Lecture Notes in Statistics

47

Albert J. Getson
Francis C. Hsuan

{2}-Inverses and Their Statistical Application

Springer-Verlag

Lecture Notes in Statistics

Edited by J. Berger, S. Fienberg, J. Gani, K. Krickeberg, and B. Singer

47

Albert J. Getson
Francis C. Hsuan

{2}-Inverses and Their Statistical Application

Springer-Verlag
New York Berlin Heidelberg London Paris Tokyo

Authors

Albert J. Getson
Merck, Sharp and Dohme Research Laboratories
West Point, PA 19486, USA

Francis C. Hsuan
Department of Statistics
School of Business Administration, Temple University
Philadelphia, PA 19122, USA

AMS Subject Classification (1980): 15-04, 15A09, 15A63, 62F99, 62H10, 65U05

ISBN 0-387-96849-0 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-96849-0 Springer-Verlag Berlin Heidelberg New York

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
2847/3140-543210

To my daughters Joanne and Stephanie, ages 7 and 5,
who believe GAG = G has no meaning.

To Kathy, thanks for all your love through the years.

A.J.G.

PREFACE

Much of the traditional approach to linear model analysis is bound up in complex matrix expressions revolving about the usual generalized inverse. Motivated by this important role of the generalized inverse, the research summarized here began as an interest in understanding, in geometric terms, the four conditions defining the unique Moore-Penrose Inverse. Such an investigation, it was hoped, might lead to a better understanding, and possibly a simplification of, the usual matrix expressions.

Initially this research was begun by Francis Hsuan and Pat Langenberg, without knowledge of Kruskal's paper published in 1975. This oversight was perhaps fortunate, since if they had read his paper they may not have continued their effort. A summary of this early research appears in Hsuan, Langenberg and Getson (1985).

This monograph is a summary of the research on {2}-inverses continued by Al Getson, while a graduate student, in collaboration with Francis Hsuan of the Department of Statistics, School of Business Administration, at Temple University, Philadelphia.

The literature on generalized inverses and related topics is extensive and some of what is present here has appeared elsewhere. Generally, this literature is not presented from the point of view of {2}-inverses. We have tried to do justice to the relevant published works and appologize for those we have either overlooked or possibly misrepresented.

While it is our intention here to present a comprehensive study of {2}-inverses in statistics, we feel that this work is be no means exhaustive. Much work remains to be done, particularly in the area of multivariate analysis.

We wish to thank Dr. D. Raghavarao, Chairman of the Statistics Department at Temple University, for his encouragement of the publication of this work. We also thank the editorial staff at Springer-Verlag for their comments and suggestions on the preparation of the manuscript.

Finally the first author would like to thank his friends and colleagues in CBARDS at Merck, Sharp and Dohme Research Laboratories for the support he received in completing his research and degree.

A.J.G, F.C.H.
Philadelphia, PA
June 1988

TABLE OF CONTENTS

CHAPTER I
INTRODUCTION

A {2}-inverse for a given matrix A is any matrix G satisfying the second of the four conditions defining the unique Moore-Penrose Inverse of A:

$$(1)\ AGA = A \qquad (1.1)$$

$$(2)\ GAG = G \qquad (1.2)$$

$$(3)\ (AG)' = AG \qquad (1.3)$$

$$(4)\ (GA)' = GA. \qquad (1.4)$$

It is possible to construct matrices satisfying only a specified subset of the above conditions, for example (i),(j),...,(k). Such matrices, known as {i,j,...,k}-inverses, will be denoted $A^{+}_{i,j,\ldots,k}$. In this notation A^{+}_{1} is the usual g-inverse. Other classes of generalized inverses have been proposed in the literature and a number of texts have treated the subject in considerable depth. These include Pringle and Rayner (1971), Rao and Mitra (1971), and Ben-Israel and Greville (1974). In these works, the focus is generally on the {1}-inverse. In contrast, {2}-inverses, as their name implies, remain the stepchild of the {1}-inverse despite their importance in numerical analysis and electrical network theory [Ben-Israel and Greville (1974), pp. 27, 76].

The main function of the {1}-inverse is in solving a system of linear equations, especially when the system has deficiencies in rank, or is plainly inconsistent. It is our intention here to provide a comprehensive study of the {2}-inverse: its geometric characterization, algebraic properties, and uses in statistics. As we shall demonstrate, the {2}-inverse has several additional uses ranging from characterizing quadratic forms to computing algorithms in linear models.

When it comes to their applications in statistics, {2}-inverses are ubiquitous but not indispensible. In the statistical literature, {2}-inverses have had only limited exposure. As symmetric {1,2}-inverses, their role in least squares estimation was explored by Mazumdar et al. (1980), and by

Searle (1984). They have also been mentioned in connection with quadratic forms by several authors, including Carpenter (1950), Khatri (1963, 1977) and Mitra (1968). However, in these works the results were not viewed specifically in terms of {2}-inverses and the actual importance of {2}-inverses was clouded.

One reason for focusing on the {2}-inverses is that they provide an elegant mathematical language to express many ideas and results which otherwise involve cumbersome and laborious matrix expressions. This monograph contains numerous examples illustrating this simplicity. In this respect it is analogous to comparing different levels of computer programming languages. Assembly language is powerful but cumbersome. Even for a simple task such as counting from one to one hundred requires a long series of statements. On the other hand, a similar program in a higher level language such as BASIC, FORTRAN , or APL requires only a few statements. Currently, most textbooks in linear models contain complex matrix expressions. The language of {2}-inverses makes the expressions much simpler and, as a consequence, makes the underlying concepts much more transparent.

This work is organized into five chapters beginning with this Introduction. In each of the next four chapters a different aspect of {2}-inverses is explored. Consequently each chapter is somewhat, but not totally, independent of the others. The first section of each chapter introduces the problem considered there, reviews the relevant literature, and contains a detailed outline of the chapter. To aid the reader in understanding how these chapters relate, the following overview is offered.

A many-to-one mapping does not have an inverse in the traditional sense. In Chapter II, a functional approach is given for constructing generalized inverses of such a mapping, The Three Phase Inversion Procedure. As applied to a linear mapping over a real vector space, the procedure makes the {2}-inverse the focal point for a series of definitions of generalized inverses which is broader than the traditional. These include not only the usual g-inverses but also projection operators and constrained inverses. Such a {2}-inverse approach provides a conceptual framework unifying these various notions. This is due in part to the natural association {2}-inverses have with certain well defined vector spaces. The chapter continues with an investigation of several specific properties of {2}-inverses which suggest their usefulness in statistics. The first of these deals with the decomposition of a {2}-inverse into a sum of {2}-inverses of lesser rank and the joint decomposition of a pair of matrices into a weighted sum of {2}-inverses. Both of these results will be used in Chapter III to establish {2}-inverses as the natural coefficient matrix in a quadratic form assuring a χ^2 distribution. Next, as a generalization of projection operators, {2}-inverses are shown to have a key role in solving a system of linear equations. This role is further explored in Chapter IV in connection with least squares. Chapter II concludes with an algorithm for the easy computation of any {2}-inverse.

Given $\underline{x} \sim \mathcal{N}(\underline{\mu}, \Sigma)$ with Σ nonsingular, then the well known necessary and sufficient condition for the χ^2 distribution of $\underline{x}'A\underline{x}$ can be restated to require A to be a {2}-inverse of Σ. In Chapter III, it is argued that the χ^2 distribution of $\underline{x}'A\underline{x}$ naturally forces A to be a {2}-inverse of Σ. With this as a basis, it is shown that for Σ singular, or otherwise, it is possible to represent, in a canonical way, all χ^2 distributed quadratic forms as $\underline{x}'A\underline{x}$, where A is a {2}-inverse of Σ. Quadratic forms are a special case of second degree polynomials, $\underline{x}'A\underline{x} + 2\underline{b}'\underline{x} + c$. The distribution of second degree polynomials has been given by Khatri (1977); however, his approach for obtaining the parameters of the distribution is computationally difficult. By an application of the joint {2}-inverse decomposition of a pair of matrices, the notion of a canonical representation is expanded to include all second degree polynomials. In terms of this canonical representation, the parameters of the distribution may be easily expressed, and independent polynomials or polynomials following a χ^2 distribution can be readily identified.

In solving for least squares solutions (LSS's), a common approach is to assume a set of nonestimable constraints in addition to the normal equations so that a unique LSS may be found. The use of nonestimable constraints is usually viewed as a matter of convenience but not as a matter of necessity. On the other hand, the use of a g-inverse is viewed as necessary and sufficient to obtain a solution. In Chapter IV it is shown that, whether one realizes it or not, by choosing a g-inverse to obtain a solution, one is in fact imposing a set of constraints on the LSS. In particular, it is shown that every LSS may be expressed in the usual way as $\hat{\underline{\beta}} = GX'\underline{y}$ where X is the design matrix and G is a symmetric {1,2}-inverse of X'X. In turn each symmetric {1,2}-inverse is uniquely associated with a set of nonestimable constraints. Expanding upon this result, it is shown that for any set of constraints, regardless of their rank or estimability, the corresponding constrained LSS's may be found in an analogous way by an appropriate choice of a {2}-inverse of X'X. By appropriately identifying a set of constaints, the constrained LSS with the smallest norm may be easily identified. Since {2}-inverses may be easily calculated, the approach advocated in Chapter IV leads to a computational algorithm for obtaining LSS's. This approach does not depend on factorization of X or on solving for the eigenvalues of X'X. Furthermore, the approach is easily extended to weighted least squares.

For a designed experiment with an equal number of replications in each cell, there is little controversy concerning the sums of squares to be used in testing the various effects. However, when the data are imbalanced there is no concensus on what the appropriate sum of squares is for testing an effect. There are at least four alternative formulations of a linear model: the Unconstrained Model, the Constrained Model where the parameters are assumed to satisfy a set of known constraints, the Reparameterized Model where the constraints are used to reduce the model to one of full column rank, and the Cell Means Model in which a set of constraints is forced on the cell means to assure the equivalence of this model to some parameteric one.

Hypotheses may be expressed in terms of the parameters of one of these four models or in terms of the sample cell means alone. When there is imbalance, these five analytical approaches may each lead to a different numerator sum of squares for testing an effect. In Chapter V, the focus is on the development of an algorithm for identifying the hypotheses in each of the five approaches which results in algebraically identical numerator sum of squares. The algorithm, which is based upon {2}-inverses, is computationally simpler and covers a broader spectrum than other algorithms found in the literature. The algorithm is illustrated by its application to the SAS Type II and III Sums of Squares.

CHAPTER II
TIME FOR {2}-INVERSES

2.0 Introduction

As noted in the previous chapter, various classes of generalized inverses have been proposed in the literature. Geometric characterizations of generalized inverses were presented by Kruskal (1975) and, more recently, by Rao and Yanai (1985). The principal aim of this chapter is to unify and expand upon these diverse approaches in a consistent way.

The approach presented here begins with a geometric characterization of generalized inverses proposed by Hsuan, Langenberg and Getson (1985), the Three Phase Inversion Procedure. Their approach, which differs from the traditional, makes the {2}-inverse the natural starting point for a series of definitions of generalized inverses. The construction of various types of such generalized inverses are outlined. Included in this class are generalized inverses not defined entirely through the Moore-Penrose conditions. Some of these non-Penrose type inverses have statistical applications which will be explored in later chapters.

Of particular importance in statistics are symmetric {2}-inverses, a point of view defended in subsequent chapters. Symmetric {2}-inverses are a particular case of Bott-Duffin Inverses after a paper by Bott and Duffin who described their application in electrical network theory. {2}-inverses in general and Bott-Duffin Inverses in particular have several interesting properties and characterizations which will be useful in the following chapters and which are summarized in this chapter. The discussion of these begins with a few observations which suggest the role of Bott-Duffin Inverses in statistics. The discussion continues with an examination of the relationship between {2}-inverses and projection operators. This latter relationship leads to a decomposition theorem of symmetric matrices in terms of Bott-Duffin Inverses which is a generalization of the well known spectral decomposition. This chapter concludes with a discusssion of a procedure for efficient computation of any specified {2}-inverse.

This chapter is organized into eleven sections following this introduction. A brief description of the highlights of each section follows.

2.1 A functional definition of a generalized inverse is given in Definition 2.1 in terms of the Three Phase Inversion Procedure. Two types are identified: null and nonnull augmented generalized inverses.

2.2 The Three Phase Inversion Procedure is applied to linear mappings A: $\Re^n \mapsto \Re^m$. In Corollary 2.1.1, the constrained inverses of Rao are shown to be equivalent to the null and nonnull augmented generalized inverses of A.

2.3 The null augmented generalized inverses are identified to be exactly the class of {2}-inverses. In Theorem 2.2 and its Corollary 2.2.1, the correspondence between {2}-inverses and a pair of spaces $\mathcal{E} \subset \Re^n$ and $\mathcal{F} \subset \Re^m$ is established.

2.4 Theorem 2.3 describes the construction of any generalized inverse by augmenting a {2}-inverse in a nonnull way. As a particular case, the construction of the {1}-inverse, or usual g-inverse, is given in Corollary 2.3.1.

2.5 The construction of any Moore-Penrose Type generalized inverse by an appropriate choice of spaces $\mathcal{E}$ and $\mathcal{F}$ is detailed in Theorem 2.4.

2.6 The geometric relationships existing among the various subspaces associated with generalized inverses are summarized in Figure 2.1.

2.7 The usual projectors and their generalization by Rao are shown to be generalized inverses of the identity matrix in Lemma 2.5. As a converse, generalized inverses, which are not themselves projectors, have associated with themselves a pair of projectors. This association is outlined in Theorem 2.6.

2.8 The role of the usual g-inverse in solving a consistent set of linear equations is well known. As an extension, the role of generalized inverses in solving a broader system of equations is outlined in Theorem 2.7.

2.9 {2}-inverses can be decomposed into a sum of {2}-inverses of lesser rank. This decomposition, given in Lemma 2.8 and its corollaries, will be used repeatedly in subsequent chapters.

2.10 Theorem 2.9 details the Joint {2}-Inverse Decomposition of a pair of matrices, A and B, into a weighted sum of {2}-inverses. This decomposition may be viewed as a generalization of both the spectral and singular value decompositions. In Chapter III, the Joint {2}-Inverse Decomposition will lead to the characterization of the distribution of arbitrary quadratic forms.

2.11 {2}-inverses may be easily calculated. One approach, an application of the G2SWEEP operator, is discussed in this section.

2.1 The Three Phase Inversion Procedure

Classically the inverse of a mapping exists if and only if the mapping is bijective, i.e. one-to-one and onto. A many-to-one mapping $f: D \mapsto R$ does not have an inverse in the strict sense. Nevertheless, generalized inverses can be defined in terms of the Three Phase Inversion Procedure as follows:

1. The reduction phase, in which a subset D_0 of D is chosen such that f restricted to D_0 is bijective. Let the resulting mapping be denoted by $h: D_0 \mapsto R_0$.
2. The inversion phase, in which the unique inverse of h is determined, say $h^{-1}: R_0 \mapsto D_0$.
3. The augmentation phase, in which a mapping $g: D \mapsto R$ is defined so that $g = h^{-1}$ on R_0.

The resulting $g: R \mapsto D$ can be called a generalized inverse of f.

The nonuniqueness of a generalized inverse arises in two possible ways: the choice of D_0 in the reduction phase, and in the definition of g on the portion of the range space outside R_0 in the augmentation phase.

In practice, the choice of D_0 is not completely arbitrary, nor is the manner in which h^{-1} is augmented. For example, if D and R are vector spaces and f a linear mapping, it is natural to require in the reduction phase that D_0 be a subspace of D. Under this restriction, R_0 is a subspace of R and in the inversion phase h^{-1} is also a linear mapping of R_0 onto D_0. In the augmentation phase, h^{-1} may be extended to R in a number of ways. If R_1 is a complementary subspace of R_0 in R, then h^{-1} may be extended by either mapping R_1 trivially to the null vector or onto some nonnull space. The above discussion leads to the following definition.

Definition 2.1: For vector spaces D and R and a linear mapping $f: D \mapsto R$ then

1. a linear mapping $g: R \mapsto D$ is a Generalized Inverse of f, if there exists a subspace $D_0 \subset D$ such that $f: D_0 \mapsto f(D_0)$ is bijective, and
$$g \circ f(\underline{d}) = \underline{d} \text{ if and only if } \underline{d} \in D_0; \tag{2.1}$$
2. a generalized inverse is a Null Augmented Generalized Inverse if g maps some complementary subspace of $f(D_0)$ to the null vector;
3. a generalized inverse which is not null augmented is a Nonnull Augmented Generalized Inverse.

●

The above definition of generalized inverse is broader than the traditional one. It includes not only the usual g-inverse and other inverses defined through the Moore-Penrose conditions, but also the usual projectors and the projectors defined by Rao (1974).

Null and nonnull augmented generalized inverses have appeared in the guise of constrained inverses as defined by Rao and Mitra (1971). The relationship between constrained and generalized inverses is outlined in the next section.

2.2 Constrained Inverses

Although the Three Phase Inversion Procedure is quite general, attention will be focused on linear operators on real vector spaces. A real m×n matrix A of rank r defines two linear mappings,
$$A : \mathfrak{R}^n \mapsto \mathfrak{R}^m \tag{2.2}$$
and its transpose
$$A' : \mathfrak{R}^m \mapsto \mathfrak{R}^n. \tag{2.3}$$
For these mappings Rao and Mitra [(1971), p. 99] defined various Constrained Inverses, G, satisfying different combinations of the following constraints:

Type 1 Constraints

$$\text{c: } G \text{ maps vectors of } \mathfrak{R}^m \text{ into } \mathcal{S} \subset \mathfrak{R}^n \tag{2.4}$$
$$\text{r: } G' \text{ maps vectors of } \mathfrak{R}^n \text{ into } \mathcal{T} \subset \mathfrak{R}^m \tag{2.5}$$

Type 2 Constraints

$$\text{C: GA is an identity in } \mathcal{E} \tag{2.6}$$

$$\text{R: G'A' is an identity in } \mathcal{F}. \tag{2.7}$$

Table 2.1, on the next page, summarizes their results. It is not clear from an examination of Table 2.1, what relationships, if any, exist among the various constrained inverses. However, simple relationships do exist among these classes of inverses, which are easily seen through the Three Phase Inversion Procedure.

The existence of a generalized inverse G of A, as defined in the previous section, implies the existence of an s-dimension subspace $\mathcal{E} \subset \Re^n$ such that:

$$GA\underline{e} = \underline{e} \text{ if and only if } \underline{e} \in \mathcal{E}. \tag{2.8}$$

In what follows, it will be shown that (2.13) implies the existence of a unique s-dimensional subspace $\mathcal{F} \subset \Re^m$

$$G'A'\underline{f} = \underline{f} \text{ if and only if } \underline{f} \in \mathcal{F}. \tag{2.9}$$

Thus as a consequence, the Type 2 constraints (2.6) and (2.7) are equivalent.

Let $\mathcal{V}$ be the eigenspace of AG corresponding to the eigenvalue 1, then (2.13) implies

$$A(\mathcal{E}) \subset \mathcal{V}. \tag{2.10}$$

If $\underline{v} \in \mathcal{V}$, then

$$GAG\underline{v} = G\underline{v} \tag{2.11}$$

which in turn implies

$$G\underline{v} \in \mathcal{E} \tag{2.12}$$

and

$$G\underline{v} \neq \underline{0}. \tag{2.13}$$

Thus $\text{Dim}(\mathcal{V}) = s$, which implies $A(\mathcal{E}) = \mathcal{V}$. Since the eigenvalues of a matrix and its transpose are identical with the same multiplicity, the eigenspace of G'A' corresponding to the eigenvalue 1, $\mathcal{F} \subset \Re^m$, is nonempty with $\text{Dim}(\mathcal{F}) = s$. Thus for no larger space

$$G'A'\underline{f} = \underline{f} \quad \text{for all } \underline{f} \in \mathcal{F}. \tag{2.14}$$

Notice that G' is the generalized inverse of A' corresponding to the space $\mathcal{F}$.

A further relationship exists between $\mathcal{E}$ and $\mathcal{F}$. Since $\mathcal{F}$ and $A(\mathcal{E})$ are the left and right eigenspaces of AG corresponding to the eigenvalue 1, then