{A}^-;\\ \mathcal{M} &= \mathrm{R}\big(\mathbf{A}^- - \mathbf{N}\big) = \mathrm{R}\big(\mathbf{A}^-\mathbf{A}\mathbf{A}^-\big);\\ \mathcal{L} &= \mathrm{N}\big(\mathbf{A}^- - \mathbf{N}\big) = \mathrm{N}\big(\mathbf{A}^-\mathbf{A}\mathbf{A}^-\big).\end{aligned}$$

Then

1. $\mathbf{A}\mathbf{N} = \mathbf{0}_{n\times n}$ *and* $\mathbf{N}\mathbf{A} = \mathbf{0}_{m\times m}$*;*
2. $\mathcal{M} \cap \mathrm{N}(\mathbf{A}) = \{\mathbf{0}\}$ *and* $\mathcal{M} \oplus \mathrm{N}(\mathbf{A}) = \mathbb{R}^m$*;*
3. $\mathcal{L} \cap \mathrm{R}(\mathbf{A}) = \{\mathbf{0}\}$ *and* $\mathcal{L} \oplus \mathrm{R}(\mathbf{A}) = \mathbb{R}^n$.

Proof. Thesis 1 is quite easy to prove since

$$\mathbf{N}\mathbf{A} = \mathbf{A}^-\mathbf{A} - \mathbf{A}^-\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}^-\mathbf{A} - \mathbf{A}^-\mathbf{A} = \mathbf{0}_{m\times m}$$

and

$$\mathbf{A}\mathbf{N} = \mathbf{A}\mathbf{A}^- - \mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}\mathbf{A}^- - \mathbf{A}\mathbf{A}^- = \mathbf{0}_{n\times n}.$$

As for thesis 2 and 3, see [15], pgs. 282–283. □

Conversely, choosing $\mathbf{N}$, $\mathcal{M}$ and $\mathcal{L}$, a g-inverse of $\mathbf{A}$ for which the conditions in the previous theorem hold may be found.

Theorem 15. *Let $\mathbf{A} \in \mathcal{M}_{n\times m}$. Choosing a matrix $\mathbf{N} \in \mathcal{M}_{m\times n}$ such that $\mathbf{A}\mathbf{N} = \mathbf{0}_{n\times n}$ and $\mathbf{N}\mathbf{A} = \mathbf{0}_{m\times m}$, a subspace $\mathcal{L}$ of $\mathbb{R}^n$ such that $\mathcal{L} \cap \mathrm{R}(\mathbf{A}) = \{\mathbf{0}\}$ and $\mathcal{L} \oplus \mathrm{R}(\mathbf{A}) = \mathbb{R}^n$, and a subspace $\mathcal{M} \cap \mathrm{N}(\mathbf{A}) = \{\mathbf{0}\}$ of $\mathbb{R}^m$ such that $\mathcal{M} \cap \mathrm{N}(\mathbf{A}) = \{\mathbf{0}\}$ and $\mathcal{M} \oplus \mathrm{N}(\mathbf{A}) = \mathbb{R}^m$. Then there exists a* g*-inverse of $\mathbf{A}$, $\mathbf{A}^-$, such that*

1. $\mathbf{N} = \mathbf{A}^- - \mathbf{A}^-\mathbf{A}\mathbf{A}^-$*;*

2. $\mathcal{M} = \mathrm{R}(\mathbf{A}^- - \mathbf{N}) = \mathrm{R}(\mathbf{A}^-\mathbf{A}\mathbf{A}^-)$*;*

3. $\mathcal{L} = \mathrm{N}(\mathbf{A}^- - \mathbf{N}) = \mathrm{N}(\mathbf{A}^-\mathbf{A}\mathbf{A}^-)$.

Proof. See [15], pg. 283. □

In virtue of the specificity of this construction, a separate definition is in order.

Definition 6. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$*. A matrix for which the assumptions of Theorem 15 hold is called a* $\mathcal{LMN}$*-inverse.*

Uniqueness of $\mathcal{LMN}$-inverses is also guaranteed.

Theorem 16. *A* $\mathcal{LMN}$*-inverse of a matrix* $\mathbf{A}$ *is unique.*

Proof. See [15], pg. 284. □

The relationship between the subspaces $\mathcal{M}$ and $\mathcal{L}$, and the matrix $\mathbf{N}$ with the matrix $\mathbf{A}$ is illustrated in the figure

Finally, a bijection between between all triplets $(\mathcal{M}, \mathcal{L}, \mathbf{N})$ and all g-inverses is established in

Theorem 17. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$*, and consider* $\mathcal{L} \subseteq \mathbb{R}^n$*,* $\mathcal{M} \subseteq \mathbb{R}^m$ *and* $\mathbf{N} \in \mathcal{M}_{m\times n}$ *as defined in Theorem 15. Then there exists a bijection between the set of all triplets* $(\mathcal{M}, \mathcal{L}, \mathbf{N})$ *and the set of all* g*-inverses,* $\{\mathbf{A}^-\}$.

Proof. See [15], pg. 286. □

Furthermore, a bijection between all pairs $(\mathcal{L}, \mathcal{M})$ and all reflexive g-inverses can be established.

Theorem 18. *Let* $\mathbf{A} \in \mathcal{M}_{n\times m}$*, and consider* $\mathcal{L} \subseteq \mathbb{R}^n$ *and* $\mathcal{M} \subseteq \mathbb{R}^m$ *as defined in Theorem 15. Then there exists a bijection between the set of all triplets* $(\mathcal{M}, \mathcal{L}, \mathbf{N})$ *and the set of all reflexive* g*-inverses.*

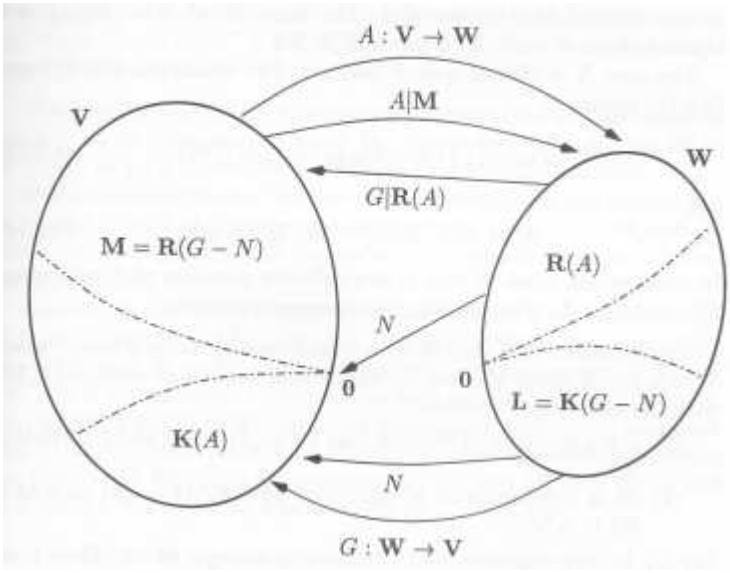

Fig. 2.1: Subspaces $\mathcal{M}$ and $\mathcal{L}$, and the matriz $\mathbf{N}$

Proof. See [15], pg. 286. □

Another interesting fact is the connection between subspaces $\mathcal{L}$ and $\mathcal{M}$, and their orthogonal projection matrices.

Theorem 19. *Let* $\mathbf{A}^-$ *be a* g*-inverse of* $\mathbf{A}$*. With*

- $\mathbf{N} = \mathbf{A}^- - \mathbf{A}^-\mathbf{A}\mathbf{A}^-$*;*
- $\mathcal{M} = \mathrm{R}(\mathbf{A}^- - \mathbf{N})$*;*
- $\mathcal{L} = \mathrm{N}(\mathbf{A}^- - \mathbf{N})$*;*

follows

$$\mathbf{A}\mathbf{A}^- = \mathbf{P}_{\mathrm{R}(\mathbf{A})} \wedge \mathbf{A}^-\mathbf{A} = \mathbf{P}_{\mathcal{M}} \Leftrightarrow \mathcal{M} \oplus \mathrm{N}(\mathbf{A}) = \mathbb{R}^m \wedge \mathrm{R}(\mathbf{A}) \oplus \mathcal{L} = \mathbb{R}^n.$$

Proof. Suppose $\mathbf{A}\mathbf{A}^- = \mathbf{P}_{\mathrm{R}(\mathbf{A})}$. Then, for all $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$, with $\mathbf{y}_1 \in \mathrm{R}(\mathbf{A})$. Now, for all $\mathbf{y}_2 = \mathbf{y} - \mathbf{y}_1$,

$$\mathbf{A}\mathbf{A}^-\mathbf{y}_2 = \mathbf{0} \Rightarrow \mathbf{A}^-\mathbf{A}\mathbf{A}^-\mathbf{y}_2 = \mathbf{0} \Rightarrow (\mathbf{A}^- - \mathbf{N})\mathbf{y}_2 = \mathbf{0}.$$

This means that any element that is linearly independent of $\mathrm{R}(\mathbf{A})$ belongs to $\mathcal{L}$. Suppose now that $\mathbf{A}^{-}\mathbf{A} = \mathbf{P}_{\mathcal{M}}$. Let $\mathbf{x} \in \mathbb{R}^m$ be expressed as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, with $\mathbf{x}_1 \in \mathcal{M}$. Then

$$\mathbf{A}^{-}\mathbf{A}\mathbf{x}_2 = \mathbf{0} \Rightarrow \mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{x}_2 = \mathbf{0} \Rightarrow \mathbf{A}\mathbf{x}_2 = \mathbf{0},$$

which means that any element that is linearly independent of $\mathcal{M}$ belongs to $\mathrm{N}(\mathbf{A}^{-} - \mathbf{N})$. This shows that $\mathbf{A}\mathbf{A}^{-} = \mathbf{P}_{\mathrm{R}(\mathbf{A})} \wedge \mathbf{A}^{-}\mathbf{A} = \mathbf{P}_{\mathcal{M}} \Rightarrow \mathcal{M} \oplus \mathrm{N}(\mathbf{A}) = \mathbb{R}^m \wedge \mathrm{R}(\mathbf{A}) \oplus \mathcal{L} = \mathbb{R}^n$. Suppose now that $\mathrm{R}(\mathbf{A}) \oplus \mathcal{L} = \mathbb{R}^n$. With $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2$, $\mathbf{y}_1 \in \mathrm{R}(\mathbf{A})$ and $\mathbf{y}_2 \in \mathcal{L}$,

$$\begin{aligned}
\mathbf{A}\mathbf{A}^{-}\mathbf{y} = \mathbf{A}\mathbf{A}^{-}\mathbf{y}_1 + \mathbf{A}\mathbf{A}^{-}\mathbf{y}_2 \Rightarrow (\mathbf{A}^{-} - \mathbf{N})\mathbf{y} &\\
= (\mathbf{A}^{-} - \mathbf{N})\mathbf{y}_1 + (\mathbf{A}^{-} - \mathbf{N})\mathbf{y}_2 &\\
= (\mathbf{A}^{-} - \mathbf{N})\mathbf{y}_1 &\\
\Rightarrow \mathbf{A}\mathbf{A}^{-}\mathbf{y} = \mathbf{A}\mathbf{A}^{-}\mathbf{y}_1 = \mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{x} &\\
= \mathbf{A}\mathbf{x} = \mathbf{y}_1, &
\end{aligned}$$

with $\mathbf{y}_1 = \mathbf{A}\mathbf{x}$. Thus, $\mathbf{A}\mathbf{A}^{-} = \mathbf{P}_{\mathrm{R}(\mathbf{A})}$. Now, for all $\mathbf{x} \in \mathbb{R}^m$, let $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, with $\mathbf{x}_1 \in \mathcal{M}$ and $\mathbf{x}_2 \in \mathrm{N}(\mathbf{A})$. Since

$$\begin{aligned}
\mathbf{A}^{-}\mathbf{A}\mathbf{x} &= \mathbf{A}^{-}\mathbf{A}\mathbf{x}_1 + \mathbf{A}^{-}\mathbf{A}\mathbf{x}_2 \\
&= \mathbf{A}^{-}\mathbf{A}\mathbf{x}_1 \\
&= \mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-}\mathbf{y} \\
&= \mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-}\mathbf{y} \\
&= \mathbf{x}_1,
\end{aligned}$$

where $\mathbf{x}_1 = (\mathbf{A}^{-} - \mathbf{N})\mathbf{y}$. Thus $\mathbf{A}^{-}\mathbf{A} = \mathbf{P}_{\mathcal{M}}$, and the proof is complete. □

A special type of $\mathcal{L}\mathcal{M}\mathbf{N}$-inverse is obtained taking $\mathbf{N} = \mathbf{0}_{m\times m}$, $\mathcal{M} = \mathrm{N}(\mathbf{A})^{\perp}$ and $\mathcal{L} = \mathrm{R}(\mathbf{A})^{\perp}$. This set of conditions is equivalent to the set of conditions

1. $\mathbf{G}$ is a reflexive g-inverse of $\mathbf{A}$;
2. $\mathbf{A}\mathbf{G} = \mathbf{P}_{\mathrm{R}(\mathbf{A})}$;

3. $\mathbf{GA} = \mathbf{P}_{\mathrm{R}(\mathbf{G})}$.

According to [18], pg. 173 (Theorem 5.2), these conditions are equivalent to having

1. $\mathbf{G}$ is a reflexive g-inverse of $\mathbf{A}$;

2. $(\mathbf{AG})' = \mathbf{AG}$;

3. $(\mathbf{GA})' = \mathbf{GA}$.

These are the properties that define the *Moore-Penrose inverse*.

Definition 7. *For a matrix $\mathbf{A} \in \mathcal{M}_{m\times n}$, a matrix $\mathbf{G}$ is denominated its Moore-Penrose inverse if the following hold:*

- $\mathbf{G}$ *is a reflexive* g*-inverse of* $\mathbf{A}$*;*
- $(\mathbf{AG})' = \mathbf{AG}$*;*
- $(\mathbf{GA})' = \mathbf{GA}$.

2.3 Jordan Algebras

This section is entirely dedicated to the algebraic structures known as Jordan algebras. These structures were first introduced by Pascual Jordan (who named these algebras), John von Neumann and Eugene P. Wigner in quantum theory, as an alternative algebraic structure for quantum mechanics. In this section Jordan algebras will be defined and their properties explored.

For completeness sake, the definition of algebras is stated.

Definition 8. *An algebra $\mathcal{A}$ is a linear space equipped with a binary operation $*$, usually denominated* product*, in which the following properties hold for all $\alpha \in \mathbb{R}$ and $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{A}$:*

- $\mathbf{a} * (\mathbf{b} + \mathbf{c}) = \mathbf{a} * \mathbf{b} + \mathbf{a} * \mathbf{c}$;
- $(\mathbf{a} + \mathbf{b}) * \mathbf{c} = \mathbf{a} * \mathbf{c} + \mathbf{b} * \mathbf{c}$;
- $\alpha(\mathbf{a} * \mathbf{b}) = (\alpha\mathbf{a}) * \mathbf{b} = \mathbf{a} * (\alpha\mathbf{b})$.

It is important to underline that associativity and commutativity are properties that are not necessary for a linear space to be an algebra. The definitions follow.

Definition 9. *An algebra $\mathcal{A}$ is associative if and only if, for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{A}$,*

$$(\mathbf{a} * \mathbf{b}) * \mathbf{c} = \mathbf{a} * (\mathbf{b} * \mathbf{c}).$$

Definition 10. *An algebra $\mathcal{A}$ is commutative if and only if, for all $\mathbf{a}, \mathbf{b} \in \mathcal{A}$,*

$$\mathbf{a} * \mathbf{b} = \mathbf{b} * \mathbf{a}.$$

As for all algebraic structures, substructures are obtained, in this case easily.

Definition 11. *A set $\mathcal{S}$ contained in an algebra $\mathcal{A}$ is a subalgebra if it is a linear subspace and if*

$$\forall \mathbf{a}, \mathbf{b} \in \mathcal{S} : \mathbf{a} * \mathbf{b} \in \mathcal{S}.$$

In order to define a Jordan algebra, one needs to define the Jordan product in an algebra.

Definition 12. *A Jordan product is a binary operation $\cdot$ defined on an algebra $\mathcal{A}$ for which the following two conditions, $\mathbf{a}, \mathbf{b} \in \mathcal{A}$:*

- $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$;
- $\mathbf{a}^{2\cdot} \cdot (\mathbf{b} \cdot \mathbf{a}) = (\mathbf{a}^{2\cdot} \cdot \mathbf{b}) \cdot \mathbf{a}$,

with $\mathbf{a}^{2\cdot} = \mathbf{a} \cdot \mathbf{a}$, hold. An algebra equipped with such a product is a Jordan algebra.

This is the primary definition for a Jordan algebra. Other more intuitive and tractable equivalent definitions will follow. A Jordan subalgebra is equally defined.

Definition 13. *A set $\mathcal{S}$ contained in a Jordan algebra $\mathcal{A}$ is a Jordan subalgebra if it is a linear subalgebra of the algebra $\mathcal{A}$ and if*

$$\forall \mathbf{a}, \mathbf{b} \in \mathcal{S} : \mathbf{a} \cdot \mathbf{b} \in \mathcal{S}.$$

Again, more natural definitions of Jordan subalgebras will be presented.

For illustration purposes, consider an example of a Jordan algebra. Let $\mathcal{S}_n$ be the space of symmetric real matrices of size n. It is a linear space with dimension $\frac{n(n+1)}{2}$, equipped with the (associative) matrix product. If the $\cdot$ product is defined as

$$\mathbf{A} \cdot \mathbf{B} = \frac{1}{2}(\mathbf{AB} + \mathbf{BA}), \tag{2.9}$$

it is quite easy to see that the $\cdot$ product is a Jordan product in S_n, and S_n is in fact a Jordan algebra. This is a very important instance of Jordan algebras, since it is the space in which covariance matrices lie. In order to show the importance of this Jordan algebra, some definitions must be given first.

Definition 14. *Algebra $\mathcal{A}$ is algebra-isomorphic to algebra $\mathcal{B}$ if and only if there exists a bijection $\phi : \mathcal{A} \mapsto \mathcal{B}$ such that, for all $\alpha, \beta \in \mathbb{R}$ and $\mathbf{a}, \mathbf{b} \in \mathcal{A}$,*

1. $\phi(\alpha\mathbf{a} + \beta\mathbf{b}) = \alpha\phi(\mathbf{a}) + \beta\phi(\mathbf{b})$;

2. $\phi(\mathbf{a} * \mathbf{b}) = \phi(\mathbf{a}) * \phi(\mathbf{b})$.

With algebra-isomorphism it is then possible to identify *special Jordan algebras*.

Definition 15. *A Jordan algebra algebra-isomorphic to a subalgebra of $\mathcal{S}_n$ is denominated a special Jordan algebra.*

Matrix algebras are the fulcral application of Jordan algebras in statistics. Most of the following results are presented in the context of such spaces. Nextly, identity elements and idempotent elements are formally defined.

Definition 16. *Let* $\mathbf{E} \in \mathcal{S}$*,* $\mathcal{S}$ *a subspace of* $\mathcal{M}_n$*. The element* $\mathbf{E}$ *is an associative identity element if* $\mathbf{ES} = \mathbf{SE} = \mathbf{S}$*,* $\forall \mathbf{S} \in \mathcal{S}$*.*

It is important to remember that, with $\mathcal{S}$ being a subset of $\mathcal{M}_n$, the identity element is not necessarily $\mathbf{I}_n$.

Definition 17. *Let* $\mathbf{E} \in \mathcal{M}_n$*. If* $\mathbf{E}^2 = \mathbf{E}$*,* $\mathbf{E}$ *is said to be idempotent.*

Here, it is also important to refer that if an element is idempotent, it is also Jordan idempotent: $\mathbf{E}^{2\cdot} = \mathbf{E}$. A similar result is obtained for Jordan identities:

Theorem 20. *Let* $\mathcal{S}$ *be a subspace of* $\mathcal{S}_n$ *and* $\mathbf{E} \in \mathcal{S}$ *an idempotent element. Then*

$$\exists \mathbf{S} \in \mathcal{S} : \mathbf{E} \cdot \mathbf{S} = \mathbf{E} \cdot \mathbf{S} \Rightarrow \mathbf{ES} = \mathbf{SE} = \mathbf{S}.$$

Proof. See [11], pages 9 and 10. □

The next theorem sets the connection between idempotency, orthogonality and Jordan orthogonality.

Theorem 21. *Let* $\mathcal{S}$ *and* $\mathcal{T}$ *be subspaces of* $\mathcal{S}_n$*. Then the following three conditions hold:*

1. $\forall \mathbf{S} \in \mathcal{S}, \mathbf{E} \cdot \mathbf{S} = \mathbf{S} \Leftrightarrow \forall \mathbf{S} \in \mathcal{S}, \mathbf{ES} = \mathbf{S} = \mathbf{SE}$*;*

2. *with* $\mathbf{E}_1$ *and* $\mathbf{E}_2$ *idempotent elements,* $\mathbf{E}_1 \cdot \mathbf{E}_2 = \mathbf{0} \Leftrightarrow \mathbf{E}_1\mathbf{E}_2 = \mathbf{0}$*;*

3. $\forall \mathbf{S} \in \mathcal{S}, \forall \mathbf{T} \in \mathcal{T}, \mathbf{S} \cdot \mathbf{T} = \mathbf{0} \Leftrightarrow \forall \mathbf{S} \in \mathcal{S}, \forall \mathbf{T} \in \mathcal{T}, \mathbf{ST} = \mathbf{TS} = \mathbf{0}$*.*

Proof. See [11], page 10. □

After these remarks on some properties of Jordan algebras, more standard equivalent definitions of Jordan algebras are presented, included the one in [22].

Theorem 22. *Let $\mathcal{S}$ be a subspace of $\mathcal{S}_n$ with identity element $\mathbf{E}$ and $\mathbf{A}$ and $\mathbf{B}$ any two elements of $\mathcal{S}$. Then $\mathcal{S}$ is a Jordan algebra if and only if any of the following equivalent definitions hold:*

1. $\mathbf{AB} + \mathbf{BA} \in \mathcal{S}$;
2. $\mathbf{ABA} \in \mathcal{S}$;
3. $\mathbf{ABC} + \mathbf{CBA} \in \mathcal{S}$;
4. $\mathbf{A}^2 \in \mathcal{S}$;
5. $\mathbf{A}^+ \in \mathcal{S}$.

Proof. For the proof of the assertions in this theorem see [6], [22], [32] and [11]. □

It is easy to see that, by the second and third conditions of the last theorem, any power of a matrix belonging to a Jordan algebra belongs to that Jordan algebra.

It is interesting to see that one of the definitions of Jordan algebra is based on the presence of the Moore-Penrose inverse of each matrix in the algebra.

Like in all linear spaces, defining a inner product in Jordan algebras is very important. Although many functions can qualify as inner products in Jordan algebras, but one of the most common functions is defined, in $\mathcal{S}$, as

$$<\mathbf{A}|\mathbf{B}>= \mathrm{tr}(\mathbf{AB}')\,. \tag{2.10}$$

Consider now that $\mathcal{S}$, spanned by $\{\mathbf{M}_1, ..., \mathbf{M}_s\}$ is included in $\mathcal{S}_n$ and let $\mathbf{I}_n \in \mathcal{S}$. Let $\mathcal{A}(\mathcal{S})$ be the smallest associative algebra in $\mathcal{M}_n$ containing $\mathcal{S}$. Also define

$$\mathcal{B}(\mathcal{S}) = \{\mathbf{B} \in \mathcal{S} : \mathbf{SBS} \in \mathcal{S}, \forall \mathbf{S} \in \mathcal{S}\}. \tag{2.11}$$

Given this,

Theorem 23. *$\mathcal{B}(\mathcal{S})$ is the maximal subspace of $\mathcal{S}$ such that*

$$\mathbf{BSB} \in \mathcal{S}, \forall \mathbf{S} \in \mathcal{S}, \forall \mathbf{B} \in \mathcal{B}(\mathcal{S}).$$

Proof. See [11], Chapter 3. □

Finally, consider $\mathcal{J}(\mathcal{S}) \subseteq \mathcal{S}_n$ the smallest Jordan algebra containing $\mathcal{S}$.

Theorem 24. *Given $\mathcal{S}$, $\mathcal{B}(\mathcal{S})$, $\mathcal{A}(\mathcal{S})$ and $\mathcal{J}(\mathcal{S})$,*

$$\mathcal{B}(\mathcal{S}) \subseteq \mathcal{S} \subseteq \mathcal{J}(\mathcal{S}) \subseteq \mathcal{A}(\mathcal{S}) \cap \mathcal{S}_n \subseteq \mathcal{A}(\mathcal{S}).$$

Proof. The proof of this theorem can be found on [11], in pages 16 through 18. □

It is important to notice that the inclusion $\mathcal{J}(\mathcal{S}) \subseteq \mathcal{A}(\mathcal{S}) \cap \mathcal{S}_n$ is in general strict, meaning that Jordan algebras generated by sets of symmetric matrices have lower dimensions than the associative algebra generated by the same set. Equality holds only when the matrices in the generating set commute, *i.e.*, $\mathbf{AB} = \mathbf{BA}$.

On the following paragraphs, emphasis will be given to the decomposition in partitions of Jordan algebras as linear spaces and conditions given on orthogonality between matrices contained in such partitions. The section will culminate with the seminal result by Seely, pointing an unique basis constituted by mutually orthogonal projection matrices for commutative Jordan algebras. The following definition is necessary:

Definition 18. *Let $\mathcal{S} = \bigotimes_{i=1}^{s} \mathcal{S}_i \subseteq \mathcal{S}_n$. For each element $\mathbf{A}$ of $\mathcal{S}$, $\mathbf{A} = \sum_{i=1}^{s} \mathbf{A}_i$, with $\mathbf{A}_i \in \mathcal{S}_i$. The set $\{i \in \mathbb{N} : \mathbf{A}_i \neq 0\}$ is called the support of $\mathbf{A}$.*

With this definition in mind, the following results are obtained:

Theorem 25. *Let $\mathcal{S} \subseteq \mathcal{S}_n$ and $\mathbf{I}_n \in \mathcal{S}$. Then, for $\mathbf{A}, \mathbf{B} \in \mathcal{B}(\mathcal{S})$, the following propositions are equivalent:*

1. $\mathbf{SASBS} = \mathbf{0}_{n \times n}, \forall \mathbf{S} \in \mathcal{S}$;

2. $\mathbf{A}\mathcal{S}\mathbf{B} = \mathbf{0}_{n\times n}$;

3. $\mathbf{A}\mathcal{B}(\mathcal{S})\mathbf{B} = \mathbf{0}_{n\times n}$;

4. $\mathbf{A}$ *and* $\mathbf{B}$ *have disjoint support in* $\mathcal{B}(\mathcal{S})$.

Proof. See [11], Chapter 3. □

Theorem 26. *Let* $\mathcal{S} \subseteq \mathcal{S}_n$ *and* $\mathbf{I}_n \in \mathcal{S}$. *Then, for* $\mathbf{A}, \mathbf{B} \in \mathcal{J}(\mathcal{S})$, *the following propositions are equivalent:*

1. $\mathbf{SASBS} = \mathbf{0}_{n\times n}, \forall \mathbf{S} \in \mathcal{J}(\mathcal{S})$;

2. $\mathbf{A}\mathcal{J}(\mathcal{S})\mathbf{B} = \mathbf{0}_{n\times n}$;

3. $\mathbf{A}$ *and* $\mathbf{B}$ *have disjoint support in* $\mathcal{J}(\mathcal{S})$.

Proof. See [11], Chapter 3. □

Theorem 27. *Let* $\mathbf{A}, \mathbf{B} \in \mathcal{S}_n$. *Then*

$$\mathbf{A}\mathcal{S}\mathbf{B} = \mathbf{0}_{n\times n} \Leftrightarrow \mathbf{A}\mathcal{J}(\mathcal{S})\mathbf{B} = \mathbf{0}_{n\times n}.$$

Proof. See [11], pages 25–26. □

Theorem 28. *Let* $\mathcal{S} \subseteq \mathcal{S}_n$ *and* $\mathbf{I}_n \in \mathcal{S}$. *Then, for* $\mathbf{A}, \mathbf{B} \in \mathcal{J}(\mathcal{S})$, *the following propositions are equivalent:*

1. $\mathbf{SASBS} = \mathbf{0}_{n\times n}, \forall \mathbf{S} \in \mathcal{S}$;

2. $\mathbf{A}\mathcal{S}\mathbf{B} = \mathbf{0}_{n\times n}$;

3. $\mathbf{A}\mathcal{J}(\mathcal{S})\mathbf{B} = \mathbf{0}_{n\times n}$;

4. $\mathbf{A}$ *and* $\mathbf{B}$ *have disjoint supports in* $\mathcal{J}(\mathcal{S})$.

Proof. The proof for this theorem follows directly from Theorem 25 and Theorem 26. □

The last part of this section is dedicated to the sets of matrices that form basis for commutative Jordan algebras. First, the existence of a basis is settled.

Theorem 29. *For every commutative Jordan algebra there exists at least one basis*

$$\mathbf{Q}_1, \ldots, \mathbf{Q}_w$$

constituted by orthogonal projection matrices such that $\mathbf{Q}_i\mathbf{Q}_j = \mathbf{0}_{n\times n}$*, for* $i \neq j$*.*

Proof. See [22], in which one may find an algorithm for the construction of this basis. □

Theorem 30 (Seely, 1971)**.** *A necessary and sufficient condition for a subspace* $\mathcal{S} \subseteq \mathcal{S}_n$ *to be a commutative Jordan algebra is the existence of a basis* $\{\mathbf{Q}_1, ..., \mathbf{Q}_s\}$ *formed by orthogonal projection matrices, such that* $\mathbf{Q}_i\mathbf{Q}_j = \mathbf{0}_{n\times n}$*, for* $i \neq j$*,* $i, j = 1, ..., s$*. Furthermore, this basis is unique.*

Proof. The existence was proven in Theorem 29. As for the uniqueness, consider another basis formed by orthogonal projection matrices, $\{\mathbf{P}_1, ..., \mathbf{P}_s\}$. Hence, for $j = 1, ..., s$,

$$\mathbf{P}_j = \sum_{i=1}^{s} \alpha_i \mathbf{Q}_i \text{ and } \mathbf{Q}_i = \sum_{j=1}^{s} \beta_{i,j} \mathbf{P}_j,$$

with the coefficients α_i and $\beta_{i,j}$, $i, j = 1, ..., s$, being unique and equal to 0 or 1, since the matrices $\mathbf{P}_j$ and $\mathbf{Q}_i$, $i, j = 1, ..., s$, are orthogonal projection matrices. Thus, choosing j,

$$\mathbf{P}_j = \beta_{i,j}\mathbf{P}_j = \mathbf{Q}_i\mathbf{P}_j = \alpha_i\mathbf{Q}_i = \mathbf{Q}_i,$$

for some i. Since this holds for all $j = 1, ..., s$, the proof is complete. □

3. INFERENCE

In this chapter the discussion will be focused on the estimation of parameters and functions of parameters of linear models, namely fixed effects, variance components and functions of fixed effects and variance components.

Estimators, in a broad sense, will be presented as well as their properties. Nextly, estimators for parameters and functions of parameters in linear models will be presented and their properties discussed.

3.1 Estimators

In order to define an estimator for any given parameter, one needs to retrieve information from a (finite) sample. Usually a function of the observations vector is taken – a *statistic* or *estimator* – in order to accomplish this mission.

It is desirable, of course, that statistics contain as much information about the parameter as possible. A usual condition is that the statistic is *sufficient*. The definition follows.

Definition 19. *Given the sample vector* $\mathbf{X} = (X_1, ..., X_n)'$ *where each component has its probability density belonging to a family of densities* $\mathcal{P} = \{f_{\boldsymbol{\theta}}(\cdot) : \boldsymbol{\theta} \in \Omega\}$, *the function* $\mathbf{T}(\mathbf{X}) = \mathbf{T}(X_1, ..., X_n)$ *is a sufficient statistic if the sampled density* $f_{\boldsymbol{\theta}}(\cdot) \in \mathcal{P}$, *given* $\mathbf{T}(\mathbf{X}) = \mathbf{t}$, *does not depend on* $\boldsymbol{\theta}$.

This definition guarantees that a sufficient statistic summarizes all the information contained in a sample about parameter $\boldsymbol{\theta}$, but it becomes quite intractable

to establish this property. Usually the following theorem is used as a sufficiency criterium.

Theorem 31 (Factorization Criterion)**.** $\mathbf{T}(\mathbf{x})$ *is a sufficient statistic to* $\boldsymbol{\theta}$ *if and only if*

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = g_{\boldsymbol{\theta}}\big(\mathbf{T}(\mathbf{x})\big)h(\mathbf{x}).$$

Proof. See [9], pp. 54 and 55. □

Sufficiency does in fact guarantee that all the necessary information is retrieved from a sample, but it doesn't stop one from shooting a fly with a cannon. Many sufficient statistics exist for a given sample from a population with density $f_{\boldsymbol{\theta}}$ – from a vector of sample observations functions to even a single one – and it is desirable to have a statistic as condensed as possible, since for every set of distributions, the whole sample is sufficient. To aid in such a goal, the concept of *minimal sufficient statistics* is introduced:

Definition 20. *A sufficient statistic* $\mathbf{T}$ *is minimal if for all sufficient statistic* $\mathbf{U}$ *there exists a function* $\mathbf{g}$ *such that* $\mathbf{T} = \mathbf{g}(\mathbf{U})$.

Besides this, there are statistics that do not carry any kind of information about the parameter, although they do carry information about the sample itself. These are called *ancillary* statistics, and they are valuable to determine the amount of "useless" information carried by a statistic. The definition:

Definition 21. *A statistic* $\mathbf{V}$ *is called ancillary if its distribution is independent of the parameter* $\boldsymbol{\theta}$.

Opposite to this definition is *completeness*, which says that a sufficient statistic doesn't carry useless – ancillary – information. Formally,

Definition 22. *A statistic* $\mathbf{T}$ *is complete if and only if*

$$\mathbb{E}[\mathbf{g}(\mathbf{T})] = 0, \forall \boldsymbol{\theta} \in \Omega \quad \Rightarrow \quad \mathbb{P}[\mathbf{g} \equiv \mathbf{0}] = 1.$$

It is, thus, fair to assume that a complete statistic is independent of any ancillary information. This is stated in the next theorem.

Theorem 32 (Basu's Theorem)**.** *Given $\boldsymbol{\theta}$, if $\mathbf{T}$ is a complete sufficient statistic then any ancillary statistic $\mathbf{V}$ is independent of $\mathbf{T}$.*

Proof. See [10], pg. 42. □

Another expected fact is that every complete sufficient statistic is a minimal sufficient statistic.

Theorem 33. *If a statistic $\mathbf{T}$ is complete sufficient then it is is minimal sufficient.*

Proof. See [17]. □

It is necessary to have some kind of efficiency measure for estimators. First, define the *loss function*.

Definition 23. *Let $\mathbf{T}$ be an estimator of $\mathbf{g}(\boldsymbol{\theta})$. The real-valued function $\ell(\mathbf{t}, \boldsymbol{\theta})$ is a loss function if it respects the following conditions:*

1. $\ell(\mathbf{t}, \boldsymbol{\theta}) \geq 0$;
2. $\ell\big(\mathbf{g}(\boldsymbol{\theta}), \boldsymbol{\theta}\big) = 0$.

This function assigns a penalization to an estimate obtained by $\mathbf{T}$ different of $\mathbf{g}(\boldsymbol{\theta})$. But of course that the loss isn't known, so, one must find a way to assess the average loss: this is accomplished using the *risk function*.

Definition 24. *Let $\mathbf{T}$ be an estimator of $\mathbf{g}(\boldsymbol{\theta})$ and $\ell(\mathbf{t}, \boldsymbol{\theta})$ its loss function. The risk function is defined as*

$$\mathcal{R}_{\boldsymbol{\theta}}(\mathbf{T}) = \mathbb{E}[\ell(\mathbf{T}, \boldsymbol{\theta})]\,.$$

It also provides an order relationship between estimators. So,

$$\mathbf{T}_1 \prec \mathbf{T}_2 \text{ if } \forall \boldsymbol{\theta} \in \Omega, \mathcal{R}_{\boldsymbol{\theta}}(\mathbf{T}_1) \leq \mathcal{R}_{\boldsymbol{\theta}}(\mathbf{T}_2) \wedge \exists \boldsymbol{\theta} \in \Omega : \mathcal{R}_{\boldsymbol{\theta}}(\mathbf{T}_1) < \mathcal{R}_{\boldsymbol{\theta}}(\mathbf{T}_2). \tag{3.1}$$

From this ordering of estimators comes the definition of *admissibility*.

Definition 25. *An estimator* $\mathbf{T}$ *is admissible if there exists no estimator* $\mathbf{U}$ *such that* $\mathbf{U} \prec \mathbf{T}$.

Given the definition of risk and sufficiency, it is possible to enunciate an important theorem that permits the improvement of an estimator not built upon a sufficient statistic: the *Rao-Blackwell theorem*.

Theorem 34 (Rao-Blackwell Theorem)**.** *Given an estimator* $\mathbf{U}$ *of* $\mathbf{g}(\boldsymbol{\theta})$ *with a loss function* $\ell(\mathbf{u}, \boldsymbol{\theta})$ *strictly convex on* $\mathbf{u}$ *and such that*

- $\mathbb{E}[\mathbf{U}] < \infty$*;*
- $\mathcal{R}_{\boldsymbol{\theta}}(\mathbf{U}) < \infty$*;*

and a sufficient statistic $\mathbf{T}$*, then* $\mathbf{f}(\mathbf{T}) = \mathbb{E}[\mathbf{U} \mid \mathbf{T}]$ *verifies*

$$\mathcal{R}_{\boldsymbol{\theta}}(\mathbf{U}) \geq \mathcal{R}_{\boldsymbol{\theta}}\big(\mathbf{f}(\mathbf{T})\big),$$

with the equality holding if and only if

$$\mathbb{P}[\mathbf{U} \equiv \mathbf{f}(\mathbf{T})] = 1.$$

Proof. See [10], pg. 48. □

The concept of information, in what statistics are concerned, as well as the amount of it the estimator carries can be interpreted using what it is called the *information matrix*:

Definition 26. *Let* $\mathbf{X} = (X_1, ..., X_n)'$ *be a sample vector belonging to the family* $\mathcal{P} = \{f_{\boldsymbol{\theta}}(\cdot) : \boldsymbol{\theta} \in \Omega\}$ *of densities, with* $\boldsymbol{\theta} = (\theta_1, ..., \theta_s)'$*, and let* $\mathbf{T}$ *be a statistic. The element* $I_{ij}(\boldsymbol{\theta})$ *of the information matrix* $\mathbf{I}(\boldsymbol{\theta})$ *is defined as*

$$I_{ij}(\boldsymbol{\theta}) = \mathbb{E}\left[\frac{\partial}{\partial\theta_i}\log\big(f_{\boldsymbol{\theta}}(\mathbf{X})\big) \cdot \frac{\partial}{\partial\theta_j}\log\big(f_{\boldsymbol{\theta}}(\mathbf{X})\big)\right].$$

This matrix measures the variation ratio of the density $f_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta}$, thus informing the change on the value of $\boldsymbol{\theta}$. It is then possible, with this concept, to obtain a bound for the variance of an estimator:

Theorem 35. *Let* $\mathbf{X}$ *be a sample vector belonging to a family* $\mathcal{P} = \{f_{\boldsymbol{\theta}}(\cdot) : \boldsymbol{\theta} \in \Omega\}$*, with* $\boldsymbol{\theta} = (\theta_1, ..., \theta_s)'$*, in which* Ω *is an open set and densities* $f_{\boldsymbol{\theta}}(\cdot)$ *have a common support, and* T *a statistic such that* $\mathbb{E}[T^2] < \infty$*. Let one of the following conditions hold:*

- $\frac{\partial}{\partial\theta_i}\mathbb{E}[T] = \mathbb{E}\left[\frac{\partial}{\partial\theta_i}T\right]$ *exists* $\forall i = 1, ..., s$*;*
- *There exist functions* $b_{\boldsymbol{\theta},i}(\cdot)$*,* $i = 1, ..., s$*, with* $\mathbb{E}\big[b^2_{\boldsymbol{\theta},i}(\mathbf{X})\big] < \infty$*, such that* $\forall\Delta > 0 : \left|\dfrac{f_{\boldsymbol{\theta}+\Delta\varepsilon_i}(x) - f_{\boldsymbol{\theta}}(x)}{\Delta}\right| < b_{\boldsymbol{\theta},i}(x)$*,* $i = 1, ..., s$*.*

Then $\mathbb{E}\left[\frac{\partial}{\partial\theta_i}\log\big(f_{\boldsymbol{\theta}}(\mathbf{X})\big)\right]$ *and*

$$\mathbb{V}[T] \geq \boldsymbol{\alpha}'\big(\mathbf{I}(\boldsymbol{\theta})\big)^{-1}\boldsymbol{\alpha},$$

with $\alpha_i = \frac{\partial}{\partial\theta_i}\mathbb{E}[T]$*,* $i = 1, ..., s$*.*

Proof. See [10], p. 127. □

The discussion will now be restricted to unidimensional parameters and unidimensional functions of the same parameter. The next step would be to find estimators with optimal properties. But in fact, the class of all estimators is too large and general to derive such an estimator, even with the adequate optimality criteria. To surround this problem, it is now introduced a criterion which restricts

the estimator's space and gives the estimators a desirable property: *unbiasedness*. Formally,

Definition 27. *An estimator T for a parameter $\theta \in \Omega$ is unbiased if*

$$\mathbb{E}[T] = \theta, \forall \theta \in \Omega.$$

Functions $g(\theta)$ and parameters θ for which there exist unbiased estimators are called *estimable* functions and parameters, respectively. Finding the best estimator in this class may be somewhat difficult, and characterizing all unbiased estimators is a very useful achievement. This is shown in

Lemma 2. *Let T be an unbiased estimator of $g(\theta)$. Every unbiased estimator U of $g(\theta)$ has the form*

$$U = T - T_0,$$

where T_0 is an unbiased estimator of 0.

Proof. It is quite simple, because

$$\mathbb{E}[U] = \mathbb{E}[T] - \mathbb{E}[T_0] = g(\theta). \tag{3.2}$$

□

Another interesting fact is the uniqueness of unbiased estimators given a complete sufficient statistic.

Lemma 3. *Let $g(\theta)$ be an estimable function and T a complete sufficient statistic. If f and h are functions such that*

$$\mathbb{E}[f(T)] = \mathbb{E}[h(T)] = g(\theta),$$

then

$$\mathbb{P}[f \equiv h] = 1.$$

Proof. See [10], pp. 87–88. □

Nextly, a widely used optimality criterion is defined: *uniformly minimum variance unbiased estimation*. The definition follows:

Definition 28 (UMVUE)**.** *An estimator T of $g(\theta)$ is an UMVUE (Uniformly Minimum Variance Unbiased Estimator) if*

1. $\mathbb{E}[T] = g(\theta)$;
2. $\mathbb{V}[T] \leq \mathbb{V}[U]$, *where U is an unbiased estimator of $g(\theta)$.*

The class of all unbiased estimators of 0 is an important class in what regards to UMVUEs. In fact this class distinguishes UMVUEs apart from other less effective estimators. This is brought by

Theorem 36. *Let T be an unbiased estimator of $g(\theta)$ such that $\mathbb{E}[T^2] < \infty$. T is an UMVUE of $g(\theta)$ if and only if*

$$\mathrm{Cov}[T, U] = 0, U \in \mathcal{U}$$

with $U \in \mathcal{U}$, the class of all unbiased estimators of 0.

Proof. See [10], pg. 86. □

It is also possible to refine the Rao-Blackwell theorem:

Theorem 37. *Given an unbiased estimator U of $g(\theta)$ with a loss function $\ell(u, \theta)$ strictly convex on u and such that*

- $\mathbb{E}[U] < \infty$;
- $\mathcal{R}_\theta(U) < \infty$;

and a sufficient statistic T, then $f(T) = \mathbb{E}[U \mid T]$ verifies

- $\mathbb{E}[f(T)] = g(\theta)$;
- $\mathcal{R}_\theta(U) \leq \mathcal{R}_\theta\big(f(T)\big)$;

with the equality holding if and only if

$$\mathbb{P}[U \equiv f(T)] = 1.$$

Proof. See [14], pg. 322. □

The relationship between completeness of statistics and UMVUEs helps to provide estimators with optimal properties as long as ancillary information is disposed of. This is the core of

Theorem 38. *Let* $\mathbf{X}$ *be a random vector with density* $f_{\boldsymbol{\theta}}$ *and let*

$$\mathbb{E}[S] = g(\boldsymbol{\theta}).$$

Let also $\mathbf{T}$ *be a complete sufficient statistic. Then there exists a function* f *such that* $f(\mathbf{T})$ *is an UMVUE of* $g(\boldsymbol{\theta})$*. Furthermore, this function is unique with probability 1.*

Proof. This theorem is adapted from theorem 1.11 in [10], taking

$$\ell_{\boldsymbol{\theta}}(u) = \big(u - g(\boldsymbol{\theta})\big)^2$$

and, consequently, the risk of S to be

$$\mathcal{R}_{\boldsymbol{\theta}}(S) = \mathbb{E}[\ell_{\boldsymbol{\theta}}(S)] = \mathbb{E}\Big[\big(S - g(\boldsymbol{\theta})\big)^2\Big].$$

When S is unbiased

$$\mathcal{R}_{\boldsymbol{\theta}}(S) = \mathbb{E}\Big[\big(S - \mathbb{E}[S]\big)^2\Big] = \mathbb{V}[S].$$

A more generalized version of this theorem can be found in [10], pg. 88, and it is the consequence of lemma 1.10 and the discussion in the precedent pages. □

3.2 Exponential Family of Densities

Up until now, inference on unknown parameters was made without any kind of restrictions on the family of distributions underlying in the model. Restricting the type of distributions may in fact narrow the scope of applications but it can lead to stronger and more useful results. This section will verse on a specific family of densities: the *exponential family* of densities.

Definition 29. *A density $f_{\boldsymbol{\theta}}(\cdot)$ belongs to the s-dimensional exponential family if it can be expressed in the form*

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = h(\mathbf{x}) \exp\left(\sum_{i=1}^{s} \eta_i T_i(\mathbf{x}) - a(\boldsymbol{\eta})\right).$$

The $\boldsymbol{\eta} = \boldsymbol{\eta}(\boldsymbol{\theta})$ are known as the *natural* parameters of the density. The natural parameter space will be denoted by Ξ.

Another consequence of the specification of this family of distributions is the easy derivation of the cumulants and, consequently, of the moments of a random variable of the exponential family.

Theorem 39. *Let $\mathbf{X}$ be a random vector with density belonging to the exponential family. Then, for each T_i, $i = 1, ..., s$, the moment generating function*

$$M_{T_i}(\mathbf{u}) = e^{a(\boldsymbol{\eta}+\mathbf{u})-a(\boldsymbol{\eta})}$$

and the cumulant generating function

$$K_{T_i}(\mathbf{u}) = a(\boldsymbol{\eta} + \mathbf{u}) - a(\boldsymbol{\eta})$$

are defined in some neighborhood of $\mathbf{0}$.

Proof. See [10], pgs. 27–28. □

Statistics $T_1, ..., T_s$ are in fact very important. It is easy to see that, due to the factorization criterion (theorem 31) these statistics will be *sufficient*. Another important fact is linear independence between the T_i, $i = 1, ..., s$, and linear independence between parameters η_i, $i = 1, ..., s$. Thus the following

Definition 30. *A density belonging to the exponential family is said to be of full rank if the T_i, $i = 1, ..., w$, are linearly independent, the η_i, $i = 1, ..., s$, are linearly independent and if Ξ contains a s-dimensional rectangle.*

It is clear that the T_i, $i = 1, ..., s$, are sufficient statistics but, in the exponential family case they are also minimal, as shown in

Theorem 40. *Let $\Xi \subseteq \mathbb{R}^s$ and $\mathbf{X}$ be a random vector belonging to the exponential family. If the density is of full rank or if the parameter space Ξ contains a set of points that generate $\mathbb{R}^s$ and do not belong to a proper affine subspace of $\mathbb{R}^s$, statistics $T_1, ..., T_s$ are minimal sufficient.*

Proof. See [10], pg. 39. □

Completeness is also achieved in this family of densities.

Theorem 41. *Let $\mathbf{X}$ be a random vector with density belonging to the exponential family with full rank. Then the statistics T_i, $i = 1, ..., s$, are complete.*

Proof. See [9], pgs. 142–143. □

An interesting property of this family of densities is connected with the information matrix of an estimator for the parameter vector.

Theorem 42. *Let $\mathbf{X}$ be a random vector with density belonging to the s-dimensional exponential family, and let also*

$$\mathbb{E}[T_i] = \tau_i, i = 1, ..., s.$$

Then $\mathbf{I}(\boldsymbol{\tau}) = (\mathbb{V}[\mathbf{T}])^{-1}$, with $\boldsymbol{\tau} = (\tau_1, ..., \tau_s)'$ and $\mathbf{T} = (T_1, ..., T_s)'$.

Proof. See [10], pg. 136. □

One of the most important results connected to the exponential family of distributions is that any unbiased estimator which is a function of the $T_1, ..., T_s$ will be an UMVUE.

Theorem 43 (Blackwell-Lehmann-Scheffé Theorem)**.** *Let X be a random variable with density belonging to the exponential family, of full rank and parameter $\boldsymbol{\theta}$. For any unbiasedly estimable function $g(\boldsymbol{\theta})$ there exists an UMVUE that is a function of $\mathbf{T}$. Furthermore, it is unique.*

Proof. The thesis is a direct consequence of theorem 38 and of theorems 40 and 41. □

3.3 Estimation on Linear Models

In this section linear models of various types will be discussed, along with the techniques for estimation of parameters, both fixed and random. The most common and important models are described, followed by fixed effects estimation theory, which is fairly established. Lastly, variance components estimation is covered.

3.3.1 Linear Models

A random variable described by a linear model has the structure

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}. \tag{3.3}$$

or, rewriting, it can be the sum of several terms:

$$\mathbf{y} = \sum_{i=1}^{w} \mathbf{X}_i \boldsymbol{\beta}_i. \tag{3.4}$$

Parameters $\boldsymbol{\beta}_i$ (or $\boldsymbol{\beta}$) are fixed unknown variables or random variables. It is on this last feature that come the first characterizations of linear models.

Definition 31. *A linear model*

$$\mathbf{y} = \sum_{i=1}^{b} \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e},$$

where matrices $\mathbf{X}_i$, $i = 1, \ldots, b$, *are known, vectors* $\boldsymbol{\beta}_i$, $i = 1, \ldots, b$, *are fixed and unknown and* $\mathbf{e}$ *is a random vector with null mean vector and covariance matrix* $\sigma^2\mathbf{V}$, *with* $\mathbf{V}$ *a known matrix and* σ^2 *unknown, is said to be a fixed effects model.*

This is the most widespread model in applications, both in analysis of variance as well in linear regression models.

In order to acquaint with random effects, one can consider models with parameters that are random variables, besides the usual error.

Definition 32. *A linear model*

$$\mathbf{y} = \mathbf{1}\mu + \sum_{i=1}^{w} \mathbf{e}_i,$$

where μ *is fixed and unknown and* $\mathbf{e}_1, \ldots, \mathbf{e}_w$ *are random with null mean vectors and covariance matrices* $\sigma_1^2\mathbf{V}_1, \ldots, \sigma_w^2\mathbf{V}_w$, *with* $\mathbf{V}_1, \ldots, \mathbf{V}_w$ *known, is said to be a random effects model.*

This model is usually applied to the analysis of variance of random effects.

Following the definition of these two models, there is, of course, the mixed model:

Definition 33. *A linear model*

$$\mathbf{y} = \sum_{i=1}^{b} \mathbf{X}_i \boldsymbol{\beta}_i + \sum_{i=b+1}^{w} \mathbf{e}_i,$$

with $b < w$, $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_b$ *fixed and unknown and* $\mathbf{e}_1, \ldots, \mathbf{e}_w$ *are random vectors with null mean is said to be a mixed effects model.*

These are the models in which this work will verse on and, in fact, both fixed and random effects models are particular cases of the mixed effects model.

3.3.2 Model Structure

A particular case of the fixed effects model is the

Definition 34 (Gauss-Markov Model)**.** *A fixed effects model*

$$\mathbf{y} = \sum_{i=1}^{w} \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}$$

with

- $\mathbb{E}[\mathbf{e}] = \mathbf{0}$*;*
- $\mathbb{V}[\mathbf{e}] = \sigma^2 \mathbf{I}$*;*

is said to be a Gauss-Markov model.

If independency of observations is dropped and only covariance ratio is known, one obtains the

Definition 35 (Generalized Gauss-Markov Model)**.** *A fixed effects model*

$$\mathbf{y} = \sum_{i=1}^{w} \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}$$

with

- $\mathbb{E}[\mathbf{e}] = \mathbf{0}$*;*
- $\mathbb{V}[\mathbf{e}] = \sigma^2 \mathbf{V}$*;*

is said to be a generalized Gauss-Markov model.

These are the models used in usual regression and controlled heteroscedasticity regression.

A very important class of models is the class of orthogonal models.

Definition 36. *A mixed model*

$$\mathbf{y} = \sum_{i=1}^{b} \mathbf{X}_i \boldsymbol{\beta}_i + \sum_{i=b+1}^{w} \mathbf{X}_i \boldsymbol{\beta}_i \tag{3.5}$$

is orthogonal when

$$\mathbf{M}_i \mathbf{M}_j = \mathbf{M}_j \mathbf{M}_i; i, j = 1, \ldots, w,$$

with $\mathbf{M}_i = \mathbf{X}_i \mathbf{X}_i'$, $i = 1, \ldots, w$, *where* $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_b$ *are fixed effects vectors and* $\boldsymbol{\beta}_{b+1}, \ldots, \boldsymbol{\beta}_w$ *are random effects vectors.*

Often, $\boldsymbol{\beta}_w$ is the technical error, having null mean and covariance structure $\sigma^2\mathbf{I}$. From this definition, more refined families of models can be obtained.

Definition 37. *A linear model has COBS (commutative orthogonal block structure) when*

$$\mathbb{V}[\mathbf{y}] = \sum_{i=1}^{w} \gamma_i \mathbf{Q}_i,$$

with $\{\mathbf{Q}_1, \ldots, \mathbf{Q}_w\}$ *a family of orthogonal projection matrices such that*

- $\mathbf{Q}_i \mathbf{Q}_j = \mathbf{0}$, $i \neq j$, $i, j = 1, \ldots, w$;
- $\sum_{i=1}^{w} \mathbf{Q}_i = \mathbf{I}$;
- *the set* Γ *of variance components* $\gamma_1, \ldots, \gamma_w$ *contains an open non empty set,*

and, with $\mathbf{T}$ *the orthogonal projection matrix on the space of the mean vector,*

$$\mathbf{T}\mathbf{Q}_i = \mathbf{Q}_i \mathbf{T}; i = 1, \ldots, w.$$

A more general family of models is obtained if the commutativity of $\mathbf{T}$ with $\mathbf{Q}_1, \ldots, \mathbf{Q}_w$ is dropped.

Definition 38. *A linear model has OBS (orthogonal block structure) when*

$$\mathbb{V}[\mathbf{y}] = \sum_{i=1}^{w} \gamma_i \mathbf{Q}_i,$$

with $\{\mathbf{Q}_1, \ldots, \mathbf{Q}_w\}$ *a family of orthogonal projection matrices such that*

- $\mathbf{Q}_i\mathbf{Q}_j = \mathbf{0}$, $i \neq j$, $i, j = 1, \ldots, w$;
- $\sum_{i=1}^{w} \mathbf{Q}_i = \mathbf{I}$;
- *the set* Γ *of variance components* $\gamma_1, \ldots, \gamma_w$ *contains an open non empty set.*

3.3.3 Fixed Effects

Estimation for fixed effects is rather well established. Considering linear estimators, one has the fundamental theorem:

Theorem 44. *Given a linear model* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$*, the following conditions are equivalent:*

1. *the function* $\boldsymbol{\psi} = \mathbf{a}'\boldsymbol{\beta}$ *is linearly estimable;*
2. $\exists \mathbf{c} : \mathbf{a} = \mathbf{X}'\mathbf{c}$;
3. $\exists \mathbf{d} : \mathbf{a} = \mathbf{X}'\mathbf{X}\mathbf{d}$;
4. $\widehat{\boldsymbol{\psi}} = \mathbf{a}'\widehat{\boldsymbol{\beta}}$*, where* $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$*, independent of the choice for* $(\mathbf{X}'\mathbf{X})^{-}$*, is an unbiased estimator.*

Proof.

1. $\Rightarrow$ *2.* It results from the definition of estimable function, and because there exists $\mathbf{c}$ such that

$$\mathbb{E}[\mathbf{c}'\mathbf{y}] = \mathbf{c}'\mathbf{X}\boldsymbol{\beta} = \mathbf{a}'\boldsymbol{\beta},$$

for all $\boldsymbol{\beta}$, and so

$$\exists \mathbf{c} : \mathbf{a} = \mathbf{X}'\mathbf{c}.$$

2. $\Rightarrow$ *3.* Let $\mathbf{v} \in \mathrm{N}(\mathbf{X}'\mathbf{X})$. Then

$$\begin{aligned} \mathbf{v} \in \mathrm{N}(\mathbf{X}'\mathbf{X}) &\Rightarrow \mathbf{X}'\mathbf{X}\mathbf{v} = \mathbf{0} \\ &\Rightarrow \mathbf{v}'\mathbf{X}'\mathbf{X}\mathbf{v} = \mathbf{0} \\ &\Rightarrow \mathbf{X}\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{v} \in \mathrm{N}(\mathbf{X})\,. \end{aligned}$$

Thus $\mathrm{N}(\mathbf{X}'\mathbf{X}) \subseteq \mathrm{N}(\mathbf{X})$ which implies $\mathrm{R}(\mathbf{X}) \subseteq \mathrm{R}(\mathbf{X}'\mathbf{X})$, and so

$$\exists \mathbf{d} : \mathbf{a} = \mathbf{X}'\mathbf{X}\mathbf{d}.$$

3. $\Rightarrow$ *4.* Let $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$. Then,

$$\begin{aligned}\mathbb{E}\left[\mathbf{a}\widehat{\boldsymbol{\beta}}\right] &= \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{d}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{d}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= \mathbf{a}'\boldsymbol{\beta}.\end{aligned}$$

4. $\Rightarrow$ *1.* It is enough to consider

$$\mathbf{c} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{+}\mathbf{a}$$

to have $\mathbb{E}[\mathbf{c}'\mathbf{y}] = \mathbf{a}'\boldsymbol{\beta} = \boldsymbol{\psi}$. □

From here, arises the

Theorem 45 (Gauss-Markov Theorem)**.** *Given the linear Gauss-Markov model* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$*, every linear estimable* $\boldsymbol{\psi} = \mathbf{a}'\boldsymbol{\beta}$ *function has the unique unbiased estimator* $\widehat{\boldsymbol{\psi}} = \mathbf{a}'\widehat{\boldsymbol{\beta}}$ *with minimum variance among all linear estimators, where* $\widehat{\boldsymbol{\beta}}$ *is the least squares estimate of* $\boldsymbol{\beta}$*.*

Of course this theorem is only valid for models where $\mathbb{V}[\mathbf{y}] = \sigma^2\mathbf{I}$ but, assuming that $\mathbb{V}[\mathbf{y}] = \sigma^2\mathbf{V}$, one takes a matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{V}\mathbf{P} = \mathbf{I}, \tag{3.6}$$

thus obtaining a Gauss-Markov model

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \mathbf{e}_*, \tag{3.7}$$

for which

$$\begin{cases} \mathbf{X}_* = \mathbf{P}'\mathbf{X} \\ \mathbb{V}[\mathbf{y}_*] = \sigma^2\mathbf{I} \end{cases}. \tag{3.8}$$

All the previous results are then applicable.

In [33], very general results were obtained in characterization and expression of best linear unbiased estimators. Consider the mixed model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $\mathbb{V}[\mathbf{y}] = \sum_{i=1}^{w} \sigma_i^2 \mathbf{V}_i$. Let $\mathcal{V}$ be the linear space spanned by $\{\mathbf{V}_1, \ldots, \mathbf{V}_w\}$ and $\mathbf{V}_0 \in \mathcal{V}$ such that

$$\mathrm{R}(\mathbf{V}_i) \subseteq \mathrm{R}(\mathbf{V}_0)\,, i = 1, \ldots, w. \tag{3.9}$$

Such a matrix always exists (see [8]). Define

$$\mathbf{W} = \begin{cases} \mathbf{V}_0 + \mathbf{X}\mathbf{X}', & \mathrm{R}(\mathbf{X}) \nsubseteq \mathcal{V}_0 \\ \mathbf{V}_0, & \mathrm{R}(\mathbf{X}) \subseteq \mathcal{V}_0 \end{cases} \tag{3.10}$$

Let also $\mathbf{P}$ be the orthogonal projection matrix on $\mathrm{R}(\mathbf{X})$ and $\mathbf{M} = \mathbf{I} - \mathbf{P}$, define

$$\boldsymbol{\Upsilon} = \mathbf{X}'\mathbf{W}^{-} \left(\sum_{i=1}^{w} \mathbf{V}_i \mathbf{M} \mathbf{V}_i \right) \mathbf{W}^{-}\mathbf{X} \tag{3.11}$$

and $\mathbf{Z} = \mathbf{I} - \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^{+}$

Theorem 46. *$T = \mathbf{c}'\mathbf{y}$ is a best linear unbiased estimator for $\mathbb{E}[\mathbf{c}'\mathbf{y}]$ if and only if*

$$\exists \mathbf{z} \in \mathrm{N}(\boldsymbol{\Upsilon}) : \mathbf{W}^{+}\mathbf{X}\mathbf{z} = \mathbf{c}.$$

Proof. See [33]. □

Existence of such optimal estimators is addressed in

Theorem 47. *There exists a best linear unbiased estimator for $\psi = \mathbf{a}'\boldsymbol{\beta}$ if and only if*

$$a \in \mathrm{R}(\mathbf{X}'\mathbf{W}\mathbf{X}\mathbf{Z})\,.$$

Proof. See [33]. □

The expression for these estimators is given by

Theorem 48. *If there exists a best linear unbiased estimator for $\psi = \mathbf{a}'\boldsymbol{\beta}$, it has the form $\hat{\psi} = \mathbf{a}'\widehat{\boldsymbol{\beta}}$, with*

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{Z}'\mathbf{X}'\mathbf{W}^{-}\mathbf{y}.$$

Proof. See [33]. □

3.3.4 Variance Components

The previous results cover estimation for fixed effects. Nextly, techniques for the estimation of variance components will be presented. Consider the linear model $\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \sum_{i=1}^{w}\mathbf{X}_i\boldsymbol{\beta}_i$. Then

$$\begin{cases} \mathbb{E}[\mathbf{y}] = \mathbf{X}_0\boldsymbol{\beta}_0 \\ \mathbb{V}[\mathbf{y}] = \sum_{i=1}^{w}\sigma_i^2\mathbf{M}_i \end{cases}, \tag{3.12}$$

were $\mathbf{M}_i = \mathbf{X}_i\mathbf{X}_i'$.

Take $\mathbf{M} = \mathbf{I} - \mathbf{P}$, with $\mathbf{P}$ the orthogonal projection matrix on $\mathrm{R}(\mathbf{X})$. Two important theorems describe the expectation and variance of quadratic forms under normality.

Theorem 49. *Let $\mathbf{y}$ be a normal vector with*

$$\begin{cases} \mathbb{E}[\mathbf{y}] = \boldsymbol{\mu} \\ \mathbb{V}[\mathbf{y}] = \mathbf{V} \end{cases}.$$

Then, given a symmetric matrix $\mathbf{A}$,

$$\begin{cases} \mathbb{E}[\mathbf{y}'\mathbf{A}\mathbf{y}] = \mathrm{tr}(\mathbf{A}\mathbf{V}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} \\ \mathbb{V}[\mathbf{y}'\mathbf{A}\mathbf{y}] = 2\mathrm{tr}((\mathbf{A}\mathbf{V})^2) + 4\boldsymbol{\mu}'\mathbf{A}\mathbf{V}\mathbf{A}\boldsymbol{\mu} \end{cases}.$$

Proof. See [18], pgs. 395–396. □

Choose now a set of quadratic forms $\mathbf{A}_1, \ldots, \mathbf{A}_w$ such that

$$\mathbf{X}_0'\mathbf{A}_i\mathbf{X}_0 = \mathbf{0}; i = 1, \ldots, w, \tag{3.13}$$

and so, with

- $\mathbf{s} = (\mathbf{y}'\mathbf{A}_1\mathbf{y}, \ldots, \mathbf{y}'\mathbf{A}_w\mathbf{y})'$;
- $\mathbf{X}'\mathbf{A}_i\mathbf{X}$, $i = 1, \ldots, w$;
- $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_w^2)'$;
- $\mathbf{C} = [\mathrm{tr}\big(\mathbf{X}_j'\mathbf{A}_i\mathbf{X}_j\big)]$,

one gets

$$\mathbb{E}[\mathbf{s}] = \mathbf{C}\boldsymbol{\sigma}^2, \tag{3.14}$$

a set of unbiased estimators

$$\widehat{\boldsymbol{\sigma}}^2 = \mathbf{C}^{-1}\mathbf{s}, \tag{3.15}$$

given that $\mathbf{C}$ is regular. This is called the *generalized ANOVA method*.

In the case of models with OBS, the quadratic forms

$$\mathbf{M}\mathbf{Q}_1\mathbf{M}, \ldots, \mathbf{M}\mathbf{Q}_w\mathbf{M}, \tag{3.16}$$

if linearly independent, and

$$\mathbf{C} = \mathrm{diag}(\mathrm{rank}(\mathbf{Q}_1), \ldots, \mathrm{rank}(\mathbf{Q}_w)) \tag{3.17}$$

guaranty a set of unbiased estimators for $\boldsymbol{\sigma}^2$.

Up until now, general methods for fixed effects and variance components have been presented. But assuming normality it is possible to obtain optimality conditions.

Theorem 50. *Consider the linear model*

$$\mathbf{y} = \mathbf{X}_0\beta_0 + \sum_{i=1}^{w} \mathbf{X}_i\boldsymbol{\beta}_i,$$

with $\boldsymbol{\beta}_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2\mathbf{I})$. *Let* $\mathcal{V}_0$ *be the set spanned by*

$$\left\{ \left(\sum_{i=1}^{w} \sigma_i^2\mathbf{M}_i \right)^{-1} : \sigma_1^2, \ldots, \sigma_w^2 \in \mathbb{R}^+ \right\}.$$

Let $\{\mathbf{W}_1, \ldots, \mathbf{W}_v\}$ *be a basis for* $\mathcal{V}_0$ *and* $\mathbf{x}_1, \ldots, \mathbf{x}_p$ *a basis for* $\mathrm{R}(\mathbf{X}_0)$. *If* $\mathbf{I} \in \mathcal{V}_0$ *then* $\mathbf{x}_1, \ldots, \mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_m$ *is a basis for the set*

$$\{\mathbf{W}\mathbf{x} : \mathbf{x} \in \mathrm{R}(\mathbf{X}_0)\,; \mathbf{W} \in \mathcal{V}_0\}.$$

and

$$\mathbf{x}_1'\mathbf{y}, \ldots, \mathbf{x}_p'\mathbf{y}, \mathbf{x}_{p+1}'\mathbf{y}, \ldots, \mathbf{x}_m'\mathbf{y}, \mathbf{y}'\mathbf{W}_1\mathbf{y}, \ldots, \mathbf{y}'\mathbf{W}_v\mathbf{y}$$

form a set of minimal sufficient statistics.

Proof. See [34]. □

Now let $\mathbf{N}_0 = \mathbf{I} - \mathbf{P}_0$ and $\mathcal{V}$ the linear space of matrices spanned by

$$\left\{ \left(\mathbf{N}_0 \left(\sum_{i=1}^{w} \sigma_i^2 \mathbf{M}_i \right) \mathbf{N}_0 \right)^+ : \sigma_1^2, \ldots, \sigma_w^2 \in \mathbb{R}^+ \right\}. \tag{3.18}$$

Then,

Theorem 51. *Let* $\{\mathbf{U}_1, \ldots, \mathbf{U}_u\}$ *be a basis for* $\mathcal{V}$. *Then*

$$\mathbf{A}_0'\mathbf{y}, \mathbf{y}'\mathbf{U}_1\mathbf{y}, \ldots, \mathbf{y}'\mathbf{U}_u\mathbf{y},$$

form a set of minimal sufficient statistics, where $\mathbf{A}_0\mathbf{A}_0' = \mathbf{P}_0$ *while* $\mathbf{P}_0$ *is the orthogonal projection matrix on* $\mathrm{R}(\mathbf{X}_0)$.

Proof. See [34]. □

Very important is the fact that completeness is also achieved.

Theorem 52. *If*

$$\mathbf{P}_0\mathbf{M}_i = \mathbf{M}_i\mathbf{P}_0; i = 1, \ldots, w$$

and $\mathcal{V}$ *is a Jordan algebra of symmetric matrices,*

$$\mathbf{A}_0'\mathbf{y}, \mathbf{y}'\mathbf{U}_1\mathbf{y}, \ldots, \mathbf{y}'\mathbf{U}_u\mathbf{y}$$

form a set of complete minimal sufficient statistics.

Proof. See [34]. □

Equivalent results were obtained by Justus Seely in [22], [23] and in [25].

But a stronger relationship between commutative Jordan algebras and models can be established. In fact, a connection between any orthogonal model and an orthogonal partition

$$\mathbb{R}^n = \boxplus_{i=1}^{w+1} \nabla_i, \tag{3.19}$$

where $\boxplus$ stands for the direct sum of orthogonal linear subspaces, can be established.

Let the orthogonal projection matrices on the subspaces ∇_i, ..., ∇_{w+1} be

$$\mathbf{Q}_1, \ldots, \mathbf{Q}_{w+1}.$$

Then $g_i = \text{rank}(\mathbf{Q}_i) = \dim(\nabla_i)$, $i = 1, ..., w+1$, as well as $\sum_{i=1}^{w+1} g_i = n$ and as $\sum_{i=1}^{w+1} \mathbf{Q}_i = \mathbf{I}_n$. If the columns of $\mathbf{A}_i$ constitute an orthonormal basis for ∇_i, $\mathbf{Q}_i = \mathbf{A}_i\mathbf{A}_i'$ as well as $\mathbf{A}_i'\mathbf{A}_i = \mathbf{I}_{g_i}$, $i = 1, ..., w+1$. Moreover, $\mathbf{A}_i'\mathbf{A}_j = \mathbf{0}_{g_i,g_j}$ and $\mathbf{A}_i'\mathbf{Q}_j = \mathbf{0}_{g_i,n}$, whenever $i \neq j$. Establish now

Theorem 53. *A normal orthogonal model associated with the orthogonal partition in* (3.19) (*as with the corresponding Jordan commutative algebra*) *has the canonical form*

$$\mathbf{Y} = \sum_{i=1}^{w+1} \mathbf{A}_i\boldsymbol{\eta}_i,$$

where vectors $\boldsymbol{\eta}_i$, $i = 1, ..., w+1$, *are normal, independent, with mean vectors* $\boldsymbol{\lambda}_i$, $i = 1, ..., w+1$, *and covariance matrices* $\gamma_i\mathbf{I}_{g_i}$, $i = 1, ..., w+1$.

Proof. Let $\mathbf{Q}^*$ be the orthogonal projection matrix in the sub-space that contains the observations mean vector $\boldsymbol{\mu}$. Since $\mathbf{Q}^*$ belongs to the algebra, $\mathbf{Q}^* = \sum_{i=1}^{w+1} c_i\mathbf{Q}_i$, with $c_i = 0$ or $c_i = 1$, $i = 1, ..., w+1$. One can assume without loss of generality that $\mathbf{Q}^* = \sum_{i=1}^{m} \mathbf{Q}_i$. Thus,

$$\boldsymbol{\mu} = \mathbf{Q}^*\boldsymbol{\mu} = \sum_{i=1}^{m} \mathbf{Q}_i\boldsymbol{\mu} = \sum_{i=1}^{m} \mathbf{A}_i\mathbf{A}_i'\boldsymbol{\mu} = \sum_{i=1}^{m} \mathbf{A}_i\boldsymbol{\lambda}_i,$$

where $\boldsymbol{\lambda}_i$ is the mean vector of $\boldsymbol{\eta}_i = \mathbf{A}_i'\mathbf{Y}$, $i = 1, ..., m$. Moreover, $\boldsymbol{\lambda}_i = \mathbf{0}$, $i = m+1, ..., w+1$, will be the mean vector of $\boldsymbol{\eta}_i = \mathbf{A}_i'\mathbf{Y}$, $i = 1, ..., m$. Then

$$\mathbf{Y} = \mathbf{I}\mathbf{Y} = \sum_{i=1}^{w+1} \mathbf{Q}_i\mathbf{Y} = \sum_{i=1}^{w+1} \mathbf{A}_i\mathbf{A}_i'\mathbf{Y} = \sum_{i=1}^{w+1} \mathbf{A}_i\boldsymbol{\eta}_i.$$

To complete the proof it is necessary only to point out that the $\boldsymbol{\eta}_i$, $i = 1, ..., w+1$, are normal and independent because, as it is easily seen, their cross-covariance matrices are null. Moreover, their covariance matrices are $\gamma_i\mathbf{I}_{g_i}$, $i = 1, ..., w+1$. □

From the previous proof it is clear that

$$\boldsymbol{\mu} = \sum_{i=1}^{m} \mathbf{A}_i\boldsymbol{\lambda}_i. \tag{3.20}$$

Moreover, since the $\boldsymbol{\eta}_i$ are independent, the covariance matrix of $\mathbf{Y}$ is

$$\mathbf{V} = \sum_{i=1}^{w+1} \gamma_i\mathbf{Q}_i. \tag{3.21}$$

Thus (see [3]),

$$\begin{cases} \det(\mathbf{V}) = \prod\limits_{i=1}^{w+1} \gamma_i^{g_i}, \\ \mathbf{V}^{-1} = \sum\limits_{i=1}^{w+1} \gamma_i^{-1}\mathbf{Q}_i. \end{cases} \tag{3.22}$$

The following result is of evident importance because it shows that, for every $\boldsymbol{\mu}$ and $\mathbf{V}$, the canonical parameters $\boldsymbol{\lambda}_i$, $i = 1, ..., m$, and γ_i, $i = 1, ..., w+1$, are unique.

Theorem 54. *The equality* $\sum_{i=1}^{m} \mathbf{A}_i\mathbf{a}_i = \sum_{i=1}^{m} \mathbf{A}_i\mathbf{b}_i$ *holds if and only if* $\mathbf{a}_i = \mathbf{b}_i$, $i = 1, ..., m$, *and* $\sum_{i=m+1}^{w+1} u_i\mathbf{Q}_i = \sum_{i=m+1}^{w+1} v_i\mathbf{Q}_i$ *when and only when* $u_i = v_i$, $i = m+1, ..., w+1$.

Proof. For either part of the thesis it is sufficient to establish the necessary condition, because the corresponding sufficient condition is self-evident. Starting with the first part, since

$$\nabla_i \bigcap \boxplus_{j \neq i} \nabla_j = \{\mathbf{0}\},$$

if $\sum_{j=1}^{w+1} \mathbf{A}_i\mathbf{a}_i = \sum_{j=1}^{w+1} \mathbf{A}_i\mathbf{b}_i$, *i.e.*, if

$$\mathbf{A}_i(\mathbf{b}_i - \mathbf{a}_i) = -\sum_{j\neq i} \mathbf{A}_j(\mathbf{b}_j - \mathbf{a}_j) \in \nabla_i \bigcap \boxplus_{j\neq i} \nabla_j,$$

then $\mathbf{A}_i(\mathbf{b}_i - \mathbf{a}_i) = \mathbf{0}$ as well as

$$\mathbf{b}_i - \mathbf{a}_i = \mathbf{A}_i'\mathbf{A}_i(\mathbf{b}_i - \mathbf{a}_i) = \mathbf{0},$$

$i = 1, ..., m$, so the first part is established. Moreover, if $\sum_{i=m+1}^{w+1} u_i\mathbf{Q}_i = \sum_{i=m+1}^{w+1} v_i\mathbf{Q}_i$, then

$$u_i\mathbf{Q}_i = \mathbf{Q}_i\left(\sum_{j=m+1}^{w+1} u_j\mathbf{Q}_j\right) = \mathbf{Q}_i\left(\sum_{j=m+1}^{w+1} v_j\mathbf{Q}_j\right) = v_i\mathbf{Q}_i,$$

thus $u_i = v_i$, $i = m+1, ..., w+1$, and the proof is complete. □

The result that follows plays a central part in the inference. This result is an alternative to the well known result by Zmyślony (see [33]) and Seely (see [25]).

Theorem 55. *The observations vector*

$$\mathbf{Y} = \sum_{j=1}^{w+1} \mathbf{A}_j\boldsymbol{\eta}_j,$$

where the random vectors $\boldsymbol{\eta}_j$ are independent, normal, with the mean vectors $\boldsymbol{\lambda}_j$, $j = 1, ..., m$, $\mathbf{0}$, $j = m+1, ..., w$, and the covariance matrices $\gamma_j\mathbf{Q}_j$, $j = 1, ..., w$, where $(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_m, \gamma_{m+1}, \ldots, \gamma_w)$ belongs to a set that has a non empty open subset, has the density

$$n(\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}) = \frac{\exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{m} \frac{\|\tilde{\boldsymbol{\eta}}_i - \boldsymbol{\lambda}_i\|^2}{\gamma_i} + \sum_{i=m+1}^{w+1} \frac{s_i}{\gamma_i}\right)\right\}}{\sqrt{(2\pi)^n \prod_{i=1}^{w+1} \gamma_i^{g_i}}}, \tag{3.23}$$

with complete sufficient statistics $\tilde{\boldsymbol{\eta}}_j = \mathbf{A}_j'\mathbf{y}$, $j = 1, ...m$, and $s_j = \|\mathbf{A}_j'\mathbf{y}\|^2$, $j = m+1, ..., w$.

Proof. As seen previously, $\mathbf{A}_i'\mathbf{Y} = \boldsymbol{\eta}_i$, $i = 1, ..., w+1$, $\mathbf{A}_i'\boldsymbol{\mu} = \boldsymbol{\lambda}_i$, $i = 1, ..., m$, and $\mathbf{A}_i'\boldsymbol{\mu} = \mathbf{0}$, $i = m+1, ..., w+1$. Also,

$$\mathbf{A}_i'\mathbf{V}^{-1}\mathbf{A}_i = \mathbf{A}_i'\left(\sum_{j=1}^{w+1}\gamma_j^{-1}\mathbf{Q}_j\right)\mathbf{A}_i = \gamma_i^{-1}\mathbf{A}_i'\mathbf{A}_i' = \gamma_i^{-1}\mathbf{I}_{g_i}; i = 1, ..., w+1 \tag{3.24}$$

so that

$$\begin{aligned}(\mathbf{y}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu}) &= \sum_{i=1}^{w+1}\frac{1}{\gamma_i}(\mathbf{y}-\boldsymbol{\mu})'\mathbf{A}_i\mathbf{A}_i'(\mathbf{y}-\boldsymbol{\mu}) \\ &= \sum_{i=1}^{m}\frac{\|\tilde{\boldsymbol{\eta}}_i - \boldsymbol{\lambda}_i\|^2}{\gamma_i} + \sum_{i=m+1}^{w+1}\frac{s_i}{\gamma_i}, \end{aligned} \tag{3.25}$$

where $\tilde{\boldsymbol{\eta}}_i = \mathbf{A}_i'\mathbf{y}$ and $s_i = \|\mathbf{A}_i'\mathbf{y}\|^2$. Thus, the model's density will be (3.23) and (see [27], pg. 31 and 32) one has the set of complete sufficient statistics $\tilde{\boldsymbol{\eta}}_i$, $i = 1, ..., m$, and s_i, $i = m+1, ..., w+1$. □

According to Theorem 43, the $\tilde{\boldsymbol{\eta}}_i$, $i = 1, ..., m$, and the $\tilde{\gamma}_i = s_i/g_i$, $i = m+1, ..., w+1$ are UMVUE for the mean vectors $\boldsymbol{\lambda}_i$, $i = 1, ..., m$, and the variance components γ_i, $i = m+1, ..., w+1$. To avoid over-parametrization, assume that

$$\gamma_i = \sum_{j=m+1}^{w+1} b_{i,j}\gamma_j, i = 1, ..., m, \tag{3.26}$$

so that the UMVUE are

$$\tilde{\gamma}_i = \sum_{j=m+1}^{w+1} b_{i,j}\tilde{\gamma}_j, i = 1, ..., m. \tag{3.27}$$

The estimable vectors will be of the form $\boldsymbol{\psi}_i = \mathbf{B}_i\boldsymbol{\lambda}_i$, $i = 1, ...m$, for which there are the UMVUE $\tilde{\boldsymbol{\psi}}_i = \mathbf{B}_i\tilde{\boldsymbol{\eta}}_i$, $i = 1, ...m$.

The joint distribution of the $\mathbf{A}_i'\mathbf{Y}$, $i = 1, ..., w+1$, is normal and, because their cross covariance matrices are null, they will be independent. Thus the $\tilde{\boldsymbol{\eta}}_i = \mathbf{A}_i'\mathbf{Y}$, $i = 1, ..., m$, and the $\tilde{\gamma}_i = \frac{1}{g_i}\|\mathbf{A}_i'\mathbf{Y}\|^2$, $i = m+1, .., w+1$, will be independent. Moreover, the $\tilde{\boldsymbol{\eta}}_i = \mathbf{A}_i'\mathbf{Y}$ and the $\tilde{\gamma}_i = \frac{1}{g_i}\|\mathbf{A}_i'\mathbf{Y}\|^2$, $i = 1, ..., m$, will also be independent, as

well as the $\tilde{\boldsymbol{\psi}}_i$ and the $\tilde{\gamma}_i$, $i = 1, ..., m$, where $\boldsymbol{\psi}_i = \mathbf{B}_i\boldsymbol{\lambda}_i$ is an estimable vector, $i = 1, ..., m$. It may be interesting to point out that taking $\mathbf{B}_i = \mathbf{I}_{g_i}$, so that $\boldsymbol{\lambda}_i$ is itself an estimable vector. If $\text{rank}(\mathbf{B}_i) = r_i$, $\mathbf{B}_i\mathbf{B}_i'$ will be positive definite and $\boldsymbol{\psi}_i = \mathbf{B}_i\boldsymbol{\lambda}_i$ will be a regular estimable vector.

4. HYPOTHESIS TESTING

In this chapter, the testing of statistical hypothesis is covered. Starting with general considerations of hypothesis testing, more specific hypothesis on fixed effects and variance components of linear mixed models will also be considered.

4.1 Hypothesis Tests

An hypothesis test is, basically, a decision to be taken or choice to be made based on (often scarce) known information. If one chooses the parametric approach that decision is made with information about an unknown parameter $\boldsymbol{\theta}$ belonging to a known set Θ, unknown, and it has, like point estimation, a loss function associated to it.

Consider now a sample $\mathbf{X} = (X_1, \ldots, X_n)'$ of independent and identically distributed random variables belonging to the class of densities

$$\mathcal{P} = \{f_{\boldsymbol{\theta}}(\cdot) : \boldsymbol{\theta} \in \Theta\}.$$

From here, it is possible to define:

Decision Rule $d = \phi(\mathbf{x})$;

Loss Function $\ell\big(\phi(\mathbf{x})\big), \boldsymbol{\theta}\big)$;

Risk Function $\mathcal{R}_{\boldsymbol{\theta}}\big(\phi(\mathbf{X})\big) = \mathbb{E}[\ell(\phi(\mathbf{X}), \boldsymbol{\theta})]$.

The two last functions are analogous to the functions defined for estimators.

Hypothesis tests are, as seen above, rules created to take a decision which has a loss and a risk associated to it. In mathematical terms, one can divide hypothesis tests in three main classes:

1. $\boldsymbol{\theta} \in \Theta_0$ or $\boldsymbol{\theta} \in \Theta_1$, $\Theta_0, \Theta_1 \subset \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$;
2. $\boldsymbol{\theta}$ belongs to one of the sets $\Theta_0, \ldots, \Theta_k$, $\Theta_0, \ldots, \Theta_k \subset \Theta$, $\Theta_i \cap \Theta_j = \emptyset$;
3. $\boldsymbol{\theta}$ as an associated loss function $\ell\big(\phi(\mathbf{x}), \boldsymbol{\theta}\big)$ and $\phi(\mathbf{x})$ is a real valued function.

The last type of hypothesis tests corresponds in fact to point estimation. Throughout this work the main interest will be focused on the first type of hypothesis tests. From this point on, the possible choice

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \tag{4.1}$$

shall be denoted the *null hypothesis* and the other possible choice

$$H_1 : \boldsymbol{\theta} \in \Theta_1 \tag{4.2}$$

shall be the *alternative hypothesis*. $\phi(\mathbf{x})$ takes the value 0 for the acceptance of the null hypothesis and the value 1 for the rejection of H_0 in favor of H_1.

Risk functions associated to decision rules, like the ones associated to point estimators, do not possess a strict ordering.

It is then, as well, impossible to choose the unambiguously better one and therefore impossible to choose the best decision rule. A solution to this problem may be, also like in point estimation, to restrict the class of decision rules in such a way that they retain desirable properties. Two of such properties are *unbiasedness* and *invariance*, which will be covered further on.

Like in any decision process, error can and will occur in hypothesis testing. In an hypothesis test with null and alternative hypothesis there are two types of error:

Type I Error Rejecting H_0 when H_0 is true;

Type II Error Accepting H_0 when H_1 is true.

Associated to an hypothesis test is the *acceptance region* and the *critical region.*

Definition 39. *The set*

$$S_0 = \{\mathbf{x} : \phi(\mathbf{x}) = 0\}$$

is called the acceptance region.

Definition 40. *The set*

$$S_1 = \{\mathbf{x} : \phi(\mathbf{x}) = 1\}$$

is called the critical region.

With the previous definitions, it is possible to now define the size of a test.

Definition 41. *The size of a test is given by*

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \mathbb{P}[\mathbf{X} \in S_1] .$$

The size of a test is in fact the maximum value the type I error can take. It is usual to restrict attention to tests such that

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \mathbb{P}[\mathbf{X} \in S_1] \leq \alpha. \tag{4.3}$$

Such a value is called the *level of significance* of an hypothesis test. Usual values for such levels of significance are 0.1, 0.05 and 0.01, although these values are still used because of an historical habit than by any other reason. Taking the converse approach, it is possible to obtain the smallest significance value for which the hypothesis would be rejected – this is called the *p-value.*

The previous discussion has covered type I error control, but another very important and sometimes neglected parameter is the *power* and the correspondent *power function.* Firstly, define

Definition 42. *The power of a test for any* $\boldsymbol{\theta} \in \Theta_1$ *is given by*

$$\beta = \mathbb{P}[\mathbf{X} \in S_1]\,.$$

Consequently,

Definition 43. *The power function of a test* ϕ *is given by*

$$\beta_\phi(\boldsymbol{\theta}) = \mathbb{P}[\mathbf{X} \in S_1]\,, \boldsymbol{\theta} \in \Theta.$$

The power function describes the efficiency of a test in detecting the departure of the data from the null hypothesis and, therefore, maximized.

The problem of hypothesis testing can generally formulated as, taking

$$\phi(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in S_0 \\ 1, & \mathbf{x} \in S_1 \end{cases}, \tag{4.4}$$

to maximize

$$\mathbb{E}[\phi(\mathbf{x})] = \beta_\phi(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \Theta_1, \tag{4.5}$$

subject to

$$\mathbb{E}[\phi(\mathbf{x})] \leq \alpha, \forall \boldsymbol{\theta} \in \Theta_0. \tag{4.6}$$

When Θ_0 and Θ_1 has only one element, the solution is given by the

Theorem 56 (Neyman-Pearson Fundamental Lemma)**.** *Let* $\Theta_0 = \{\boldsymbol{\theta}_0\}$ *and* $\Theta_1 = \{\boldsymbol{\theta}_1\}$*. There exists a constant* c *such that*

1. $\phi(\mathbf{x}) = \begin{cases} 0, & \dfrac{f_{\boldsymbol{\theta}_1}(\mathbf{x})}{f_{\boldsymbol{\theta}_0}(\mathbf{x})} < c \\ 1, & \dfrac{f_{\boldsymbol{\theta}_1}(\mathbf{x})}{f_{\boldsymbol{\theta}_0}(\mathbf{x})} > c \end{cases}$;
2. $\boldsymbol{\theta} = \boldsymbol{\theta}_0 \Rightarrow \mathbb{E}[\phi(\mathbf{X})] = \alpha$;

and ϕ *is the most powerful test for* $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ *vs.* $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ *at significance level* α*. Furthermore, if* ϕ_* *is the most powerful test then if it satisfies 2., it also satisfies 1. with probability 1.*

Proof. See [10], pgs. 74–76. □

There is also a relation between the power of the test and α.

Corollary 2. *Let β denote the power of the most powerful test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ vs. $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ at significance level α. Then, $\alpha < \beta$ unless $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$.*

Proof. See [10], pg. 76. □

After the two previous results, it is necessary to extend the definition of *most powerful test.*

Definition 44. *A test $\phi(\mathbf{x})$ with size α that maximizes $\beta_\phi(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta_1$ is denominated a uniformly most powerful test (UMP).*

Sufficiency plays again a fundamental role in hypothesis testing. This is shown in the following theorem:

Theorem 57. *Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be a sample from a family of densities $\mathcal{P}$ with parameter $\boldsymbol{\theta} \in \Theta$ and let T be a sufficient statistic for $\boldsymbol{\theta}$. For every test ϕ there exists a test ϕ_0, function of T, such that*

$$\beta_\phi(\boldsymbol{\theta}) = \beta_{\phi_0}(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \Theta.$$

Proof. See [14], pg. 408. □

With this definition in mind, it is important to refer special case, when $\boldsymbol{\theta} = \theta$ is a real number and the hypothesis to test is $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$. Consider now

Definition 45. *A density f_θ is said to have monotone likelihood ratio in $t(x)$ if, for $\theta < \theta_*$,*

$$\frac{f_{\theta_*}(x)}{f_\theta(x)}$$

is a non-decreasing for a function $t(x)$.

When the density has monotone likelihood ratio, it is possible to derive a UMP test. This is proven in the next theorem.

Theorem 58. *Let $\mathcal{P}$ be a family of densities with monotone likelihood ratio in $t(x)$. Considering the hypothesis $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$, there exists a test*

$$\phi(x) = \begin{cases} 1, & t(x) > c \\ 0, & t(x) < c \end{cases},$$

with

$$\theta = \theta_0 \Rightarrow \mathbb{E}[\phi(x)] = \beta_\phi(\theta_0) = \alpha,$$

$0 < \beta_\phi(\theta) < 1$ *being an increasing function of θ. Furthermore, the test minimizes the type I error for $\theta \in \Theta_0$,* i. e., *$\theta \leq \theta_0$.*

Proof. See [9], pg. 79. □

This is in fact the class of hypothesis that usual F tests belong to. They are, in fact, UMP tests for the most usual hypothesis on fixed effects and variance components.

For more complex hypothesis, namely hypothesis of the type

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \text{ vs. } H_1 : \theta < \theta_1 \vee \theta_2 < \theta, \tag{4.7}$$

with $\theta_1 < \theta_2$. For the one-dimensional exponential family there exist UMP tests, as proven in

Theorem 59. *Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be a sample of a random variable of the exponential family with parameter $\theta \in \mathbb{R}$ and $T(\mathbf{X}) = T$ its sufficient statistic. Then the test*

$$\phi(t) = \begin{cases} 1, & c_1 < t < c_2 \\ 0, & t < c_1 \text{ or } c_2 < t \end{cases},$$

with

$$\theta = \theta_i, i = 1, 2 \Rightarrow \mathbb{E}[\phi(T)] = \alpha,$$

is a UMP test for the hypothesis $H_0 : \theta_1 \leq \theta \leq \theta_2$ *vs.* $H_1 : \theta < \theta_1 \vee \theta_2 < \theta$. *Furthermore, there exists* $\theta_1 < \theta_0 < \theta_2$ *such that* $\beta\phi(\theta_0) \leq \beta_\phi(\theta)$ *for all* $\theta \in \Theta$.

Proof. See [9], pgs. 102–103. □

Like mentioned earlier, it is usual to restrict the class of tests to a "smaller" class possessing some interesting or convenient property. One of such properties is *unbiasedness*.

Definition 46. *A test* ϕ *such that*

$$\beta_\phi(\boldsymbol{\theta}) \leq \alpha, \forall \boldsymbol{\theta} \in \Theta_0$$

and

$$\beta_\phi(\boldsymbol{\theta}) \geq \alpha, \forall \boldsymbol{\theta} \in \Theta_1$$

is denominated an unbiased test.

This is to say that the power of the test for $\boldsymbol{\theta} \in \Theta_1$ is always bigger than the type I error. When the power test of a function is continuous,

$$\beta_\phi(\boldsymbol{\theta}) = \alpha, \tag{4.8}$$

for all $\boldsymbol{\theta} \in B$, where B is the common boundary of Θ_0 and Θ_1. From this arises the definition of

Definition 47. *A test* ϕ *such that*

$$\beta_\phi(\boldsymbol{\theta}) = \alpha, \forall \boldsymbol{\theta} \in B,$$

with B *the common boundary of* Θ_0 *and* Θ_1*, is called a similar test.*

Nextly, a theorem that makes the bridge between UMP tests, unbiasedness and similarity is presented.

Theorem 60. *Let $\mathcal{P}$ be a family of densities such that for every test, the power function is continuous. If, in the class of similar tests, ϕ is a UMP unbiased (UMPU) test, ϕ is UMP.*

Proof. See [9], pg. 135. □

As far as two sided hypothesis on one parameter exponential families go, the matter is discussed in [9], pgs. 135–137. The extension to multi-parameter is discussed in pgs. 145–151.

One very important property is invariance. In a simplistic description, it guarantees that the test to be used for an hypothesis remains unchanged for a class of transformations, namely bijections, on the parameter space. To make things more precise, consider

Definition 48. *An hypothesis H_0 vs. H_1 is said to be invariant for a bijection $g : \Theta \to \Theta$ if*

$$h(\Theta_0) = \Theta_0.$$

This definition can be extended to a group $\mathscr{H}$ of such transformations like, for example, linear transformations.

Definition 49. *An hypothesis H_0 vs. H_1 is said to be invariant for a group of bijections $\mathscr{H}$ if*

$$h(\Theta_0) = \Theta_0, \forall h \in \mathscr{H}.$$

It is now important to refer that there must exist a one to one correspondence between the bijections in the parameter space and the bijections in the sample space, because a transformation in the sample causes a change in the sample density.

Definition 50. *A bijection $h \in \mathscr{H}$ corresponds to a bijection $g \in \mathscr{G}$ if and only if*

$$\mathbb{P}[\mathbf{x} \in A \mid \boldsymbol{\theta}] = \mathbb{P}[\mathbf{x} \in h(A) \mid g(\boldsymbol{\theta})]\,.$$

It is clear that only transformations that change solely the parameter of the density are considered. More is explained in

Theorem 61. *Let $h_i \in \mathscr{H}$ and $g_i \in \mathscr{G}$, $i = 1, 2$ be two corresponding bijections. Then:*

1. *$h_1 h_2$ corresponds to $g_1 g_2$;*
2. *h_i^{-1} corresponds to g_i^{-1}, $i = 1, 2$.*

It is important then to obtain functions that remain unchanged with respect to classes of bijections.

Definition 51. *A function $m(\mathbf{x})$ is said to be invariant if*

$$m\big(g(\mathbf{x})\big) = m(\mathbf{x}), \forall g \in \mathscr{G}.$$

From here, arises

Definition 52. *A function $m(\mathbf{x})$ is said to be maximal invariant if*

1. *it is invariant;*
2. *$m(\mathbf{x}_1) = m(\mathbf{x}_2) \Rightarrow \exists g \in \mathscr{G} : \mathbf{x}_1 = g(\mathbf{x}_2)$.*

It is now possible to assert

Theorem 62. *A test ϕ is invariant for a group of bijections $\mathscr{G}$ if and only if it is a function of a maximal invariant $m(\mathbf{x})$.*

Proof. See [9], pg. 285. □

It is often the case that one needs to consider more than one sort (group) of transformations. It becomes then useful to determine the maximal invariant in the group generated by several groups of transformations.

Theorem 63. *Let $\mathscr{A}$ and $\mathscr{B}$ be two groups of bijections on the sample space, and $\mathscr{G}$ the group generated by them. Let s be a maximal invariant in $\mathscr{A}$ such that*

$$s(\mathbf{x}_1) = s(\mathbf{x}_2) \Rightarrow s\big(a(\mathbf{x}_1)\big) = s\big(a(\mathbf{x}_2)\big), a \in \mathscr{A}$$

and t_* a maximal invariant in

$$\mathscr{B}_* = \Big\{b_* : b_*(\mathbf{y}) = s\big(b(\mathbf{x}_1)\big), \mathbf{y} = s(\mathbf{x}), b \in \mathscr{B}\Big\}.$$

Then $z(\mathbf{x}) = t_*\big(s(\mathbf{x})\big)$ is a maximal invariant in $\mathcal{G}$.

Proof. See [9], pgs. 288–289. □

The next theorem shows an interesting and important property: when unbiasedness and invariance originate optimum tests, these tests coincide.

Theorem 64. *Let ϕ_u be a UMPU test that is unique (with probability 1) for an hypothesis H_0 and ϕ_i an unique UMPU test (with probability 1) for the same hypothesis. Then*

$$\mathbb{P}[\phi_u = \phi_i] = 1.$$

Proof. A proof of a slightly more general version of this theorem can be found in [9], pgs. 302–303. □

Such tests are called UMPUI (*uniformly most powerful unbiased invariant*).

4.2 Linear Models and F Tests

Given the results in the previous section consider the linear orthogonal mixed model

$$\mathbf{y} = \sum_{i=1}^{m} \mathbf{X}_i \boldsymbol{\beta}_i + \sum_{i=m+1}^{v} \mathbf{X}_i \boldsymbol{\beta}_i \tag{4.9}$$

in its canonical form

$$\mathbf{Y} = \sum_{j=1}^{b} \mathbf{A}_j \boldsymbol{\eta}_j + \sum_{j=b+1}^{w} \mathbf{A}_j \boldsymbol{\eta}_j, \tag{4.10}$$

where indexes $1, \ldots, b$ stand for fixed effects and indexes $b+1, \ldots, w$ stand for normal vectors with null mean and covariance matrix $\sigma_i^2\mathbf{I}$ and $\gamma_j\mathbf{I}$, respectively. For these models it is possible to obtain tests of hypothesis for most relevant hypothesis. These hypothesis usually focus on the nullity of fixed parameters and variance components, based on F tests. F tests are UMP, unbiased and therefore, optimality is achieved for these tests (for a more detailed discussion see [7], Chap. 2, section 2.5).

If one takes a COBS model with the density for the model is given by

$$n(\mathbf{y}) = \frac{-\frac{1}{2}\exp\left(\sum_{j=1}^{m}\frac{\|\mathbf{A}_j\mathbf{y}-\mathbf{A}_j\boldsymbol{\eta}_j\|^2}{\gamma_j} + \sum_{i=b+1}^{w}\frac{\|\mathbf{A}_{v+j}\mathbf{y}\|^2}{\gamma_j}\right)}{\sqrt{2^n\prod_{j=1}^{w}\gamma_j^{g_j}}}, \tag{4.11}$$

with g_j the rank of $\mathbf{A}_j$, $j = 1, \ldots, 2v$, tests for the hypothesis of nullity both for fixed parameters and variance components are easy to obtain. It is clear that, if $\gamma_j = \gamma_k$, with $j \leq m < k$, the statistic

$$T = \frac{g_k}{g} \cdot \frac{(\mathbf{G}\mathbf{A}_j\mathbf{y} - \mathbf{b})'(\mathbf{G}\mathbf{G}')^{-1}(\mathbf{G}\mathbf{A}_j\mathbf{y} - \mathbf{b})}{\|\mathbf{A}_k\mathbf{y}\|^2} \tag{4.12}$$

has an F distribution with g and g_j degrees of freedom with non centrality parameter

$$\delta = \frac{(\mathbf{G}\mathbf{A}_j\boldsymbol{\eta}_i - \mathbf{b})'(\mathbf{G}\mathbf{G}')^{-1}(\mathbf{G}\mathbf{A}_j\boldsymbol{\eta}_j - \mathbf{b})}{\gamma_k}, \tag{4.13}$$

where g is the rank of $\mathbf{G}$, a full rank matrix. Under the null hypothesis

$$H_0 : \mathbf{G}\mathbf{A}_j\boldsymbol{\eta}_j = \mathbf{b}, \tag{4.14}$$

the statistic T has a central F distribution. As far as variance components go, for any hypothesis of the form

$$H_0 : \gamma_{j_1} = \gamma_{j_2}, \tag{4.15}$$

one can consider the statistic

$$S = \frac{g_{j_2}}{g_{j_1}} \cdot \frac{\|\mathbf{A}_1\mathbf{y}\|^2}{\|\mathbf{A}_2\mathbf{y}\|^2}. \tag{4.16}$$

S has a central F distribution proportional to $\frac{\gamma_{j_1}}{\gamma_{j_2}}$ and a usual central F distribution under the null hypothesis.

As for variance components, following the approach that can be followed on [12] and [13], the estimator for a variance component in an orthogonal model has the form

$$\hat{\sigma}^2_{i_0} = \sum_{j \in I_+} c_j S_j - \sum_{j \in I_-} c_j S_j, \tag{4.17}$$

where $S_j = \|\mathbf{A}_j \mathbf{y}\|^2$, and I_+ and I_- are sets of indexes. Thus, a test for the hypothesis

$$H_0 : \sigma_{i_0} = 0 \text{ vs. } H_1 : \sigma_{i_0} > 0 \tag{4.18}$$

can be derived using the statistic $S = \frac{\sum_{j \in I_+} c_j S_j}{\sum_{j \in I_-} c_j S_j}$.

Theorem 65. *In a normal orthogonal model,*

$$S = \frac{\sum_{j \in I_+} c_j S_j}{\sum_{j \in I_-} c_j S_j}$$

has the distribution

$$F_G = \frac{\sum_{j \in I_+} c_j \gamma_j \frac{\chi^2_{(j)}}{g_j}}{\sum_{j \in I_-} c_j \gamma_j \frac{\chi^2_{(j)}}{g_j}},$$

where all the chi squares are independent.

Proof. It is trivial. □

Given this statistic, one needs only to consider a quantile of probability $1 - \alpha$, $q_{1-\alpha}$, and the critical value associated to it, c_α, to obtain a test procedure for the considered one sided hypothesis. Also easy to check is the assertion in

Theorem 66. *If* $\#(I_+) = \#(I_-) = 1$*, the statistic*

$$S = \frac{\sum_{j \in I_+} c_j S_j}{\sum_{j \in I_-} c_j S_j}$$

has F *distribution.*

Proof. It is trivial. □

A sufficient condition for the existence of a usual F test is given in

Theorem 67. *Consider a model with COBS with covariance matrix*

$$\mathbf{V} = \sum_{i=1}^{w} \sigma_i^2 \mathbf{V}_i + \mathbf{I}$$

and the Jordan algebra generated by $\{\mathbf{MV}_1, ..., \mathbf{M}\}$*, with* $\mathbf{M} = \mathbf{I} - \mathbf{T}$*,* $\mathbf{T}$ *the orthogonal projection on the mean vector space. If the linear space spanned by* $\{\mathbf{MV}_1, ..., \mathbf{M}\} \backslash \{\mathbf{MV}_{i_0}\}$ *is a commutative Jordan algebra,* $\#(I_+) = \#(I_-) = 1$*.*

Proof. See [12]. □

In [3] a converse result of the previous result was presented.

Theorem 68. *If*

$$S = \frac{\sum_{j \in I_+} c_j S_j}{\sum_{j \in I_-} c_j S_j}$$

has F *distribution when* H_0 *holds, the corresponding sub-model will have COBS.*

Proof. Let $\{\mathbf{E}_1, ..., \mathbf{E}_k\}$ be the principal basis for the commutative Jordan algebra generated by $\{\mathbf{MV}_1, ..., \mathbf{M}\}$. Thus, from the assumption of F distribution the UMVUE for $\sigma_{i_0}^2$ should have the structure

$$\mathbf{y}'\mathbf{A}\mathbf{y} = c \left(\frac{\mathbf{y}'\mathbf{E}_i\mathbf{y}}{\text{rank}(\mathbf{E}_i)} - \frac{\mathbf{y}'\mathbf{E}_j\mathbf{y}}{\text{rank}(\mathbf{E}_j)} \right), \quad i \neq j. \tag{4.19}$$

Clearly $c \neq 0$. Note that $\mathbf{V}_l$ can be expressed as follows:

$$\mathbf{V}_l = \sum_{m=1}^{k} \lambda_{l,m} \mathbf{E}_m \tag{4.20}$$

From (4.19) and the unbiasedness assumption it follows that for $l \neq i_0$ and $\sigma_l^2 \geq 0$, $\sigma_l^2 c\lambda_{l,i} = \sigma_l^2 c\lambda_{l,j}$. Thus, whenever $\sigma_l^2 > 0$, $c\sigma_l^2 \neq 0$ as well as $\lambda_{l,i} = \lambda_{l,j}$ for $l \neq i_0$. This means that $\mathcal{V}_{i_0} = \text{sp}\{\mathbf{E}_1, ..., \mathbf{E}_i + \mathbf{E}_j, ..., \mathbf{E}_k\}$ is a commutative Jordan algebra, and so the corresponding sub-model will have COBS. □

4.3 Generalized F Distribution

Let $F(u|\mathbf{c}_1, \mathbf{c}_2, \mathbf{g}_1, \mathbf{g}_2)$ be the distribution of

$$\mathcal{F} = \frac{\sum_{j=1}^{r} c_{1j}\chi^2_{(g_{2j})}}{\sum_{j=1}^{s} c_{2j}\chi^2_{(g_{2j})}} \tag{4.21}$$

when the chi-squares are independent. Clearly the c_{1j} and g_{1j}, $j = 1, ...r$ [c_{2j} and g_{2j}, $j = 1, ...s$] will be the components of $\mathbf{c}_1$ and $\mathbf{g}_1$ [$\mathbf{c}_2$ and $\mathbf{g}_2$]. The generalized F distributions will belong to this family, namely they correspond to the case in which $\mathbf{c}_1 > \mathbf{0}$, $\mathbf{c}_2 > \mathbf{0}$ and $\sum_{j=1}^{r} c_{1j}g_{1j} = \sum_{j=1}^{s} c_{2j}g_{2j}$.

Start by assuming that $\mathbf{c}_1 > \mathbf{0}$, $\mathbf{c}_2 > \mathbf{0}$ and that $\mathbf{g}_1 = 2\mathbf{m}$. These results extend directly to the case in which $\mathbf{g}_2 = 2\mathbf{m}$, since

$$F(u|\mathbf{c}_1, \mathbf{c}_2, \mathbf{g}_1, 2\mathbf{m}) = 1 - F(u^{-1}|\mathbf{c}_2, \mathbf{c}_1, 2\mathbf{m}, \mathbf{g}_1). \tag{4.22}$$

The cases of generalized F distributions will be covered whenever one of the evenness conditions hold.

To lighten the writing put $a_i = c_{1i}$ and $g_i = g_{1i}$, $i = 1, ..., r$, as well as $a_{i+r} = c_{2i}$ and $g_{i+r} = g_{2i}$, $i = 1, ..., s$. Define also $X_i = a_i\chi^2_{(g_i)}$, $i = 1, ..., r + s$. Successive integrations are now carried out. The density of X_i will be

$$\begin{cases} f_i(x) = \dfrac{x^{m_i-1}}{(m_i-1)!(2a_i)^{m_i}}\, e^{-\frac{x}{2a_i}} & , x > 0, i = 1, ..., r \\ f_i(x) = \dfrac{x^{\frac{g_i}{2}-1}}{\Gamma(\frac{g_i}{2})(2a_i)^{\frac{g_i}{2}}}\, e^{-\frac{x}{2a_i}} & , x > 0, i = r+1, ..., r+s \end{cases} \tag{4.23}$$

Then

$$\begin{aligned} F(u|\mathbf{c}_1, \mathbf{c}_2, 2\mathbf{m}, \mathbf{g}_2) =& \mathbb{P}\left[\sum_{i=1}^{r} X_i \leq u \sum_{i=r+1}^{r+s} X_i\right] \\ =& \int_0^{+\infty} \cdots \int_0^{+\infty} \int_0^{u\sum_{i=r+1}^{r+s} x_i} \cdots \\ & \cdots \int_0^{u\sum_{i=r+1}^{r+s} x_i - \sum_{i=2}^{r} x_i} \prod_{j=1}^{r+s} f_j(x_j)dx_j. \end{aligned} \tag{4.24}$$

Note that, with b a non-negative integer, through successive integration by parts we get

$$\int_0^d x^b e^{-\frac{x}{a}} dx = a^{b+1} b! \left(1 - e^{-\frac{d}{a}} \sum_{j=0}^{b} \frac{d_j}{a^j j!}\right). \tag{4.25}$$

This expression can be rewritten as

$$\int_0^d x^b e^{-\frac{x}{a}} dx = a^{b+1} b! \sum_{k=0}^{1} e^{-\frac{kd}{a}} \sum_{j=0}^{kb} \frac{d^j}{a^j j!}. \tag{4.26}$$

Using (4.26), with $y = u \sum_{i=r+1}^{r+s} x_i$ and $b_i = m_i - 1$, $i = 1, ..., r$, and in the first place, $a = 2a_1$ and $d = y - \sum_{i=2}^{r} x_i$, one gets

$$\begin{aligned}
\int_0^{y-\sum_{i=2}^{r} x_i} f_1(x_1)\, dx_1 &= \int_0^{y-\sum_{i=2}^{r} x_i} \frac{x_1^{b_1} e^{-\frac{x_1}{a}}}{b_1!(2a_1)^{b_1+1}}\, dx_1 \\
&= \sum_{k_1=0}^{1} (-1)^{k_1} e^{-\frac{k_1}{2a_1}\left(y-\sum_{i=2}^{r} x_i\right)} \sum_{j_1=0}^{k_1 b_1} \frac{\left(y - \sum_{i=2}^{r} x_i\right)^{j_1}}{j_1!(2a_1)^{j_1}} \\
&= \sum_{k_1=0}^{1} (-1)^{k_1} e^{-\frac{k_1}{2a_1}\left(y-\sum_{i=2}^{r} x_i\right)} \sum_{j_1=0}^{k_1 b_1} \frac{1}{(2a_1)^{j_1}} \\
&\qquad \sum_{\left(\sum_{i=1}^{j_1} t_{1,i} = j_1\right)} (-1)^{j_1 - t_{1,1}} \frac{y^{t_{1,1}}}{t_{1,1}!} \prod_{i=2}^{r} \frac{x_i^{t_{1,i}}}{t_{1,i}!}
\end{aligned} \tag{4.27}$$

were $\left(\sum_{i=1}^{j_1} t_{1,i} = j_1\right)$ indicates summation for all sets of non negative integers $t_{1,1}, ..., t_{1,r}$

with sum j_1. Thus, it is possible to apply again (4.26) to get

$$\int_0^{y-\sum_{i=3}^{r} x_i}\int_0^{y-\sum_{i=2}^{r} x_i} f_1(x_1)f_2(x_2)\,dx_1dx_2 =$$
$$\sum_{k_1=0}^{1}(-1)^{k_1}e^{-\frac{k_1}{2a_1}\left(y-\sum_{y=3}^{r} x_i\right)}\sum_{j_1=0}^{k_1b_1}\sum_{\left(\sum_{i=1}^{r} t_{1,i}=j_1\right)}\frac{y^{t_{1,1}}}{t_{1,1}!}\prod_{i=3}^{r}\frac{x_i^{t_{1,i}}}{t_{1,i}!}\cdot\frac{1}{b_2!t_{1,2}!}\int_0^{y-\sum_{i=3}^{r} x_i}\frac{x_2^{b_2+t_{1,2}}}{(2a_2)^{b_2+1}}$$
$$e^{-\left(\frac{1}{2a_2}-\frac{k_1}{2a_1}\right)}\,dx_2 = \sum_{k_1=0}^{1}\frac{(-1)^{k_1}}{(2a_2)^{b_2+1}}\sum_{j_1=0}^{k_1b_1}\frac{1}{(2a_1)^{j_1}}\sum_{\left(\sum_{i=1}^{r} t_{1,i}=j_1\right)}(-1)^{j_1-t_{1,1}}$$
$$\left(\frac{1}{2a_2}-\frac{k_1}{2a_1}\right)^{-(b_2+t_{1,2}+1)}\frac{(b_2+t_{1,2})!}{b_2!t_{1,2}!}\sum_{k_2=0}^{1}(-1)^{k_2}e^{-\left(\frac{k_1}{2a_1}+k_2\left(\frac{1}{2a_2}-\frac{k_1}{2a_2}\right)\right)\left(y-\sum_{i=3}^{r} x_i\right)}$$
$$\sum_{j_2=0}^{k_2(b_2+t_{1,2})}\left(\frac{1}{2a_2}-\frac{k_1}{2a_1}\right)^{j_2}\sum_{\left(\sum_{i=1}^{r-1} t_{2,1}=j_2\right)}(-1)^{j_2-t_{2,1}}\frac{y^{t_{1,1}+t_{2,1}}}{t_{1,1}!t_{2,1}!}\prod_{i=3}^{r}\frac{x_i^{t_{1,i}+t_{2,i-1}}}{t_{1,i}!t_{2,i-1}!}. \quad (4.28)$$

Now, with $b_j^+ = b_j + \sum_{u=1}^{j-1} t_{u,j+1-u}$, $j = 1, ..., w$, $b_1^+ = b_1$ and $b_2^+ = b_2 + t_{1,2}$. Besides this, taking $\mathbf{k} = (k_1, ..., k_n)'$, comes $\mathbf{k}_2 = (k_1, k_2)'$. Let $d(\mathbf{0}) = 0$ and, for $\mathbf{k} \neq \mathbf{0}$, $d(\mathbf{k}) = \frac{1}{a_{w_n}}$ with w_n the largest index for non null components of $\mathbf{k}$, $n = 1, ..., r$. It is easy to check that $\frac{k_1}{2a_1} + k_2(\frac{1}{2a_2} - \frac{k_1}{2a_1}) = d(\mathbf{k}_2)$, and that $d(\mathbf{k}) + k_{n+1}\left(\frac{1}{2a_{n+1}} - d(\mathbf{k})\right) = d(\mathbf{k})$, the last $\mathbf{k}$ having one more component.

Rewriting (4.28):

$$\int_0^{y-\sum_{i=3}^{r} x_i}\int_0^{y-\sum_{i=2}^{r} x_i} f_1(x_1)f_2(x_2)dx_1dx_2 =$$
$$\frac{1}{(2a_2)^{b_2+1}}\sum_{k_1=0}^{1}\sum_{j_1=0}^{k_1b_1^+}\frac{1}{(2a_1)^{j_1}}\sum_{\left(\sum_{i=1}^{r} t_{1,i}=j_1\right)}\sum_{k_2=0}^{1}\sum_{j_2=0}^{k_2b_2^+}\sum_{\left(\sum_{i=1}^{r} t_{2,i}=j_2\right)}(-1)^{\sum_{i=1}^{2}(k_i+j_i-t_{i,1})}$$
$$\frac{b_2^+!}{b_2!t_{1,2}!}\left(\frac{1}{2a_2}-\frac{k_1}{2a_1}\right)^{j_2-b_2^+-1}\frac{e^{-d(\mathbf{k}_2)y}y^{t_{1,1}+t_{2,1}}}{t_{1,1}!t_{2,1}!}\prod_{i=3}^{r}\frac{e^{-d(\mathbf{k})x_i}x_i^{t_{1,i}+t_{2,i-1}}}{t_{1,i}!t_{2,i-1}!}. \quad (4.29)$$

Establish now

Lemma 4. *With*

$$L_{r,m}(y, x_{m+1}, ..., x_r) = \int_0^{y-\sum_{i=m+1}^{r} x_i} \cdots \int_0^{y-\sum_{i=2}^{r} x_i} \prod_{v=1}^{m} f_v(x_v) \prod_{v=1}^{m} dx_v$$

one gets

$$L_{r,m}(y, x_m + 1, ..., x_r) =$$
$$\int_0^{y-\sum_{i=m+1}^{r} x_i} \cdots \int_0^{y-\sum_{i=2}^{r} x_i} \prod_{v=1}^{m} f_v(x_v) \prod_{v=1}^{m} dx_v = \frac{1}{\prod_{v=2}^{m} (2a_v)^{b_v+1}}$$
$$\sum_{k_1=0}^{1} \sum_{j_1=0}^{k_1 b_1^+} \sum_{\left(\sum_{i=1}^{r} t_{1,i}=j_1\right)} \cdots \sum_{k_m=0}^{1} \sum_{j_m} \sum_{\left(\sum_{i=1}^{r+1-m} t_{m,i}=j_m\right)} (-1)^{\sum_{i=1}^{m}(k_i+j_i-t_{i,1})} \frac{1}{(2a_1)^{j_1}}$$
$$\prod_{v=2}^{m} \left(\frac{b_v^+!}{b_v! \prod_{i=1}^{v-1} t_{i,v+1-i}!} \left(\frac{1}{2a_v} - d(\mathbf{k}_{v-1}) \right)^{j_v - b_v^+ - 1} \right) \frac{e^{-d(\mathbf{k}_m)y} y^{\sum_{i=1}^{m} t_{i,1}}}{\prod_{i=1}^{m} t_{i,1}!}$$
$$\prod_{i=m+1}^{r} \frac{e^{d(\mathbf{k}_m)x_i} x_i^{\sum_{v=1}^{m} t_{v,i+1-v}}}{\prod_{v=1}^{m} t_{v,i+1-v}!}.$$

Proof. Induction is used in m. If $m = 2$ the thesis holds since, then, one has (4.29). Assume that it holds for $m \leq w < r$, since $b_{w+1}^+ = b_w + \sum_{v=1}^{w} t_{v,w+1-v}$. Thus, using

(4.26),

$$\int_0^{y-\sum_{i=w+2}^{r} x_i} e^{d(\mathbf{k}_w)x_{w+1}} \frac{x_{w+1}^{\sum_{v=1}^{w} t_{v,w+2-v}}}{\prod_{v=1}^{w} t_{v,w+2-v}!} f_{w+1}(x_{w+1})\, dx_{w+1} =$$

$$\frac{1}{b_{w+1}! \prod_{v=1}^{w} t_{v,w+2-v}!} \cdot \frac{1}{(2a_{w+1})^{b_{w+1}+1}} \int_0^{y-\sum_{i=w+2}^{r} x_i} x_{w+1}^{b_{w+1}+\sum_{v=1}^{w}(t_{v,w+2-v})}$$

$$e^{-\left(\frac{1}{2a_{w+1}}-d(\mathbf{k}_w)\right)x_{w+1}}\, dx_{w+1} = \frac{b_{w+1}^{+}!}{b_{w+1}! \prod_{v=1}^{w} t_{v,w+2-v}!} \cdot \frac{1}{(2a_{w+1})^{b_{w+1}+1}}$$

$$\left(\frac{1}{2a_{w+1}} - d(\mathbf{k}_w)\right)^{-b_{w+1}^{+}-1} \sum_{k_{w+1}=0}^{1} \sum_{j_{w+1}=0}^{k_{w+1}b_{w+1}^{+}} e^{-k_{w+1}\left(\frac{1}{2a_{w+1}}-d(\mathbf{k}_w)\right)\left(y-\sum_{i=w+2}^{r} x_i\right)}$$

$$\frac{(-1)^{k_{w+1}}}{j_{w+1}!} \left(\frac{1}{2a_{w+1}} - d(\mathbf{k}_w)\right)^{j_{w+1}} \left(y - \sum_{i=w+2}^{r} x_i\right)^{j_{w+1}} =$$

$$\frac{b_{w+1}^{+}!}{b_{w+1}! \prod_{v=1}^{w} t_{v,w+2-v}!} \cdot \frac{1}{(2a_{w+1})^{b_{w+1}+1}} \sum_{k_{w+1}=0}^{1} \sum_{j_{w+1}=0}^{k_{w+1}b_{w+1}^{+}}$$

$$\sum_{\left(\sum_{i=1}^{r+1-(w+1)} t_{w+1,i}=j_{w+1}\right)} (-1)^{k_{w+1}+j_{w+1}-t_{w+1,1}} \left(\frac{1}{2a_{w+1}} - d(\mathbf{k}_w)\right)^{j_{w+1}-b_{w+1}^{+}-1}$$

$$e^{-k_{w+1}\left(\frac{1}{2a_{w+1}}-d(\mathbf{k}_w)\right)y} \frac{y^{t_{w+1,1}}}{t_{w+1,1}!} \prod_{i=w+2}^{r} \frac{e^{k_{w+1}\left(\frac{1}{2a_{w+1}}-d(\mathbf{k}_w)\right)x_i} x_i^{t_{w+1,(i+1)-(w+1)}}}{t_{w+1,(i+1)-(w+1)}!}.$$

Since $d(\mathbf{k}_w) + k_{w+1}\left(\frac{1}{2a_{w+1}} - d(\mathbf{k}_w)\right) = d(\mathbf{k}w+1)$,

$$L_{w+1.r}(y, x_{w+2}, ..., x_r) = \int_0^{y-\sum_{w+2}^{r} x_i} L_{w.r}(y, x_{w+1}, ..., x_r) f_{w+1}(x_{w+1})\, dx_{w+1}$$

$$= \frac{1}{\prod_{v=2}^{w} (2a_v)^{b_v+1}} \sum_{k_1=0}^{1} \sum_{j_1=0}^{k_1 b_1^+} \sum_{\left(\sum_{i=1}^{r} t_{1,i}=j_1\right)} \cdots \sum_{k_w=0}^{1} \sum_{j_w=0}^{k_w b_w^+} \sum_{\left(\sum_{i=1}^{r+1-w} t_{w,i}=j_w\right)} \frac{(-1)^{\sum_{i=1}^{w}(k_i+j_i-t_{i,1})}}{(2a_1)^{j_1}}$$

$$\prod_{v=2}^{w} \left(\frac{b_v^+!}{b_v! \prod_{i=1}^{v-1} t_{i,v+1-1}!} \left(\frac{1}{2a_v} - d(\mathbf{k}_{v-1})\right)^{j_v - b_v^+ - 1} \right) \frac{e^{-d(\mathbf{k}_w)y} y^{\sum_{i=1}^{w} t_{i,1}}}{\prod_{i=1}^{w} t_{i,1}!}$$

$$\prod_{i=w+2}^{r} \frac{e^{-d(\mathbf{k}_w)x_i} x_i^{\sum_{v=1}^{w} t_{v,i+1-v}}}{\prod_{v=1}^{w} t_{v,i+1-v}!} \int_0^{y-\sum_{i=w+2}^{r} x_i} \frac{e^{d(\mathbf{k}_w)x_{w+1}} x_{w+1}^{\sum_{v=1}^{w} t_{v,w+2-v}}}{\prod_{v=1}^{w} t_{v,w+2-v}!} f_{w+1}(x_{w+1})\, dx_{w+1}$$

$$= \frac{1}{\prod_{v=2}^{w+1} (2a_v)^{b_v+1}} \sum_{k_1=0}^{1} \sum_{j_1=0}^{k_1 b_1^+} \sum_{\left(\sum_{i=1}^{r} t_{1,i}=j_1\right)} \cdots \sum_{k_w=0}^{1} \sum_{j_w=0}^{k_w b_w^+} \sum_{\left(\sum_{i=1}^{r+1-w} t_{w,i}=j_w\right)} \sum_{k_{w+1}=0}^{1} \sum_{j_{w+1}=0}^{k_{w+1} b_{w+1}^+}$$

$$\sum_{\left(\sum_{i=1}^{r+1-(w+1)} t_{w+1,i}=j_{w+1}\right)} \frac{(-1)^{\sum_{i=1}^{w+1} k_i+j_i-t_{i,1}}}{(2a_1)^{j_1}} \prod_{v=2}^{w+1} \left(\frac{b_v^+!}{b_v! \prod_{i=1}^{v-1} t_{i,v+1-1}!} \right.$$

$$\left. \left(\frac{1}{2a_v} - d(\mathbf{k}_{v-1})\right)^{j_v - b_v^+ - 1} \right) \frac{e^{-d(\mathbf{k}_{w-1})y} y^{\sum_{i=1}^{w+1} t_{i,1}}}{\prod_{i=1}^{w+1} t_{i,1}!} \prod_{i=w+2}^{r} \frac{e^{-d(\mathbf{k}_{w-1})x_i} x_i^{\sum_{v=1}^{w+1} t_{v,i+1-v}}}{\prod_{v=1}^{w+1} t_{v,i+1-v}!}$$

which completes the proof. □

It is now possible to establish

Proposition 1.

$$F(u|\mathbf{c}_1,\mathbf{c}_2,2\mathbf{m},\mathbf{g}_2)=\frac{1}{\prod\limits_{v=2}^{r}(2a_v)^{b_v+1}}\sum_{k_1=0}^{1}\sum_{j_1=0}^{k_1b_1^+}\sum_{\left(\sum\limits_{i=1}^{r}t_{1,i}=j_1\right)}\cdots\sum_{k_r=0}^{1}\sum_{j_r=0}^{k_rb_r^+}$$
$$\sum_{\left(\sum\limits_{i=1}^{1}t_{r,i}=j_r\right)}\frac{(-1)^{\sum\limits_{i=1}^{r}(k_i+j_i-t_{i,1})}}{(2a_1)^{j_1}}\prod_{i=2}^{r}\left(\frac{b_i^+!}{b_i!\prod\limits_{v=1}^{i-1}t_{v,i+1-v}!}\left(\frac{1}{2a_i}-d(\mathbf{k}_{i-1})\right)^{j_i-b_i^+-1}\right)$$
$$\frac{\left(\sum\limits_{i=1}^{r}t_{i,1}\right)!u^{\sum\limits_{v=1}^{r}rt_{v,1}}}{\prod\limits_{i=1}^{r}t_{i,1}!}\sum_{\left(\sum\limits_{i=r+1}^{r+s}l_i=\sum\limits_{v=1}^{r}t_{v,1}\right)}\prod_{i=r+1}^{r+s}\left(\frac{\Gamma\left(l_i+\frac{g_i}{2}\right)}{l_i!\Gamma\left(\frac{g_i}{2}\right)}\cdot\frac{1}{(2a_i)^{\frac{g_i}{2}}}\right.$$
$$\left.\left(\frac{1}{2a_i}+d(\mathbf{k}_r)u\right)^{-\left(l_i+\frac{g_i}{2}\right)}\right).$$

Proof. Using Lemma 4, take

$$F(u|\mathbf{c}_1,\mathbf{c}_2,2\mathbf{m},\mathbf{g}_2)=\int_0^{+\infty}\cdots\int_0^{+\infty}L_{r,r}\left(u\sum_{i=r+1}^{r+s}x_i\right)\prod_{i=r+1}^{r+s}f_i(x_i)dx_i=$$
$$\int_0^{+\infty}\cdots\int_0^{+\infty}\left(\frac{1}{\prod\limits_{v=2}^{r}(2a_v)^{b_v+1}}\sum_{k_1=0}^{1}\sum_{j_1=0}^{k_1b_1^+}\sum_{\left(\sum\limits_{i=1}^{r}t_{1,i}=j_1\right)}\cdots\sum_{k_r=0}^{1}\sum_{j_r=0}^{k_rb_r^+}\sum_{\left(\sum\limits_{i=1}^{1}t_{r,i}=j_r\right)}\frac{(-1)^{\sum\limits_{v=1}^{r}k_v+j_v-t_{v,1}}}{(2a_1)^{j_1}}\right.$$
$$\left.\prod_{w=2}^{r}\left(\frac{b_w^+!}{b_w!\prod\limits_{v=1}^{w-1}t_{v,w+1-v}!}\right)\left(\frac{1}{2a_w}-d(\mathbf{k}_{w-1})\right)^{j_v-b_v^+-1}\frac{e^{-d(\mathbf{k}_r)u}u^{\sum\limits_{i=1}^{r}t_{i,1}}}{\prod\limits_{i=1}^{r}t_{i,1}!}\right)\prod_{i=r+1}^{r+s}f_i(x_i)dx_i$$

(Note that $\prod\limits_{i=r+1}^{r}\frac{e^{d(\mathbf{k}_r)x_i}x_i^{\sum\limits_{i=1}^{r}t_{i,1}}}{\prod\limits_{i=1}^{r}t_{i,1}!}=1$ since the lower index exceeds the upper one) thus

getting

$$F(u|\mathbf{c}_1,\mathbf{c}_2,2\mathbf{m},\mathbf{g}_2)=$$

$$\frac{1}{\prod\limits_{v=2}^{r}(2a_v)^{b_v+1}}\sum_{k_1=0}^{1}\sum_{j_1=0}^{k_1b_1^+}\sum_{\left(\sum\limits_{i=1}^{r}t_{1,i}=j_1\right)}\cdots\sum_{k_r=0}^{1}\sum_{j_r=0}^{k_rb_r^+}\sum_{\left(\sum\limits_{i=1}^{1}t_{r,i}=j_r\right)}\frac{(-1)^{\sum\limits_{v=1}^{r}(k_v+j_v-t_{v,1})}}{(2a_1)^{j_1}}$$

$$\prod_{w=2}^{r}\left(\frac{b_w^+!}{b_w!\prod\limits_{v=1}^{w-1}t_{v,w+1-v}!}\right)\left(\frac{1}{2a_w}-d(\mathbf{k}_{w-1})\right)^{j_v-b_v^+-1}\int_0^{+\infty}\cdots\int_0^{+\infty}e^{-d(\mathbf{k}_r)u\sum\limits_{i=r+1}^{r+s}x_i}$$

$$\frac{\left(u\sum\limits_{i=r+1}^{r+s}x_i\right)^{\sum\limits_{i=1}^{r}t_{i,1}}}{\prod\limits_{i=1}^{r}t_{i,1}!}\prod_{i=r+1}^{r+s}f_i(x_i)dx_i$$

Now

$$\int_0^{+\infty}\cdots\int_0^{+\infty}e^{-d(\mathbf{k}_r)u\sum\limits_{i=r+1}^{r+s}x_i}\frac{\left(u\sum\limits_{i=r+1}^{r+s}x_i\right)^{\sum\limits_{i=1}^{r}t_{i,1}}}{\prod\limits_{i=1}^{r}t_{i,1}!}\prod_{i=r+1}^{r+s}f_i(x_i)dx_i=$$

$$\sum_{\left(\sum\limits_{i=r+1}^{r+s}l_i=\sum\limits_{v=1}^{r}t_{v,1}\right)}\frac{\left(\sum\limits_{v=1}^{r}t_{v,1}\right)!u^{\sum\limits_{v=1}^{r}t_{v,1}}}{\prod\limits_{v=1}^{r}t_{v,1}!\prod\limits_{v=r+1}^{r+s}l_i!}\prod_{i=r+1}^{r+s}\int_0^{+\infty}\frac{e^{-\left(\frac{1}{2a_i}+d(\mathbf{k}_r)\right)x_i}}{\Gamma\left(\frac{g_i}{2}\right)(2a_i)^{\frac{g_i}{2}}}x_i^{l_i+\frac{g_i}{2}-1}dx_i=$$

$$\sum_{\left(\sum\limits_{i=r+1}^{r+s}l_i=\sum\limits_{v=1}^{r}t_{v,1}\right)}\frac{\left(\sum\limits_{v=1}^{r}t_{v,1}\right)!u^{\sum\limits_{v=1}^{r}t_{v,1}}}{\prod\limits_{v=1}^{r}t_{v,1}!\prod\limits_{v=r+1}^{r+s}l_i!}\prod_{i=r+1}^{r+s}\left(\frac{\Gamma\left(l_i+\frac{g_i}{2}\right)}{\Gamma\left(\frac{g_i}{2}\right)(2a_i)^{\frac{g_i}{2}}}\left(\frac{1}{2a_i}+d(\mathbf{k}_r)u\right)^{-\left(l_i+\frac{g_i}{2}\right)}\right).$$

To compute the last integrals it suffices to use the transformation

$$z_i=\left(\frac{1}{2a_i}+d(\mathbf{k}_r)u\right)x_i, \tag{4.30}$$

$i=r+1,...,r+s$, and use the well known gamma function. To complete the

proof it is only needed to substitute the result just obtained in the expression of $F(u|\mathbf{c}_1, \mathbf{c}_2, 2\mathbf{m}, \mathbf{g}_2)$. □

5. CROSS NESTED MODELS

This chapter is started by presenting the models in this section, the balanced cross nested random models. They are considered a special case of regular models.

5.1 Models

With Γ the set of vectors $\mathbf{h}$ with integer components $h_l = 0, ..., u_l$, $l = 1, ...L$, we assume, for the vector $\mathbf{y}$ of observations, the model

$$\mathbf{y} = \sum_{\mathbf{h}\in\Gamma} \mathbf{X}(\mathbf{h})\boldsymbol{\beta}(\mathbf{h}) + \mathbf{e} \tag{5.1}$$

where $\mathbf{X}(\mathbf{0}) = \mathbf{1}^n$, $\boldsymbol{\beta}(\mathbf{0}) = \mu$, and, with $\mathbf{h} \neq \mathbf{0}$, $\mathbf{X}(\mathbf{h}) = \bigotimes_{l=1}^{L} \mathbf{X}_l(h_l) \otimes \mathbf{1}^r$. The vectors $\boldsymbol{\beta}(\mathbf{h})$, with $\mathbf{h} \neq \mathbf{0}$, and $\mathbf{e}$ will be normal, independent, with null mean vectors and covariance matrices $\sigma^2(\mathbf{h})\mathbf{I}_{c(\mathbf{h})}$ and $\sigma^2\mathbf{I}_n$, respectively. The crossing was integrated in the model through the use of the Kronecker matrix product, while, to have nesting, we require that $\mathrm{R}(\mathbf{X}_l(h_l)) \subset \mathrm{R}(\mathbf{X}_l(h_l + 1))$, $h_l = 0, ..., u_l - 1$, $l = 1, ..., L$. This means that $\mathrm{R}(\mathbf{X}_l(h_l))$ is strictly included in $\mathrm{R}(\mathbf{X}_l(h_l + 1))$. Lastly, model orthogonality is achieved by assuming that $\mathbf{M}_l(h_l) = \mathbf{X}_l(h_l)\mathbf{X}_l'(h_l) = b(h_l)\mathbf{Q}_l(h_l)$, with $\mathbf{Q}_l(h_l)$ an orthogonal projection matrix, $h_l = 0, ..., u_l$, $l = 1, ..., L$. In order to have $\mathbf{X}(\mathbf{0}) = \mathbf{1}^n$ one must have $\mathbf{X}_l(0) = \mathbf{1}^{n_l}$, $l = 1, ..., L$, as well as $n = r\prod_{l=1}^{L} n_l$. Then $\mathbf{Q}_l(0) = \frac{1}{n_l}\mathbf{J}_{n_l}$, $l = 1, ..., L$, when $\mathbf{J}_v = \mathbf{1}^v\mathbf{1}^{v\prime}$. As we shall see, this last assumption will enable us to derive a Jordan algebra $\mathcal{A}$ automatically associated to the model. Such algebra corresponds, as it is well known, to orthogonal partitions and so, in this way, the orthogonality of the model is ensured.

On account of nesting we will also have the orthogonal projection matrices $\mathbf{B}_l(0) = \mathbf{Q}_l(0)$ and $\mathbf{B}_l(t_l) = \mathbf{Q}_l(t_l) - \mathbf{Q}_l(t_l - 1)$, $t_l = 1, ..., u_l$, $l = 1, ..., L$. The Kronecker matrix product enables the construction of the orthogonal projection matrices:

$$\begin{cases} \mathbf{Q}(\mathbf{h}) = \bigotimes_{l=1}^{L} \mathbf{Q}_l(h_l) \otimes \frac{1}{r}\mathbf{J}_r, & \mathbf{h} \in \Gamma \\ \mathbf{B}(\mathbf{t}) = \bigotimes_{l=1}^{L} \mathbf{B}_l(t_l) \otimes \frac{1}{r}\mathbf{J}_r, & \mathbf{t} \in \Gamma \end{cases} . \tag{5.2}$$

If matrices $\mathbf{Q}_l(h_l)$, $h_l = 0, ..., u_l$, have ranks $r_l(0) = 1$ and $r_l(h_l)$, $h_l = 1, ..., u_l$, matrices $\mathbf{B}_l(t_l)$, $t_l = 0, ..., u_l$, will have ranks $g_l(0) = 1$ and $g_l(t_l) = r_l(t_l) - r_l(t_l - 1)$, $t_l = 1, ..., u_l$, $l = 1, ..., L$. Moreover, matrices $\mathbf{Q}(\mathbf{h})$ $[\mathbf{B}(\mathbf{t})]$ with $\mathbf{u} \in \Gamma$ $[\mathbf{t} \in \Gamma]$ will have rank $r(\mathbf{h}) = \prod_{l=1}^{L} r_l(h_l)$ $\left[g(\mathbf{t}) = \prod_{l=1}^{L} g_l(t_l)\right]$.

In order to illustrate these concepts, an example: a three factor random model, in which the second factor nests the third. We then get the model

$$\mathbf{Y} = \sum_{\mathbf{h}_2 \in \Gamma} \mathbf{X}(\mathbf{h}_2)\boldsymbol{\beta}(\mathbf{h}_2) + \mathbf{e}, \tag{5.3}$$

were $\mathbf{h}_2 = (i_1, i_2)$, $i = 0, 1$, $j = 0, 1, 2$. Taking $a_1(1)$ as the number of levels of the first factor, $a_2(1)$ the number of levels of the second and $a_2(2)$ the number of levels of the third (nested on the second), with $a_1(0) = a_0(2) = 1$. The following matrices are then obtained:

$$\begin{cases} \mathbf{X}(1,0) = \mathbf{I}_{a_1(1)} \otimes \mathbf{1}^{a_2(1)a_2(2)} \\ \mathbf{X}(0,1) = \mathbf{1}^{a_1(1)} \otimes \mathbf{I}_{a_2(1)} \otimes \mathbf{1}^{a_2(2)} \\ \mathbf{X}(0,2) = \mathbf{1}^{a_1(1)} \otimes \mathbf{I}_{a_2(1)a_2(2)} \\ \mathbf{X}(1,1) = \mathbf{I}_{a_1(1)a_2(1)} \otimes \mathbf{1}^{a_2(2)} \\ \mathbf{X}(1,2) = \mathbf{I}_{a_1(1)a_2(1)a_2(2)} \end{cases} , \tag{5.4}$$

having also $\boldsymbol{\beta}(0,0) = \mu$ and $\boldsymbol{\beta}(i,j) \sim \mathcal{N}(\mathbf{0}, \sigma^2(i,j)\mathbf{I}_{c(i,j)})$, $i = 0, 1$, $j = 0, 1, 2$,

$(i,j) \neq (0,0)$. Consequently

$$\begin{cases} \mathbf{Q}(1,0) = \frac{1}{a_2(1)a_2(2)}\mathbf{I}_{a_1(1)} \otimes \mathbf{J}_{a_2(1)a_2(2)} \\ \mathbf{Q}(0,1) = \frac{1}{a_1(1)a_2(2)}\mathbf{J}_{a_1(1)} \otimes \mathbf{I}_{a_2(1)} \otimes \mathbf{J}_{a_2(2)} \\ \mathbf{Q}(0,2) = \frac{1}{a_1(1)}\mathbf{J}_{a_1(1)} \otimes \mathbf{I}_{a_2(1)a_2(2)} \\ \mathbf{Q}(1,1) = \frac{1}{a_2(2)}\mathbf{I}_{a_1(1)a_2(1)} \otimes \mathbf{J}_{a_2(2)} \\ \mathbf{Q}(1,2) = \mathbf{I}_{a_1(1)a_2(1)a_2(2)} \end{cases} . \tag{5.5}$$

Consequently:

$$\begin{cases} \mathbf{B}(1,0) = \frac{1}{a_2(1)a_2(2)r}\mathbf{K}_{a_1(1)} \otimes \mathbf{J}_{a_2(1)a_2(2)r} \\ \mathbf{B}(0,1) = \frac{1}{a_1(1)a_2(2)r}\mathbf{J}_{a_1(1)} \otimes \mathbf{K}_{a_2(1)} \otimes \mathbf{J}_{a_2(2)r} \\ \mathbf{B}(0,2) = \frac{1}{a_1(1)r}\mathbf{J}_{a_1(1)} \otimes \mathbf{I}_{a_2(1)}\mathbf{K}_{a_2(2)} \otimes \mathbf{J}_r \\ \mathbf{B}(1,1) = \frac{1}{a_2(2)r}\mathbf{K}_{a_1(1)} \otimes \mathbf{K}_{a_2(1)} \otimes \mathbf{J}_{a_2(2)r} \\ \mathbf{B}(1,2) = \frac{1}{r}\mathbf{K}_{a_1(1)} \otimes \mathbf{I}_{a_2(1)} \otimes \mathbf{K}_{a_2(2)} \otimes \mathbf{J}_r \end{cases} , \tag{5.6}$$

with r being the number of repetitions and $\mathbf{K}_v = \mathbf{I}_v - \frac{1}{v}\mathbf{J}_v$, and

$$\begin{cases} \text{rank}(\mathbf{B}(1,0)) = a_1(1) - 1 \\ \text{rank}(\mathbf{B}(0,1)) = a_2(1) - 1 \\ \text{rank}(\mathbf{B}(0,2)) = a_2(1)(a_2(2) - 1) \\ \text{rank}(\mathbf{B}(1,1)) = (a_1(1) - 1)(a_2(1) - 1) \\ \text{rank}(\mathbf{B}(1,2)) = (a_1(1) - 1)a_2(1)(a_2(2) - 1) \end{cases} . \tag{5.7}$$

This is in fact the model used in [2].

Now, in order to relate matrices $\mathbf{Q}(\mathbf{h})$, $\mathbf{u} \in \Gamma$, and $\mathbf{B}(\mathbf{t})$, $\mathbf{t} \in \Gamma$,we are going to consider certain sub-sets of Γ. Let $\mathbf{u} \wedge \mathbf{v}$ $[\mathbf{u} \vee \mathbf{v}]$ have components $\min\{u_l, v_l\}$ $[\max\{u_l, v_l\}]$, $l = 1, ..., L$, and put $\mathbf{t}L- = (\mathbf{t} - \mathbf{1}^L) \vee \mathbf{0}$ and $\mathbf{h}L+ = (\mathbf{h} + \mathbf{1}^L) \wedge \mathbf{u}$, as well as

$$\begin{cases} \sqcap(\mathbf{t}) = \{\mathbf{h} : \mathbf{t}L- \leq \mathbf{h} \leq \mathbf{t}\} \\ \sqcup(\mathbf{h}) = \{\mathbf{t} : \mathbf{h} \leq \mathbf{t} \leq \mathbf{h}L+\} \end{cases} . \tag{5.8}$$

in order to get, first

Proposition 2. $\mathbf{Q}(\mathbf{h}) = \sum_{\mathbf{t} \leq \mathbf{h}} \mathbf{B}(\mathbf{t})$, *as well as*

$$\mathbf{B}(\mathbf{t}) = \sum_{\mathbf{h} \in \sqcap(\mathbf{t})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \mathbf{Q}(\mathbf{h})$$

with $\mathrm{m}(\mathbf{h}, \mathbf{t})$ *the number of components of* $\mathbf{h}$ *that are smaller than the corresponding components of* $\mathbf{t}$.

Proof. The first part of the thesis follows from $\mathbf{Q}_l(h_l) = \sum_{t_l \leq h_l} \mathbf{B}_l(t_l)$, $h_l = 0, ..., u_l$, $l = 1, ..., L$, and from the distributive properties of the Kronecker product. Nextly, the first part of the thesis enables the writing of

$$\sum_{\mathbf{h} \in \sqcap(\mathbf{t})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \mathbf{Q}(\mathbf{h}) = \sum_{\mathbf{h} \in \sqcap(\mathbf{t})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \sum_{\mathbf{v} \leq \mathbf{h}} \mathbf{B}(\mathbf{v}).$$

If $\mathbf{v} < \mathbf{t}$, $\mathbf{B}(\mathbf{v})$ will enter in the terms of the $\mathbf{h}$ such that $\mathbf{v} \leq \mathbf{h} \leq \mathbf{t}$. Since it is possible to choose in $\binom{\mathrm{m}(\mathbf{v},\mathbf{t})}{a}$ ways a of the components of $\mathbf{h}$ that are smaller than the corresponding components of $\mathbf{t}$, this being the number of vectors $\mathbf{h}$ such that $\mathrm{m}(\mathbf{h}, \mathbf{t}) = a$. If $\mathrm{m}(\mathbf{v}, \mathbf{t}) > 0$ the coefficient of $\mathbf{B}(\mathbf{v})$ in the previous expression of $\sum_{\mathbf{h} \in \sqcap(\mathbf{t})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \mathbf{Q}(\mathbf{h})$ will be

$$\sum_{a=0}^{\mathrm{m}(\mathbf{v},\mathbf{t})} \binom{\mathrm{m}(\mathbf{v}, \mathbf{t})}{a} (-1)^a = 0$$

since $a = \mathrm{m}(\mathbf{h}, \mathbf{t})$. If $\mathrm{m}(\mathbf{h}, \mathbf{t}) = 0$, $\mathbf{v} = \mathbf{h} = \mathbf{t}$ and $\mathbf{B}(\mathbf{v})$ only enters, with coefficient one, in $\sum_{\mathbf{h} \in \sqcap(\mathbf{t})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \sum_{\mathbf{v} \leq \mathbf{h}} \mathbf{B}(\mathbf{v})$. Thus

$$\mathbf{B}(\mathbf{t}) = \sum_{\mathbf{v} \in \sqcap(\mathbf{t})} (-1)^{\mathrm{m}(\mathbf{v},\mathbf{t})} \mathbf{Q}(\mathbf{v}).$$

□

Corollary 3. *With* $\mathrm{S}(\mathbf{t}) = \|\mathbf{B}(\mathbf{t})\mathbf{y}\|^2$,

$$\mathrm{S}(\mathbf{t}) = \sum_{\mathbf{h} \in \sqcap(\mathbf{t})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \|\mathbf{Q}(\mathbf{h})\mathbf{y}\|^2 = \sum_{\mathbf{h} \in \sqcap(\mathbf{t})} \frac{(-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})}}{b(\mathbf{h})} \|\mathbf{X}'(\mathbf{h})\mathbf{y}\|^2$$

with $b(\mathbf{h}) = r \prod_{l=1}^{L} b_l(h_l)$.

Proof. The first part of the thesis follows directly from Proposition 2. As to the second part, since $\mathbf{X}(\mathbf{h})\mathbf{X}'(\mathbf{h}) = b(\mathbf{h})\mathbf{Q}(\mathbf{h})$ and orthogonal projection matrices are idempotent and symmetric,

$$\begin{aligned}\|\mathbf{Q}(\mathbf{h})\mathbf{y}\|^2 = \mathbf{y}'\mathbf{Q}'(\mathbf{h})\mathbf{Q}(\mathbf{h})\mathbf{y} = \mathbf{y}'\mathbf{Q}(\mathbf{h})\mathbf{y} \\ = \mathbf{y}'\left(\frac{1}{b(\mathbf{h})}\mathbf{X}(\mathbf{h})\mathbf{X}'(\mathbf{h})\right)\mathbf{y} = \frac{1}{b(\mathbf{h})}\|\mathbf{X}'(\mathbf{h})\mathbf{y}\|^2,\end{aligned}$$

the rest of the proof being straightforward. □

Besides matrices $\mathbf{B}(\mathbf{t})$, $\mathbf{t} \in \Gamma$, take

$$\mathbf{B} = \mathbf{I}_n - \mathbf{B}(\mathbf{u}) = \mathbf{I}_n - \sum_{\mathbf{t}\in\Gamma} \mathbf{B}(\mathbf{t}) \tag{5.9}$$

to obtain a basis for a Jordan algebra $\mathcal{A}$. This basis is the sole basis of $\mathcal{A}$ constituted by mutually orthogonal orthogonal projection matrices. This will be the principal basis of $\mathcal{A}$. Moreover matrices $\mathbf{Q}(\mathbf{h})$, $\mathbf{h} \in \Gamma$, and $\mathbf{I}_n$ belong to $\mathcal{A}$. $\mathrm{R}(\mathbf{B}(\mathbf{t}))$, with $\mathbf{t} \in \Gamma$, and $\mathrm{R}(\mathbf{B})$ constitute an orthogonal partition of $\mathbb{R}^n$ into mutually orthogonal sub-spaces.

The mean vector and covariance matrix of $\mathbf{y}$ will be, with $\sigma^2(\mathbf{0}) = 0$

$$\begin{cases} \mathbb{E}[\mathbf{y}] = \mathbf{1}^n\mu \\ \mathbb{V}[\mathbf{y}] = \sum_{\mathbf{h}\in\Gamma} \sigma^2(\mathbf{h})\mathbf{M}(\mathbf{h}) + \sigma^2\mathbf{I}_n = \sum_{\mathbf{t}\in\Gamma} \gamma(\mathbf{t})\mathbf{B}(\mathbf{t}) + \sigma^2\mathbf{B} \end{cases} \tag{5.10}$$

with

$$\gamma(\mathbf{t}) = \sigma^2 + \sum_{\mathbf{h}\geq\mathbf{t}} b(\mathbf{h})\sigma^2(\mathbf{h}), \tag{5.11}$$

since $\sigma^2(\mathbf{h})\mathbf{M}(\mathbf{h}) = b(\mathbf{h})\sigma^2(\mathbf{h}) \sum_{\mathbf{t}\leq\mathbf{h}} \mathbf{B}(\mathbf{t})$, $\mathbf{h} \in \Gamma$. It is the fact that $\mathbb{V}[\mathbf{y}]$ is a linear combination of the matrices in the basis of $\mathcal{A}$, which are mutually orthogonal, that gives the model it's orthogonality.

Still using model (5.3), it is easy to check that

$$\begin{cases} \gamma(1,0) = a_2(1)a_2(2)r\sigma^2(1,0) + a_2(2)r\sigma^2(1,1) + r\sigma^2(1,2) + \sigma^2 \\ \gamma(0,1) = a_1(1)a_2(2)r\sigma^2(0,1) + a_1(1)a_2(1)r\sigma^2(0,2) + a_2(2)r\sigma^2(1,1) \\ \qquad + r\sigma^2(1,2) + \sigma^2 \\ \gamma(0,2) = a_1(1)a_2(1)r\sigma^2(0,2) + a_2(2)r\sigma^2(1,1) + r\sigma^2(1,2) + \sigma^2 \\ \gamma(1,1) = a_2(2)r\sigma^2(1,1) + r\sigma^2(1,2) + \sigma^2 \\ \gamma(1,2) = r\sigma^2(1,2) + \sigma^2 \end{cases} . \quad (5.12)$$

Going over to the balanced case in which will be L groups of $u_1, ..., u_L$ factors, *i.e.*, the generalization of model (5.3). When $u_l > 1$ there is balanced nesting for the factors of the l-th group. This is, for each of the $a_l(1)$ levels of the first factor in the group there will be $a_l(2)$ levels of the second factor. Then, for each of the $a_l(1) \times a_l(2)$ levels of the second factor there will be $a_l(3)$ of the third factor, and so on.

$$\mathbf{X}_l(h_l) = \mathbf{I}_{c_l(h_l)} \otimes \mathbf{1}^{b_l(h_l)}, \quad h_l = 0, ..., u_l,\ l = 1, ..., L \quad (5.13)$$

with $c_l(h_l) = \prod_{t=1}^{h_l} a_l(t)$ and $b_l = {}^{c_l(u_l)}/_{c_l(h_l)}$, as well as

$$\mathbf{M}_l(h_l) = \mathbf{I}_{c_l(h_l)} \otimes \left(\frac{1}{b_l(h_l)}\mathbf{J}_{b_l(h_l)}\right) = \frac{1}{b_l(h_l)}\mathbf{Q}_l(h_l),$$
$$h_l = 0, ..., u_l,\ l = 1, ..., L \quad (5.14)$$

where $\mathbf{Q}_l(h_l) = \mathbf{I}_{c_l(h_l)} \otimes \left(\frac{1}{b_l(h_l)}\mathbf{J}_{b_l(h_l)}\right)$, $h_l = 0, ..., u_l$, $l = 1, ..., L$. Defining $\mathbf{K}_v = \mathbf{I}_v - \frac{1}{v}\mathbf{J}_v$ we also get

$$\mathbf{B}_l(t_l) = \mathbf{I}_{c_l(t_l-1)} \otimes \mathbf{K}_{a_l(t_l)} \otimes \left(\frac{1}{b_l(t_l)}\mathbf{J}_{b_l(t_l)}\right) = \mathbf{A}_l'(t_l)\mathbf{A}_l(t_l),$$
$$t_l = 1, ..., u_l,\ l = 1, ..., L \quad (5.15)$$

with

$$\mathbf{A}_l(t_l) = \frac{1}{\sqrt{b_l(t_l)}}\, \mathbf{I}_{c_l(t_l-1)} \otimes \mathbf{K}_{a_l(t_l)} \otimes \mathbf{1}^{b_l(t_l)\prime}, \quad t_l = 1, ..., u_l,\ l = 1, ..., L.$$

Putting $\mathbf{A}_l(0) = \mathbf{1}^{n_l\prime}$, $l = 1, ..., L$, and, if there are r replicates for each factor combination, define the following matrices

$$\mathbf{A}(\mathbf{t}) = \frac{1}{\sqrt{r}} \bigotimes_{l=1}^{L} \mathbf{A}_l(t_l) \otimes \mathbf{1}^{r\prime}, \quad \mathbf{t} \in \Gamma.$$

The sums of squares and corresponding degrees of freedom for a given factor or interaction will be

$$\begin{cases} \mathrm{S}(\mathbf{t}) = \|\mathbf{A}(\mathbf{t})\mathbf{y}\|^2 \quad , \mathbf{t} \in \Gamma \\ g(\mathbf{t}) = \prod_{l=1}^{L} g_l(t_l) = \prod_{l=1}^{L} \left(c_l(t_l) - c_l(t_l - 1)\right) \quad , \mathbf{t} \in \Gamma \end{cases} \tag{5.16}$$

with $c_l(-1) = 0$ for $l = 1, ..., L$.

Just for the record, consider another special case. Let $\mathbf{P}_l$ be a $n_l \times n_l$ orthogonal projection matrix whose first column vector is $\frac{1}{\sqrt{n_l}}\, \mathbf{1}^{n_l}$, $l = 1, .., L$. With $c_l(0) = 1$ and $c_l(h_l) < c_l(h_l + 1) \leq n_l$, $h_l = 1, ..., u_l - 1$, take $\mathbf{X}_l(h_l)$ to be constituted by the first $c_l(h_l)$ columns of $\mathbf{P}_l$, $l = 1, ..., L$. It is easy to see that the assumptions hold in this case which is distinct from the balanced one. Putting $\mathbf{A}(0) = \mathbf{X}_l'(0)$, $l = 1, ..., L$

$$\mathbf{X}_l(t_l) = \left[\begin{array}{cc} \mathbf{X}_l(t_l - 1) & \mathbf{A}_l'(t_l) \end{array} \right] \quad t_l = 1, ..., u_l,\ l = 1, ..., L \tag{5.17}$$

and $\mathbf{A}(\mathbf{t}) = \bigotimes_{l=1}^{L} \mathbf{A}_l(t_l) \otimes \mathbf{1}^{r\prime}$ expressions in (5.16) continue to hold.

5.2 Estimators

Note that for the models considered, the minimal complete statistics will be: $y_\bullet = \frac{1}{n}\mathbf{1}^{n\prime}\mathbf{y}$, $\mathrm{S}(\mathbf{t})$, with $\mathbf{t} \in \Gamma$, and $\mathrm{S} = \|\mathbf{B}\mathbf{y}\|^2$. The statistics $\mathrm{S}(\mathbf{t})$, with $\mathbf{t} \in \Gamma$, are given by the first expression in (5.16).

When $\mathbf{0} < \mathbf{t} \leq \mathbf{u}$, $\mathbf{B}(\mathbf{t})\mathbf{y}$ will be normal with null mean vector and covariance matrix $\gamma(\mathbf{t})\mathbf{B}(\mathbf{t})$, so that $\mathrm{S}(\mathbf{t})$ will be the product by $\gamma(\mathbf{t})$ of a central chi-square with $g(\mathbf{t}) = \mathrm{rank}(\mathbf{B}(\mathbf{t}))$ degrees of freedom, denoted by $\mathrm{S}(\mathbf{t}) \sim \gamma(\mathbf{t})\chi^2_{(g(\mathbf{t}))}$. Likewise $\mathrm{S} \sim \sigma^2\chi^2_{(g)}$, with $g = n - r(\mathbf{u})$. Moreover,

$$\mathcal{F}(\mathbf{t}_1, \mathbf{t}_2) = \frac{g(\mathbf{t}_2)}{g(\mathbf{t}_1)} \cdot \frac{\mathrm{S}(\mathbf{t}_1)}{\mathrm{S}(\mathbf{t}_2)} \tag{5.18}$$

will be the product by $\dfrac{\gamma(\mathbf{t}_1)}{\gamma(\mathbf{t}_2)}$ of a random variable with central F distribution with $g(\mathbf{t}_1)$ and $g(\mathbf{t}_2)$ degrees of freedom. Whenever $g(\mathbf{t}_2) > 2$, $\mathbb{E}[\mathcal{F}(\mathbf{t}_1, \mathbf{t}_2)] = \dfrac{g(\mathbf{t}_2)}{g(\mathbf{t}_2) - 2} \cdot \dfrac{\gamma(\mathbf{t}_1)}{\gamma(\mathbf{t}_2)}$ and so we have the UMVUE

$$\widehat{\left(\frac{\gamma(\mathbf{t}_1)}{\gamma(\mathbf{t}_2)}\right)} = \frac{g(\mathbf{t}_2) - 2}{g(\mathbf{t}_2)} \cdot \frac{\mathrm{S}(\mathbf{t}_1)}{\mathrm{S}(\mathbf{t}_2)}. \tag{5.19}$$

Later on these estimators will be used. If $f_{q:r,s}$ is the quantile, for probability q, of the central F distribution with r and s degrees of freedom we have, for $\dfrac{\gamma(\mathbf{t}_1)}{\gamma(\mathbf{t}_2)}$, the $(1-q) \times 100\%$ confidence interval $\left[\dfrac{\mathcal{F}(\mathbf{t}_1, \mathbf{t}_2)}{f_{1-\frac{q}{2}:g_1,g_2}}, \dfrac{\mathcal{F}(\mathbf{t}_1, \mathbf{t}_2)}{f_{\frac{q}{2}:g_1,g_2}}\right]$ with $g_i = g(\mathbf{t}_i)$, $i = 1, 2$. The $1 - q$ level upper bound will likewise be $\dfrac{\mathcal{F}(\mathbf{t}_1, \mathbf{t}_2)}{f_{1-q:g_1,g_2}}$

The sets $\sqcap(\mathbf{t})$ and $\sqcup(\mathbf{t})$ were defined in the preceding section and established Proposition 2.

Proposition 3. *Whenever* $\mathrm{m}(\mathbf{h}) = \mathrm{m}(\mathbf{h}, \mathbf{u}) > 0$,

$$\sigma^2(\mathbf{h}) = \frac{1}{b(\mathbf{h})} \sum_{\mathbf{t} \in \sqcup(\mathbf{h})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \gamma(\mathbf{t}).$$

Proof. From the definition of $\gamma(\mathbf{t})$,

$$\sum_{\mathbf{t} \in \sqcup(\mathbf{h})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \gamma(\mathbf{t}) = \sum_{\mathbf{t} \in \sqcup(\mathbf{h})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})} \left(\sigma^2 + \sum_{\mathbf{v} \geq \mathbf{t}} b(\mathbf{v})\sigma^2(\mathbf{v})\right).$$

Now, in $\sqcup(\mathbf{h})$ we have $\binom{\mathrm{m}(\mathbf{h})}{z}$ (with z ranging from 0 to $\mathrm{m}(\mathbf{h})$) vectors $\mathbf{t}$ such that $\mathrm{m}(\mathbf{h},\mathbf{t}) = z$, since this is the number of choices of z components of $\mathbf{h}$, from the $\mathrm{m}(\mathbf{h})$ components that are smaller than those of $\mathbf{u}$. So, these components of $\mathbf{h}$ can be increased by 1, the resulting vector still belonging to $\sqcup(\mathbf{h})$. Thus, the coefficient of σ^2 will be $\sum_{z=0}^{\mathrm{m}(\mathbf{h})} \binom{\mathrm{m}(\mathbf{h})}{z}(-1)^z = 0$. Likewise, if $\mathbf{v} > \mathbf{h}$, $\sigma^2(\mathbf{v})$ appears in the terms associated with the $\mathbf{t} \in \sqcup(\mathbf{h})$ such that $\mathbf{t} \leq \mathbf{v}$. It is possible to reason as before to show that there are $\binom{\mathrm{m}(\mathbf{h},\mathbf{v})}{z}$, $z = 0, ..., \mathrm{m}(\mathbf{h},\mathbf{v})$, such vectors $\mathbf{t} \in \sqcup(\mathbf{h})$ with $\mathrm{m}(\mathbf{h},\mathbf{v}) = z$, and so, the coefficient of $b(\mathbf{h})\sigma^2(\mathbf{h})$ will be $\sum_{z=0}^{\mathrm{m}(\mathbf{h},\mathbf{v})} \binom{\mathrm{m}(\mathbf{h},\mathbf{v})}{z}(-1)^z = 0$. Lastly when $\mathbf{v} = \mathbf{h}$, $b(\mathbf{v})\sigma^2(\mathbf{v}) = b(\mathbf{h})\sigma^2(\mathbf{h})$ only enters in the term associated with $\mathbf{t} = \mathbf{h}$ with coefficient 1, so that $\sum_{\mathbf{t}\in\sqcup(\mathbf{h})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})}\gamma(\mathbf{t}) = b(\mathbf{h})\sigma^2(\mathbf{h})$ and the thesis is established. □

Corollary 4. $\hat{\gamma}(\mathbf{t}) = \dfrac{\mathrm{S}(\mathbf{t})}{g(\mathbf{t})}$, $\mathbf{t} \in \Gamma$ *and, if* $\mathrm{m}(\mathbf{h}) > 0$,

$$\hat{\sigma}^2(\mathbf{h}) = \frac{1}{b(\mathbf{h})} \sum_{\mathbf{t}\in\sqcup(\mathbf{h})} (-1)^{\mathrm{m}(\mathbf{h},\mathbf{t})}\hat{\gamma}(\mathbf{t})$$

will be UMVUE.

Moreover, for σ^2 and $\sigma^2(\mathbf{u})$ there are the UMVUE

$$\begin{cases} \hat{\sigma}^2 = \dfrac{\mathrm{S}}{g} \\ \hat{\sigma}^2(\mathbf{u}) = \dfrac{1}{b(\mathbf{u})}(\hat{\gamma}(\mathbf{u}) - \hat{\sigma}^2), \end{cases} \tag{5.20}$$

since $\mathrm{S} \sim \sigma^2\chi^2_{(g)}$ and $\gamma(\mathbf{u}) = \sigma^2 + b(\mathbf{u})\sigma^2(\mathbf{u})$. Note that $\mathrm{m}(\mathbf{h}) = 0$ if and only if $\mathbf{h} = \mathbf{u}$.

Let $\sqcup(\mathbf{h})_+$ and $\sqcup(\mathbf{h})_-$ be the vectors of $\sqcup(\mathbf{h})$ for which $\mathrm{m}(\mathbf{h},\mathbf{t})$ is even or is odd. If $\mathrm{m}(\mathbf{h}) > 0$, with $\sigma^2(\mathbf{h})_+ = \dfrac{1}{b(\mathbf{h})} \sum_{\mathbf{t}\in\sqcup(\mathbf{h})_+} \gamma(\mathbf{t})$ and $\sigma^2(\mathbf{h})_- = \dfrac{1}{b(\mathbf{h})} \sum_{\mathbf{t}\in\sqcup(\mathbf{h})_-} \gamma(\mathbf{t})$ we have

$$\sigma^2(\mathbf{h}) = \sigma^2(\mathbf{h})_+ - \sigma^2(\mathbf{h})_-. \tag{5.21}$$

We will also have the UMVUE estimators

$$\begin{cases} \hat{\sigma}^2(\mathbf{h})_+ = \dfrac{1}{b(\mathbf{h})} \displaystyle\sum_{\mathbf{t}\in\sqcup(\mathbf{h})_+} \hat{\gamma}(\mathbf{t}) \\ \hat{\sigma}^2(\mathbf{h})_- = \dfrac{1}{b(\mathbf{h})} \displaystyle\sum_{\mathbf{t}\in\sqcup(\mathbf{h})_-} \hat{\gamma}(\mathbf{t}). \end{cases} \tag{5.22}$$

Besides through completeness of the statistics, it will now be shown through an alternative method that the estimators obtained are efficient in the Fisher sense.

It is proved in [3] that

$$\det\left(\mathbb{V}[\mathbf{y}]\right) = \left(\prod_{\mathbf{t}\in\Gamma} \gamma(\mathbf{t})^{g(\mathbf{t})}\right) (\sigma^2)^g, \tag{5.23}$$

so that, if $\boldsymbol{\eta}$ has components μ, $\gamma(\mathbf{t})$, $\mathbf{0} < \mathbf{t} \leq \mathbf{u}$, and σ^2 one gets the log-likelihood

$$l(\boldsymbol{\eta}) = -\frac{n}{2\gamma(\mathbf{0})}(y_\bullet - \mu)^2 - \frac{1}{2}\sum_{\mathbf{0}\leq\mathbf{t}\leq\mathbf{u}} \frac{\mathrm{S}(\mathbf{t})}{\gamma(\mathbf{t})} - \frac{\mathrm{S}}{2\sigma^2} - \frac{n}{\log}(2\pi) \\ - \frac{1}{2}\sum_{\mathbf{t}\in\Gamma} g(\mathbf{t}) \log\gamma(\mathbf{t}) - \frac{g}{2}\log\sigma^2. \tag{5.24}$$

$\gamma(\mathbf{0})$ was omitted from the components of $\boldsymbol{\eta}$ since this parameter is clearly a function of the $\gamma(\mathbf{t})$, $\mathbf{0} < \mathbf{t} \leq \mathbf{u}$. It is now straightforward to show that $\tilde{\mu} = y_\bullet$, $\tilde{\gamma}(\mathbf{t})$, $\mathbf{0} < \mathbf{t} \leq \mathbf{u}$, and $\tilde{\sigma}^2$ are maximum likelihood estimators and that their Fisher information matrix

$$\mathbf{Z}(\boldsymbol{\eta}) = \mathbb{E}\big[\mathrm{grad}(l(\boldsymbol{\eta}))\,(\mathrm{grad}(l(\boldsymbol{\eta})))'\big] \tag{5.25}$$

is the diagonal matrix with principal elements $\dfrac{n}{2\gamma(\mathbf{0})}$, $\dfrac{n}{2\gamma(\mathbf{0})}$, $\dfrac{g(\mathbf{t})}{2\gamma(\mathbf{t})}$, $\mathbf{0} < \mathbf{t} \leq \mathbf{u}$, and $\dfrac{g}{2\sigma^2}$. Since these are the inverses of the variances of these estimators we see that the Cramer-Rao lower bound for the variance of unbiased estimators is attained for these estimators,

$$\mathbb{V}[\hat{\boldsymbol{\eta}}]^{-1} = \mathbf{Z}(\boldsymbol{\eta}). \tag{5.26}$$

Let $\boldsymbol{\theta}$ have components μ, $\sigma^2(\mathbf{t})$, $\mathbf{0} < \mathbf{t} \leq \mathbf{u}$, and σ^2. Then $\boldsymbol{\theta} = \mathbf{K}\boldsymbol{\eta}$ with $\mathbf{K}$ a regular matrix so that $\boldsymbol{\eta} = \mathbf{K}^{-1}\boldsymbol{\theta}$ and, using the new parameters, we have for the gradient

of the log-likelihood

$$\text{grad}(l(\boldsymbol{\theta})) = (\mathbf{K}^{-1})'\text{grad}(l(\boldsymbol{\eta}))\,. \tag{5.27}$$

Thus the Fisher information matrix for the new parameters will be

$$\mathbf{Z}(\boldsymbol{\theta}) = (\mathbf{K}^{-1})'\mathbf{Z}(\boldsymbol{\eta})\mathbf{K}^{-1} \tag{5.28}$$

and, with $\boldsymbol{\theta} = \mathbf{K}\boldsymbol{\eta}$, we have the covariance matrix

$$\mathbb{V}[\boldsymbol{\theta}] = \mathbf{K}\mathbb{V}[\tilde{\boldsymbol{\eta}}]\,\mathbf{K}' = \mathbf{K}\mathbf{Z}(\boldsymbol{\eta})\mathbf{K}' = \mathbf{Z}(\boldsymbol{\theta}). \tag{5.29}$$

According to (5.26), to (5.29) and to the invariance principle of the maximum likelihood estimators it is established

Proposition 4. *$\hat{\boldsymbol{\eta}}$ and $\boldsymbol{\theta}$ are maximum likelihood estimators for which the Cramer-Rao lower bounds for covariance matrices are attained.*

5.3 Hypothesis testing

5.3.1 F Tests

In this section we derive F tests for the hypothesis

$$\begin{cases} H_0(\mathbf{h}):\ \sigma^2(\mathbf{h}) = 0, & \mathrm{m}(\mathbf{h}) \leq 1 \\ \overline{H}_0(\mathbf{h}):\ \gamma(\mathbf{h}) = \sigma^2, & \mathbf{h} \in \Gamma. \end{cases} \tag{5.30}$$

We start with $H_0(\mathbf{u})$. Since $\mathrm{S}(\mathbf{u}) \sim (r\sigma^2(\mathbf{u}) + \sigma^2)\,\chi^2_{(g(\mathbf{u}))}$ independent from $\mathrm{S} \sim \sigma^2\chi^2_{(g)}$. The test statistic

$$\mathcal{F}(\mathbf{u}) = \frac{g}{g(\mathbf{u})} \cdot \frac{\mathrm{S}(\mathbf{u})}{\mathrm{S}} \tag{5.31}$$

will have central F distribution with $g(\mathbf{u})$ and g degrees of freedom when $H_0(\mathbf{u})$ holds. Likewise, if $\mathrm{m}(\mathbf{h}) = 1$, $\sqcup(\mathbf{h})_+ = \{\mathbf{h}\}$ and $\sqcup(\mathbf{h})_- = \{\mathbf{u}\}$. Since

$$\mathrm{S}(\mathbf{h}) \sim \left(rb(\mathbf{h})\sigma^2(\mathbf{h}) + r\sigma^2(\mathbf{u}) + \sigma^2\right)\chi^2_{(g(\mathbf{h}))}$$

independent from $\mathrm{S}(\mathbf{u})$,

$$\mathcal{F}(\mathbf{u}) = \frac{g(\mathbf{u})}{g(\mathbf{h})} \cdot \frac{\mathrm{S}(\mathbf{h})}{\mathrm{S}(\mathbf{u})} \tag{5.32}$$

will have a central F distribution with $g(\mathbf{h})$ and $g(\mathbf{u})$ degrees of freedom, when $H_0(\mathbf{h})$ holds.

When $\mathrm{m}(\mathbf{u}) > 1$ there is more than one vector in $\sqcup(\mathbf{h})_+$ and in $\sqcup(\mathbf{h})_-$ so that we cannot pursue this straightforward derivation of F tests is not possible. In the next section we will consider how to test $H_0(\mathbf{h})$ when $\mathrm{m}(\mathbf{h}) > 1$. The possibilities are, see [7], using either Bartlett-Scheffé F tests or the Satterthwaite approximation of the distribution of F statistics. These approaches will not be considered, since the first one discards information, for instance see [7] pg. 43, while the second may lead to difficulties (see [7] pg. 40, in the control of test size).

No such problems arise wile testing $\overline{H}_0(\mathbf{h})$, $\mathbf{h} \in \Gamma$. Since $\mathrm{S}(\mathbf{h}) \sim \gamma(\mathbf{h})\chi^2_{(g(\mathbf{h}))}$ independent of $\mathrm{S} \sim \sigma^2\chi^2_{(g)}$ one gets the F statistic

$$\overline{\mathcal{F}(\mathbf{h})} = \frac{g}{g(\mathbf{h})} \cdot \frac{\mathrm{S}(\mathbf{h})}{\mathrm{S}}, \quad \mathbf{h} \in \Gamma \tag{5.33}$$

with $g(\mathbf{h})$ and g degrees of freedom, when $\overline{H}_0(\mathbf{h})$ holds.

As a parting remark note that $\overline{H}_0(\mathbf{h})$ holds if and only if the $H_0(\mathbf{v})$, with $\mathbf{h} \leq \mathbf{v} \leq \mathbf{u}$, hold. From Theorem 68 it follows that this statistic can have F distribution if and only if $\mathrm{m}(\mathbf{h}) > 1$ we will have to consider generalized F tests. We now present such tests.

5.3.2 Generalized F Tests

Test Statistics

These statistics introduced by Michalski and Zmyślony, see [12] and [13], can be written as

$$\mathcal{F}(\mathbf{h}) = \frac{\hat{\sigma}^2(\mathbf{h})_+}{\hat{\sigma}^2(\mathbf{h})_-} \tag{5.34}$$

with $\hat{\sigma}^2(\mathbf{h})_+$ and $\hat{\sigma}^2(\mathbf{h})_-$ defined in (5.22).

It is interesting to observe that $\sigma^2(\mathbf{h}) = \sigma^2(\mathbf{h})_+ - \sigma^2(\mathbf{h})_- \geq 0$, thus, with $\lambda(\mathbf{h}) = \frac{\sigma^2(\mathbf{h})_+}{\sigma^2(\mathbf{h})_-}$, both $H_0(\mathbf{h})$ and the corresponding alternative may be written as

$$\begin{cases} H_0(\mathbf{h}) : \lambda(\mathbf{h}) = 1 \\ H_1(\mathbf{h}) : \lambda(\mathbf{h}) > 1. \end{cases} \tag{5.35}$$

Consider a two-dimensional presentation of the behavior of the test statistic (Figure 1). Let f be the critical value.

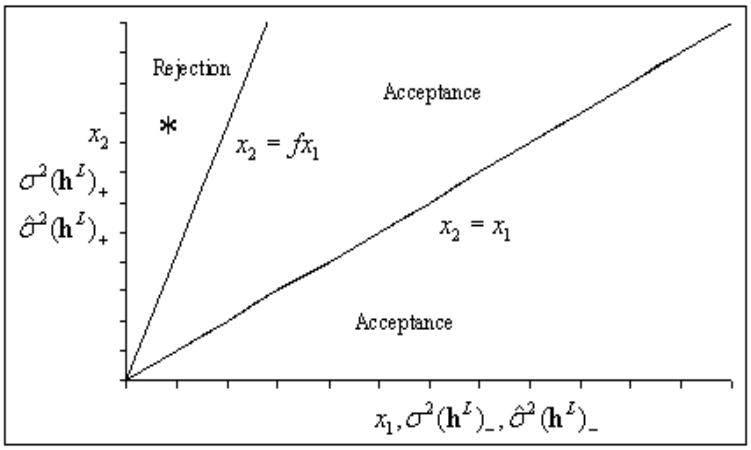

Fig. 5.1: Acceptance and Rejection Regions

In this presentation lie the points, with coordinates (x_1, x_2), $(\sigma^2(\mathbf{h})_-, \sigma^2(\mathbf{h})_+)$ and $(\hat{\sigma}^2(\mathbf{h})_-, \hat{\sigma}^2(\mathbf{h})_+)$. The semi-straight line $x_1 = fx_2$ separates the rejection from the acceptance region, while (*) indicates a clear alternative to $H_0(\sigma^2(\mathbf{h}))$. Since $\hat{\sigma}^2(\mathbf{h})_+$ and $\hat{\sigma}^2(\mathbf{h})_-$ are UMVUE of $\sigma^2(\mathbf{h})_+$ and $\sigma^2(\mathbf{h})_+$ it is expected that, when this alternative holds, $(\hat{\sigma}^2(\mathbf{h})_-, \hat{\sigma}^2(\mathbf{h})_+)$ will be close to (*), thus inside the rejection region. This heuristic argument clearly points towards the use of the test statistic $\mathcal{F}(\mathbf{h})$.

After the figure, a closer look at $\mathcal{F}(\mathbf{h})$ is necessary in order to be able to use it properly. To lighten the writing put $m = \mathrm{m}(\mathbf{h})$ taking $w = 2^{m-1}$, $\sqcup(\mathbf{h})_+ =$

$\{\mathbf{t}_1, ..., \mathbf{t}_w\}$ and $\sqcup(\mathbf{h})_- = \{\mathbf{t}_{w+1}, ..., \mathbf{t}_{2w}\}$. Let $\mathbf{g}_1$ and $\mathbf{g}_2$ have components $g_{1,j} = g(\mathbf{t}_j)$ and $g_{2,j} = g(\mathbf{t}_{j+w})$, $j = 1, ..., w$, where we can assume that $g_{i,1} = \min\{g_{i,1}, ..., g_{i,w}\}$ and $g_{i,w} = \max\{g_{i,1}, ..., g_{i,w}\}$, $i = 1, 2$. Then

$$\mathcal{F}(\mathbf{h}) = \lambda(\mathbf{h}) \frac{\sum_{j=1}^{w} \frac{p_{1,j}}{g_{1,j}} \chi^2_{(g_{1,j})}}{\sum_{j=1}^{w} \frac{p_{2,j}}{g_{2,j}} \chi^2_{(g_{2,j})}}, \tag{5.36}$$

with $p_{1,j} = \dfrac{\gamma(\mathbf{t}_j)}{\sigma^2(\mathbf{h})_+}$ and $p_{2,j} = \dfrac{\gamma(\mathbf{t}_{j+w})}{\sigma^2(\mathbf{h})_-}$, $j = 1, ..., w$, and, as before, $\lambda(\mathbf{h}) = \dfrac{\sigma^2(\mathbf{h})_+}{\sigma^2(\mathbf{h})_-}$. Since $\sum_{j=1}^{w} p_{1,j} = \sum_{j=1}^{w} p_{2,j} = 1$, $\mathcal{F}(\mathbf{h})$ will be the product by $\lambda(\mathbf{h})$ of the quotient of two convex combinations of independent central chi-squares divided by their degrees of freedom. Thus, it is natural to consider $\mathcal{F}(\mathbf{h})$ as a generalized F statistic. The $p_{1,j}$ and $p_{2,j}$, $j = 1, ..., w$ will be the components of vectors $\mathbf{p}_1$ and $\mathbf{p}_2$. These vectors will be nuisance parameters. Since

$$\begin{cases} p_{1,j} = \dfrac{\frac{\gamma(\mathbf{t}_j)}{\gamma(\mathbf{t}_w)}}{\frac{\sum_{j'=1}^{w} \gamma(\mathbf{t}_{j'})}{\gamma(\mathbf{t}_w)}}, & j = 1, ..., w \\ p_{2,j} = \dfrac{\frac{\gamma(\mathbf{t}_{j+w})}{\gamma(\mathbf{t}_{2w})}}{\frac{\sum_{j'=1}^{w} \gamma(\mathbf{t}_{j'+w})}{\gamma(\mathbf{t}_{2w})}}, & j = 1, ..., w, \end{cases} \tag{5.37}$$

$\left(\widehat{\dfrac{\gamma(\mathbf{t}_j)}{\gamma(\mathbf{t}_w)}}\right)$ and $\left(\widehat{\dfrac{\gamma(\mathbf{t}_{j+w})}{\gamma(\mathbf{t}_{2w})}}\right)$ can be used to estimate $p_{1,j}$ and $p_{2,j}$, $j = 1, ..., w$. Now, when $H_0(\mathbf{h})$ holds,

$$\mathcal{F}(\mathbf{h}) = \frac{Y_1}{Y_2}, \tag{5.38}$$

with $Y_1 = \sum_{j=1}^{w} p_{i,j} \dfrac{\chi^2_{(g_{i,j})}}{g_{i,j}}$, $i = 1, 2$. Since the Y_1 and Y_2 are independent with mean value 1, and the partial derivatives of $\dfrac{Y_1}{Y_2}$, at point $(1, 1)$, are 1 and -1, one gets, when $H_0(\mathbf{h})$ holds

$$\mathcal{F}(\mathbf{h}) \approx 1 + (Y_1 - 1) - (Y_2 - 1), \tag{5.39}$$

so that, according to the independence of Y_1 and Y_2

$$\begin{cases} \mathbb{E}[\mathcal{F}(\mathbf{h})] \approx 1 \\ \mathbb{V}[\mathcal{F}(\mathbf{h})] \approx \mathbb{V}[Y_1] + \mathbb{V}[Y_2] \end{cases}. \tag{5.40}$$

Also

$$\mathbb{V}[Y_i] = 2\sum_{j=1}^{w} \frac{p_{i,j}^2}{g_{i,j}}, \quad i = 1, 2. \tag{5.41}$$

Looking at the Figure 2 one is led to think that, when $H_0(\mathbf{h})$ holds, the probability

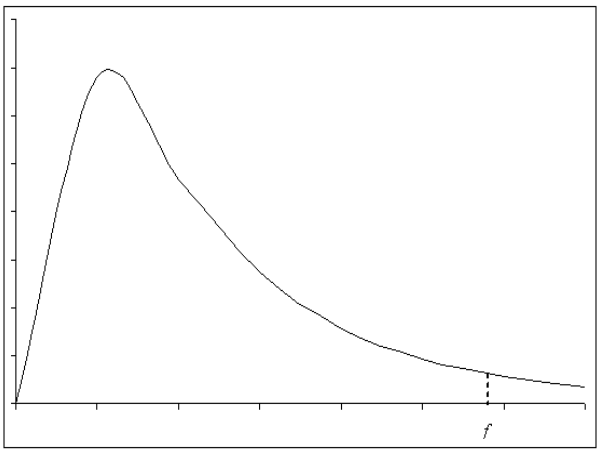

Fig. 5.2: Critical Value

of rejection increases with $\sum_{i=1}^{2} \mathbb{V}[Y_i]$. Putting $g_i = \sum_{j=1}^{w} g_{i,j}$ and $u_{i,j} = p_{i,j} - \frac{1}{g_i} g_{i,j}$, $j = 1, ..., w$, $i = 1, 2$,

$$\mathbb{V}[Y_i] = \frac{2}{g_i} + 2\sum_{j=1}^{w} \frac{u_{i,j}^2}{g_{i,j}}, \quad i = 1, 2 \tag{5.42}$$

so that, according to (5.41) and (5.42), $\frac{2}{g_i} < \mathbb{V}[Y_i] < \frac{2}{g_{i,1}}$, $i = 1, 2$. It is interesting to observe that when $\mathbb{V}[Y_i]$ attains it´s lower bound, Y_i is a central chi-square with g_i degrees of freedom divided by this number and when $\mathbb{V}[Y_i]$ attains it´s upper bound, Y_i is a central chi-square with $g_{i,1}$ degrees of freedom divided by the same number. Thus, when both lower or both upper bounds are attained, $\mathcal{F}(\mathbf{h})$ has, when $H_0(\mathbf{h})$ holds, the central F distribution with g_1 and g_2 or $g_{1,1}$ and $g_{2,1}$ degrees of freedom.

In the balanced case there are some restrictions on the $p_{i,j}$, $j = 1, ..., w$, $i = 1, 2$. When $H_0(\mathbf{h})$ holds $\sigma^2(\mathbf{h})_+ = \sigma^2(\mathbf{h})_-$, and so $\gamma(\mathbf{h}) = \gamma(\mathbf{t}_1) = \max\{\gamma(\mathbf{t}_1), ..., \gamma(\mathbf{t}_w)\} \geq \max\{\gamma(\mathbf{t}_{w+1}), ..., \gamma(\mathbf{t}_{2w})\}$, so that

$$\begin{cases} p_{1,1} = \max\{p_{1,1}, ..., p_{1,w}\} \geq \dfrac{1}{w} > \dfrac{g(\mathbf{h})}{g_1} \\ p_{1,1} \geq \max\{p_{2,1}, ..., p_{2,w}\} \end{cases} . \tag{5.43}$$

When $w = 2$ one can take $p_{1,1} = p_1$, $p_{1,2} = 1 - p_1$, $p_{2,1} = p_2$ and $p_{2,2} = 1 - p_2$ to get

$$\max\{p_2, 1 - p_2\} \leq p_1. \tag{5.44}$$

Getting back to the general case in order to control the first type error, one can

- place under the least favorable situation in which, as we saw, we have an F test with $g_{1,1}$ and $g_{2,1}$ degrees of freedom;
- use more completely the set of complete sufficient statistics. One then obtains, besides the test statistic, estimates to the nuisance parameters. In this case one will be carrying out an adaptative procedure similar to the one carried out in [1] and [35]. Likewise one may think we are using ancillary statistics, see [26].

Restricting to the balanced case, the goal is to see when one or both of the evenness conditions hold. Since

$$g(\mathbf{h}) = \prod_{l=1}^{L} g_l(h_l) \tag{5.45}$$

it suffices that one of the $g_l(h_l)$, $l = 1, ..., L$, is even for $g(\mathbf{h})$ to be even. Besides this, with $0 \leq h_l \leq u_l$,

$$\begin{cases} g_l(0) = 1, & l = 1, ..., L \\ g_l(h_l) = \left(\prod_{t_l=1}^{h_l - 1} a_l(t_l) \right) (a_l(h_l) - 1), & h_l = 1, ..., u_l,\ l = 1, ..., L, \end{cases} \tag{5.46}$$

were $a_l(h_l)$ is the number of factors of the h_l-th factor in the l-th group of nested factors, and $\prod_{t=1}^{0} a_l(t) = 1$, $l = 1, , ..., L$. Thus if $g_l(h_l)$, with $l \geq 1$, is odd, $a_l(h_l)$ must be even as well as $g_l(h_*)$ with $h_l \leq h_* \leq u_l$. Thus in the sequence $g_l(1), ..., g_l(u_l)$ there can be only one odd number. This observation clearly points towards evenness conditions holding quite often.Consider the sets of indexes $\mathcal{C}(\mathbf{h}) = \{j : h_j < u_j\}$ and $\mathcal{D}(\mathbf{h}) = \{l : h_l > 0\}$, putting $u(\mathbf{h}) = \#(\mathcal{C}(\mathbf{h}))$ and $o(\mathbf{h}) = \#(\mathcal{D}(\mathbf{h})) - 1$. With $q_l(h_l)$ the number of even numbers in the pair $(g_l(h_l), g_l(h_l + 1))$, one get $q_l(0) \leq 1$ since $g_l(0) = 1$ so that only $g_l(1)$ may be even. Put also

$$\begin{cases} k(\mathbf{h}) = \max\limits_{l \in \mathcal{C}(\mathbf{h})} q_l(h_l) \\ t(\mathbf{h}) = \min\limits_{l \in \mathcal{C}(\mathbf{h})} q_l(h_l). \end{cases} \tag{5.47}$$

Establish

Proposition 5. *If $k(\mathbf{h}) = 2$ or if there exists $l' \notin \mathcal{C}(\mathbf{h})$ such that $g_{l'}(u_{l'})$ is even both evenness conditions are satisfied.*

Proof. If $k(\mathbf{h}) = 2$ there will be $l \in \mathcal{C}(\mathbf{h})$ such that $g_l(h_l)$ and $g_l(h_l + 1)$ are even. Since either of these will be a factor in $g(\mathbf{h}_*)$ whatever $\mathbf{h}_* \in \sqcup(\mathbf{h})$, the first part of the thesis is established. The second part follows from $g_{l'}(u_{l'})$ being, if it exists, a factor from whatever $g(\mathbf{h}_*)$ with $\mathbf{h}_* \in \sqcup(\mathbf{h})$. □

Nextly one gets

Proposition 6. *If $k(\mathbf{h}) = t(\mathbf{h}) = 1$ one of the two evenness conditions holds.*

Proof. If $k(\mathbf{h}) = t(\mathbf{h}) = 1$, whenever $l \in \mathcal{C}(\mathbf{h})$, $g_l(h_l))$ or $g_l(h_l + 1)$ is even but not both since $k(\mathbf{h}) = 1$. For $g(\mathbf{h}L*)$, with $\mathbf{h}L* \in \sqcup(\mathbf{h})$, to be odd one must have, for $l \in \mathcal{C}(\mathbf{h})$, $h_{*l} = h_l$ when $g_l(h_l)$ is odd or $h_l + 1$ when $g_l(h_l)$ is even. Thus there is at most one vector $\mathbf{h}_0 \in \sqcup(\mathbf{h})$ such that $g(\mathbf{h}_0)$ is odd.

Then, if $\mathbf{h}_0 \in \sqcup(\mathbf{h})_+$ the first evenness condition holds and if $\sqcup(\mathbf{h})_-$ the second evenness condition holds. □

Corollary 5. *If $o(\mathbf{h}) = L - 1$ at least one of the evenness conditions holds.*

Proof. If $o(\mathbf{h}) = L-1$ we have $h_l \geq 1$, $l = 1, ..., L$ and so $q_l(h_l) \geq 1$, for all $l \in \mathcal{C}(\mathbf{h})$, since, when $g_l(h_l)$ is odd $g_l(h_l + 1)$ is even, thus $t(\mathbf{h}) \geq 1$ and the thesis follows from Propositions 5 and 6. □

Corollary 6. *If $a_{l,1}$ is odd, $l = 1, ..., L$, at least one of the evenness conditions holds for all $\mathbf{h} \in \Gamma$.*

Proof. $q_l(0) = 1$, $l = 1, ..., L$, as well as $q_l(h_l) \geq 1$, $h_l = 1, ..., u_{l-1}$. Thus $t(\mathbf{h}) \geq 1$ for all $\mathbf{h} \in \Gamma$. To complete the proof, apply Propositions 5 and 6. □

These results show that quite often one can get exact expressions for $F(z|\mathbf{c}_1, \mathbf{c}_2, \mathbf{g}_1, \mathbf{g}_2)$. This is interesting in itself and enables accuracy checking applying Monte-Carlo methods to evaluate $F(z|\mathbf{c}_1, \mathbf{c}_2, \mathbf{g}_1, \mathbf{g}_2)$ (see [1]).

6. FINAL REMARKS

Although some proposals were made regarding pendent issues on linear models, specially in model structure, estimation of fixed effects and variance components and hypothesis testing, a lot of questions remain unanswered.

- When BLUE are not obtainable, what to do? How to derive explicit formulas for LBLUE (locally best unbiased estimators)? An analogous problem (but somewhat more complex) arises with variance components.

- What approach to take with general unbalanced data? The results obtained work under the assumption of model orthogonality, but a systematic approach for unbalanced models is needed. A possible solution is the use of U statistics in the estimation of parameters.

- Normality. At what extend is it necessary? And can it be attained asymptotically for general linear models and if so, in what conditions?

- It is logical to think of the generalization of model algebraic structure and the generalized F test to the multivariate case. It is a pendent subject.

These are merely some suggestions for future research. It is a truism that the more answers are obtained the more questions arise.

BIBLIOGRAPHY

[1] Miguel Fonseca, João Tiago Mexia, and Roman Zmyślony. Exact distribution for the generalized F tests. *Discuss. Math., Probab. Stat.*, 22(1-2):37–51, 2002.

[2] Miguel Fonseca, João Tiago Mexia, and Roman Zmyślony. Estimating and Testing of Variance Components: an Application to a Grapevine Experiment. *Biometrical Letters*, 40:1–7, 2003.

[3] Miguel Fonseca, João Tiago Mexia, and Roman Zmyślony. Estimators and tests for variance components in cross nested orthogonal designs. *Discuss. Math., Probab. Stat.*, 23(2):175–201, 2003.

[4] Miguel Fonseca, João Tiago Mexia, and Roman Zmyślony. Binary operations on Jordan algebras and orthogonal normal models. *Linear Algebra Appl.*, 417(1):75–86, 2006.

[5] S. Gnot, W. Klonecki, and R. Zmyslony. Best unbiased linear estimation, a coordinate free approach. *Probab. Math. Stat.*, 1:1–13, 1980.

[6] P. Jordan, J. von Neumann, and Eugene P. Wigner. On an algebraic generalization of the quantum mechanical formalism. 1934.

[7] André I. Khuri, Thomas Mathew, and Bimal K. Sinha. *Statistical tests for mixed linear models.* Wiley Series in Probability and Mathematical Statistics, Applied Section. New York, NY: Wiley., 1998.

[8] Lynn Roy LaMotte. A canonical form for the general linear model. *Ann. Stat.*, 5:787–789, 1977.

[9] E.L. Lehmann. *Testing statistical hypotheses. Reprint of the 2nd ed. publ. by Wiley 1986.* New York, NY: Springer. , 1997.

[10] E.L. Lehmann and George Casella. *Theory of point estimation. 2nd ed.* Springer Texts in Statistics. New York, NY: Springer. , 1998.

[11] James D. Malley. *Statistical applications of Jordan algebras.* Lecture Notes in Statistics (Springer). 91. New York, NY: Springer- Verlag. , 1994.

[12] Andrzej Michalski and Roman Zmyślony. Testing hypotheses for variance components in mixed linear models. *Statistics*, 27(3-4):297–310, 1996.

[13] Andrzej Michalski and Roman Zmyślony. Testing hypotheses for linear functions of parameters in mixed linear models. *Tatra Mt. Math. Publ.*, 17:103–110, 1999.

[14] Alexander M. Mood, Franklin A. Graybill, and Duane C. Boes. *Introduction to the theory of statistics. 3rd ed.* McGraw-Hill Series in Probability and Statistics. New York etc.: McGraw- Hill Book Company. , 1974.

[15] C.Radhakrishna Rao and M.Bhaskara Rao. *Matrix algebra and its applications to statistics and econometrics.* Singapore: World Scientific Publishing. , 1998.

[16] Henry Scheffé. *The analysis of variance. Repr. of the 1959 orig.* Wiley Classics Library. New York, NY: Wiley. , 1999.

[17] Mark J. Schervish. *Theory of statistics.* Springer Series in Statistics. New York, NY: Springer-Verlag., 1995.

[18] James R. Schott. *Matrix analysis for statistics.* Wiley Series in Probability and Mathematical Statistics. New York, NY: Wiley. , 1997.

[19] Shayle R. Searle, George Casella, and Charles E. McCulloch. *Variance components.* Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons., 2006.

[20] J. Seely. Linear spaces and unbiased estimation. *Ann. Math. Stat.*, (41):1725—1734, 1970.

[21] J. Seely. Linear spaces and unbiased estimation. an aplication to the mixed linear model. *Ann. Math. Stat.*, (41):1735—1748, 1970.

[22] J. Seely. Quadratic subspaces and completeness. *Ann. Math. Stat.*, (42):710—721, 1971.

[23] J. Seely and G. Zyskind. Linear spaces and minimum variance unbiased estimation. *Ann. Math. Stat.*, (42):691—703, 1971.

[24] Justus Seely. Completeness for a family of multivariate normal distributions. *Ann. Math. Stat.*, 43:1644–1647, 1972.

[25] Justus Seely. Minimal sufficient statistics and completeness for multivariate normal families. *Sankhyā, Ser. A*, 39:170–185, 1977.

[26] Thomas A. Severini. *Likelihood methods in statistics.* Oxford Statistical Science Series. 22. Oxford: Oxford University Press. , 2000.

[27] S.D. Silvey. *Statistical inference. Reprinted with corrections.* Monographs on Applied Probability and Statistics. London: Chapman and Hall Ltd. 192 p. £2.30 , 1975.

[28] Samaradasa Weerahandi. Generalized confidence intervals. *J. Am. Stat. Assoc.*, 88(423):899–905, 1993.

[29] Samaradasa Weerahandi. *Exact statistical methods for data analysis. 2nd print.* Berlin: Springer-Verlag., 1996.

[30] David Williams. *Probability with martingales.* Cambridge etc.: Cambridge University Press. , 1991.

[31] Leping Zhou and Thomas Mathew. Some tests for variance components using generalized p values. *Technometrics*, 36(4):394–402, 1994.

[32] R. Zmyslony. On estimation of parameters in linear models. *Zastosow. Mat.*, 15:271–276, 1976.

[33] Roman Zmyslony. A characterization of best linear unbiased estimators in the general linear model. 1980.

[34] Roman Zmyslony. Completeness for a family of normal distributions. Mathematical statistics, Banach Cent. Publ. 6, 355-357 (1980)., 1980.

[35] Stefan Zontek. Robust estimation in linear model for spatially located sensors and random input. *Tatra Mt. Math. Publ.*, 17:301–310, 1999.