Group Invariance in Statistical Inference

This page is intentionally left blank

Group Invariance in Statistical Inference

Narayan C. Giri

University of Montreal, Canada

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 912805
USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

GROUP INVARIANCE IN STATISTICAL INFERENCE

ISBN 981-02-1875-3

Printed in Singapore.

To

NILIMA

NABANITA

NANDIAN

This page is intentionally left blank

CONTENTS

Group Invariance in Statistical Inference

Chapter 0

GROUP INVARIANCE

0.0. Introduction

One of the unpleasant facts of statistical problems is that they are often too big or too difficult to admit of practical solutions. Statistical decisions are made on the basis of sample observations. Sample observations often contain information which is not relevant to the making of the statistical decision. Some simplifications are introduced by characterizing the decision rules in terms of the sufficient statistic (minimal) which discard that part of sample observations which is of no value for any decision making concerning the parameter and thereby reducing the dimension of the sample space to that of the minimal sufficient statistic. This, however, does not reduce the dimension of the parametric space. By introducing the *group invariance* principle and restricting attention to invariant decision rules a reduction to the dimension of the parametric space is possible. In view of the fact that sufficiency and group invariance are both successful in reducing the dimension of the statistical problems, one is naturally interested in knowing whether both principles can be used simultaneously and if so, in what order. Hall, Wijsman and Ghosh (1965) have shown that under certain conditions this reduction can be carried out by using both principles simultaneously and the order in which they are used is immaterial in such cases. However, one can avoid verifying these conditions by replacing the sample space by the space of sufficient statistic and then using group invariance on the space of sufficient statistic. In this monograph we treat multivariate problems only where the reduction in dimension is very significant. In what follows we use the term invariance to indicate group invariance.

In statistics the term *invariance* is used in the mathematical sense to denote a property that remains unchanged (invariant) under a group of transformations. In actual practice many statistical problems possess such a property. As in other branches of applied sciences it is a generally accepted principle in statistics that if a problem with an unique solution is invariant under a group of transformations, then the solution should also be invariant under it. This notion has an old origin in statistical sciences. Apart from this natural justification for the use of invariant decision rules, the unpublished work of Hunt and Stein towards the end of Second World War has given this principle a strong support as to its applicability and meaningfulness to prove various optimum properties like minimax, admissibility etc. of statistical decision rules.

Although a great deal has been written concerning this principle in statistical inference, no great amount of literature exists concerning the problem of discerning whether or not a given statistical problem is actually invariant under a certain group of transformations. Brillinger (1963) gave necessary and sufficient conditions that a statistical problem must satisfy in order that it be invariant under a fairly large class of group of transformations including Lie groups.

In our treatment in this monograph we treat invariance in the framework of statistical decision rules only. De Finetti (1964) in his theory of exchangeability treats invariance of the distribution of sample observations under finite permutations. It provides a crucial link between his theory of subjective probability and the frequency approach of probability. The classical statistical methods take as basic a family of distributions, the true distribution of the sample observations is an unknown member of this family about which the statistical inference is required. According to De Finetti's approach no probability is unknown. If $x_1, x_2, \ldots$ are the outcomes of a sequence of experiments conducted under similar conditions, subjective uncertainty is expressed directly by ascribing to the corresponding random variables $X_1, X_2, \ldots$ a known joint distribution. When some of the X's are observed, predictive inference about others is made by conditioning the original distributions on the observations. De Finetti has shown that these approaches are equivalent when the subjectivist's joint distribution is invariant under finite permutation.

Two other related principles, known in the literature, are the weak invariance and the strong invariance principles. The weak invariance principle is used to demonstrate the sufficiency of the classical assumptions associated with the weak convergence of stable laws (Billingsley, 1968). This is popularly known as Donsker's theorem (Donsker, 1951). Let $X_1, X_2, \ldots$ be independently distributed random variable with the same mean zero and the variance σ^2 and let $S_j = \sum_{i=1}^{j} X_i$, $X_n(t) = \frac{S_i}{\sigma\sqrt{n}}$, $t = \frac{i}{n}$, $i = 1, \ldots, n$. Donsker proved that $\{X_n(t)\}$ converges weakly to Brownian motion. Strong invariance principle has been introduced to prove the strong convergence result (Tusnady, 1977). Here the term invariance is used in the sense that if $X_1, X_2, \ldots$ are independently distributed random variables with the same mean 0 and the same variance σ^2, and if h is a continuous function on $[0, 1]$ then the limiting distribution of $h(X_i)$ does not depend on any other property of X_i.

0.1. Examples

We now give an example to show how the solution to a statistical problem can be obtained through direct application of group theoretic results.

Example 0.1.1. Let $X^\alpha = (X_{\alpha 1}, \ldots, X_{\alpha p})'$, $\alpha = 1, \ldots, N(> p)$ be independently and identically distributed p-variate normal vectors with the same mean $\mu = (\mu_1, \ldots, \mu_p)'$, and the same positive definite covariance matrix Σ. The parametric space Ω is the space of all (μ, Σ). The problem of testing $H_0 : \mu = 0$ against the alternatives $H_1 : \mu \neq 0$ remains unchanged (invariant) under the full linear group $G_l(p)$ of $p \times p$ nonsingular matrices g transforming each $X_i \to gX_i$, $i = 1, \ldots, N$. Let

$$\bar{X} = \frac{1}{N} \sum_{\alpha=1}^{N} X^\alpha, \quad S = \sum_{\alpha=1}^{N} (X^\alpha - \bar{X})(X^\alpha - \bar{X})' . \tag{0.1}$$

It is well-known (Giri, 1977) that $(\bar{X}, S)$ is sufficient for (μ, Σ). The induced transformation on the space of $(\bar{X}, S)$ is given by

$$(\bar{X}, S) \to (g\bar{X}, gSg'), \quad g \in G_l(p) . \tag{0.2}$$

Since this transformation permits arbitrary changes of $\bar{X}, S$ and any reasonable statistical test procedure should not depend on any such arbitrary change by g, we conclude that a reasonable statistical test procedure should depend on $(\bar{X}, S)$ only through

$$T^2 = N(N-1)\bar{X}'S^{-1}\bar{X} . \tag{0.3}$$

It is well-known that (Giri, 1977) the distribution of T^2 is given by

$$f_{T^2}(t^2|\delta^2) = \frac{\exp\{-\frac{1}{2}\delta^2\}}{(N-1)\Gamma(\frac{1}{2}(N-p))} \times \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\delta^2)^j (t^2/(N-1))^{\frac{p}{2}+j-1}\Gamma(\frac{1}{2}N+j)}{j!\Gamma(\frac{1}{2}p+j)(1+t^2/(N-1))^{\frac{1}{2}N+j}} \quad \text{if } t^2 \geq 0 \tag{0.4}$$

where $\delta^2 = N\mu'\Sigma^{-1}\mu$. Under $H_0\, \delta^2 = 0$ and under $H_1\, \delta^2 > 0$. Applying Neyman and Pearson's Lemma we conclude from (0.4) that the uniformly most powerful test based on T^2 of H_0 against H_1 is the well-known Hotelling's T^2 test, which rejects H_0 for large values of T^2.

Note. In this problem the dimension of Ω is $p + \frac{p(p+1)}{2} = \frac{p(p+3)}{2}$, which is also the dimension of the $(\bar{X}, S)$. For the distribution of T^2 the parameter is a scalar quantity.

One main reason of the intuitive appeal, that for an invariant problem with an unique solution, the solution should be invariant, is probably the belief that there should be a unique way of analysing a collection of statistical data. As a word of caution we should point out that, if in cases where the use of invariant decision rule conflicts violently with the desire to make a correct decision or to have a smaller risk, it must be abandoned. We give below one such example which is due to Charles Stein as reported by Lehmann (1959, p. 338).

Example 0.1.2. Let $X = (X_1, \ldots, X_p)'$, $Y = (Y_1, \ldots, Y_p)'$ be independently distributed normal p-vectors with the same mean 0 and positive definite covariance matrices $\Sigma, \delta\Sigma$ respectively where δ is an unknown scalar constant. Consider the problem of testing $H_0 : \delta = 1$ against $H_1 : \delta > 1$. This problem is invariant under $G_l(p)$ transforming $X \to gX, Y \to gY, g \in G_l(p)$. Since this group is transitive (see Chapter 1) on the space of values of (X, Y) with probability one, the uniformly most powerful invariant test of level α under $G_l(p)$ is the trivial test $\Phi(X, Y) = \alpha$ which rejects H_0 with constant probability α for all values (x, y) of (X, Y). Hence the maximum power that can be achieved over the alternatives H_1 by any invariant test under $G_l(p)$ is also α. But the test which rejects H_0 whenever

$$\frac{X_1^2}{Y_1^2} \geq C\,, \tag{0.5}$$

where the constant C depends on level α, has strictly increasing power $\beta(\delta)$ whose minimum over the set $\delta \geq \delta_1 > 1$ is $\beta(\delta_1) > \beta(1) = \alpha$. For more discussions and results refer to Giri (1983a, 1983b).

Exercises

1. Let $X_1, \ldots, X_n$ be independently and identically distributed normal random variables with the same mean θ and the same unknown variance σ^2 and let $H_0 : \theta = 0$ and $H_1 : \theta \neq 0$.

   (a) Find the largest group of transformations which leaves the problem of testing H_0 against H_1 invariant.

   (b) Using the group theoretic notion show that the two-sided student t-test is uniformly most powerful among all tests based on t.

2. **Univariate General Linear Hypothesis**. Let $X_1, \ldots, X_n$ be independently distributed normal random variables with $E(X_i) = \theta_i$, $\operatorname{Var}(X_i) = \sigma^2$, $i = 1, \ldots, n$. Let Ω be the linear coordinate space of dimension of n and let Π_Ω and Π_ω be two linear subspaces of Ω such that dim $\Pi_\omega = k$ and dim $\Pi_\Omega = l$, $l > k$. Consider the problem of testing $H_0 : \theta = (\theta_1, \ldots, \theta_p)' \in \Pi_\omega$ against the alternatives $H_1 : \theta \in \Pi_\Omega$.

   (a) Find the largest groups of transformations which leave the problem invariant.

   (b) Using the group theoretic notions show that the usual F-test is uniformly most powerful for testing H_0 against H_1.

3. Let $X_1, \ldots, X_n$ be independently distributed normal random variables with the same mean θ and the same variance σ^2. For testing $H_0 : \sigma^2 = \sigma_1^2$ against $H_1 : \sigma^2 = \sigma_2^2 < \sigma_1^2$, where σ_1^2, σ_2^2 are known, find the most powerful invariant test using the group theoretic notions.

References

1. P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
2. D. Brillinger, *Necessary and sufficient conditions for a statistical problem to be invariant under a Lie group*, Ann. Math. Statist., **34** (1963) 492–500.
3. B. De Finetti, *Studies in Subjective Probability*, H. E. Kyburg and H. E. Smoker, eds., Wiley, New York, (1964) 93–158.
4. M. Donsker, *An invariance principle for certain probability limit theorems*, Amer. Math. Soc., **6** (1951).
5. N. Giri, *Hunt–Stein theorem*, Encyclopedia of Statistical Sciences, Vol. 3, Kotz–Johnson, eds., Wiley, New York, 1983a, 686–689.
6. N. Giri, *Invariance Concepts in Statistics*, Encyclopedia of Statistical Sciences, Vol. 4, Kotz–Johnson, eds., Wiley, New York, 1983b, 219–224.
7. N. Giri, *Multivariate Statistical Inference*, Academic Press, New York, 1977.

8. W. J. Hall, R. A. Wijsman and J. K. Ghosh, *The relationship between sufficiency and invariance with application in sequential analysis*, Ann. Math. Statist. **36** (1965) 575–614.
9. E. L. Lehmann, *Testing of Statistical Hypothesis*, Wiley, New York, 1959.
10. G. Tusnady, in *Recent Developments in Statistics*, Barra, J. R., Brodeau, F., Romier, G. and Van Cutsem, B. eds, North-Holland, Amsterdam, (1977) 289–300.

Chapter 1

MATRICES, GROUPS AND JACOBIANS

1.0. Introduction

The study of group invariance requires knowledge of matrix algebra and group theory. We present here some basic results on matrices, groups and Jacobians without proofs.

Proofs can be obtained from Giri (1993, 1996) or any textbook on these topics.

1.1. Matrices

A $p \times q$ matrix $C = (c_{ij})$ is a rectangular array of real number c_{ij} written as

$$C = \begin{pmatrix} c_{11} \dots c_{1q} \\ \vdots \ \ddots \ \vdots \\ c_{p1} \dots c_{pq} \end{pmatrix} \tag{1.1}$$

where c_{ij} is the element in the ith row and the jth column. The transpose C' of C in a $q \times p$ matrix obtained by interchanging the rows and the columns of C. If $q = p, C$ is called a square matrix of order p. If $q = 1, C$ is a $1 \times p$ row vector and if $p = 1, C$ is a $q \times 1$ column vector.

A square matrix C is *symmetric* if $C = C'$.

A square matrix $C = (c_{ij})$ of order p is a *diagonal matrix* $D(c_{11}, \dots, c_{pp})$ with diagonal elements $c_{11}, \dots, c_{pp}$ if all off-diagonal elements of C are zero.

A diagonal matrix with unit diagonal elements is an *indentity matrix* and is denoted by I. Sometimes it will be necessary to write it as I_k to denote an identity matrix of order k.

A square matrix $C = (c_{ij})$ of order p is a *lower triangular matrix* if $c_{ij} = 0$ for $j > i$. The determinant of the lower triangular matrix det $C = \prod_{i=1}^{p} c_{ii}$. We shall also write det $C = |C|$ for convenience.

A square matrix $C = (c_{ij})$ of order p is a *upper triangular matrix* if $c_{ij} = 0$ for $i > j$ and det $C = \prod_{i=1}^{p} c_{ii}$.

A square matrix of order p is *nonsingular* if det $C \neq 0$. If det $C = 0$ then C, is a singular matrix.

A nonsingular matrix C of order p is orthogonal if $CC' = C'C = I$.

The inverse of a nonsingular matrix C of order p is the unique matrix C^{-1} such that $CC^{-1} = C^{-1}C = I$. From this it follows that det $C^{-1} = (\text{det } C)^{-1}$.

A square matrix $C = (c_{ij})$ of order p or the associated quadratic form $x'Cx = \sum_i \sum_j c_{ij}x_ix_j$ is positive definite if $x'Cx > 0$ for $x = (x_1, \ldots, x_p)' \neq 0$. If C is positive definite C^{-1} is positive definite and for any nonsingular matrix A of order p ACA' is also positive definite.

1.1.1. *Characteristic roots and vectors*

The characteristic roots of a square matrix C of order p are given by the roots of the characteristic equation

$$\det (C - \lambda I) = 0 \tag{1.2}$$

where λ is real. As

$$\det (\theta C \theta' - \lambda I) = \det (C - \lambda I)$$

for any orthogonal matrix θ of order p, the characteristic roots of C remain invariant under the transformation of $C \to \theta C \theta'$. The vector $x = (x_i, \ldots, x_p)' \neq 0$ satisfying

$$(C - \lambda I)x = 0 \tag{1.3}$$

is the characteristic vector of C corresponding to its characteristic root λ. If x is a characteristic vector of C corresponding to its characteristic root λ, then any scalar multiple $ax, a \neq 0$, is also a characteristic vector of C corresponding to λ.

Some Results on Characteristic Roots and Vectors

1. The characteristic roots of a real symmetric matrix are real.
2. The characteristic vectors corresponding to distinct characteristic roots of a symmetric matrix are orthogonal.

3. The characteristic roots of a symmetric positive definite matrix C are all positive.
4. Given any real square symmetric matrix C of order p, there exists an orthogonal matrix θ of order p such that $\theta C \theta'$ is a diagonal matrix $D(\lambda_1, \ldots, \lambda_p)$ where $\lambda_1, \ldots, \lambda_p$ are the characteristic roots of C. Hence det $C = \prod_{i=1}^{p} \lambda_i$ and tr $C = \sum_{i=1}^{p} \lambda_i$. Note that tr C is the sum of diagonal elements of C.

1.1.2. *Factorization of matrices*

In this sequal we shall use frequently the following factorizations of matrices.

For every positive definite matrix C of order p there exists a nonsingular matrix A of order p such that $C = AA'$ and, hence, there exists a nonsingular matrix $B\,(B = A^{-1})$ of order p such that $BCB' = I$.

Given a symmetric nonsingular matrix C of order p, there exists a nonsingular matrix A of order p such that

$$ACA' = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \tag{1.4}$$

where the order of I is equal to the number of positive characteristic roots of C and the order of $-I$ is equal to the number of negative characteristic roots of C.

Given a symmetric positive definite matrix C of order p there exists a nonsingular lower traingular matrix A (an upper triangular matrix B) of the same order p such that

$$C = AA' = B'B\,. \tag{1.5}$$

Cholesky Decomposition For every positive definite matrix C there exists an unique lower triangular matrix of positive diagonal elements D such that $C = DD'$.

1.1.3. *Partitioned matrices*

Let $C = (c_{ij})$ be a $p \times q$ matrix and let C be partitioned into submatrices C_{ij} as

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

where $C_{11} = (c_{ij})(i = 1, \ldots, m;\ j = 1, \ldots, n); C_{12} = (c_{ij})(i = 1, \ldots, m;\ j =$

$n+1,\ldots,q); C_{21} = (c_{ij})(i = m+1,\ldots,p;\ j = 1,\ldots,n); C_{22} = (c_{ij})(i = m+1,\ldots,p;\ j = n+1,\ldots,q)$. For any square matrix

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

where C_{11}, C_{22} are square matrices and C_{22} is nonsingular
(1) det $(C) =$ det (C_{22}) det $(C_{11} - C_{12}C_{22}^{-1}C_{21})$,
(2) C is positive definite if and only if $C_{11}, C_{22} - C_{21}C_{11}^{-1}C_{21}$ or equivalently $C_{22}, C_{11} - C_{12}C_{22}^{-1}C_{21}$ are positive definite.

Let $C^{-1} = B$ be similarly partitioned into submatrices $B_{ij}, i, j = 1, 2$. Then

$$C_{11}^{-1} = B_{11} - B_{12}B_{22}^{-1}B_{21}, \;\; C_{22}^{-1} = B_{22} - B_{21}B_{11}^{-1}B_{12},$$
$$C_{11}^{-1}C_{12} = -B_{12}B_{22}^{-1}. \tag{1.6}$$

1.2. Groups

A group G is a set with an operation τ satisfying the following axioms.
A_1. For any two elements $a, b \in G, a\tau b \in G$.
A_2. For any three elements $a, b, c \in G; (a\tau b)\tau c = a\tau(b\tau c)$.
A_3. There exists an element e (identity element) such that for all $a \in G, a\tau e = a$.
A_4. For any $a \in G$ there exists an element a^{-1} (inverse element) such that $a\tau a^{-1} = e$.

In what follows we will write for the convenience of notation $a\tau b = ab$. In such writing the reader may not confuse it with the arithmetic product ab.

A group G is abelian if for any pair of elements a, b belonging to $G, ab = ba$.

A non-empty subset H of G is a subgroup if the restriction of the group operation τ to H satisfies the axioms $A_1, \ldots, A_4$,

Examples of Groups

Example 1.1. A. Let $\mathfrak{X}$ be a set and G be the set of all one-to-one mappings

$$g : \mathfrak{X} \to \mathfrak{X}$$

with $g(x) = g(y); x, y \in \mathfrak{X}$; implies $x = y$ and for $x \in \mathfrak{X}$ there exists $y \in \mathfrak{X}$ such that $y = g(x)$. With the group operation defined by

$$g_1g_2(x) = g_1(g_2(x)); g_1, g_2 \in G,$$

G forms a permutation group.

Example 1.2. The additive group of real numbers is the set of all reals with the group operation $ab = a + b$. The multiplicative group of all nonzero reals with the group operation $ab = a$ multiplied by b.

Example 1.3. Let $\mathfrak{X}$ be a linear space of dimension n.

Define for $x_0 \in \mathfrak{X}, g_{x_0}(x) = x + x_0, x \in \mathfrak{X}$. The set of all $\{g_{x_0}\}$ forms an additive abelian group and is called the translation group.

Example 1.4. Let $\mathfrak{X}$ be a linear space of dimension n and let $G_l(n)$ be the set of all nonsingular linear transformations $\mathfrak{X}$ onto $\mathfrak{X}$. $G_l(n)$ with matrix multiplication as the group operation is called the full linear group.

Example 1.5. The affine group is the set of pairs $(g, x), x \in \mathfrak{X}, g \in G_l(n)$ with the group operation defined by $(g_1, x_1)(g_2, x_2) = (g_1 g_2, g_1 x_2 + x_1)$ with $g_1, g_2 \in G_l(n)$ and $x_1, x_2 \in \mathfrak{X}$. The identity element and the inverses are $(I, 0)$ and $(g, x)^{-1} = (g^{-1}, -g^{-1}x)$.

Example 1.6. The set of all nonsingular lower triangular matrices of order p with the usual matrix multiplication as the group operation forms a group $G_T(p)$ (with identity matrix as the unit element). The set of all upper triangular nonsingular matrices of order p forms the group $G_{UT}(p)$ with identity matrix as the unit element.

Example 1.7. The set of all orthogonal matrices θ of order p with the matrix multiplication as the group operation forms a group $O(p)$.

Example 1.8. Let $o \subset \mathfrak{X}_1 \subset \mathfrak{X}_2 \subset \cdots \subset \mathfrak{X}_k = \mathfrak{X}$ be a strictly increasing sequence of linear subspaces of $\mathfrak{X}$ and let G be a subgroup of $G_l(p)$ such that $g \in G_l(p)$ if and only if $g\mathfrak{X}_i = \mathfrak{X}_i, i = 1, \ldots, k$. The group G is the group of nonsingular linear transformations on $\mathfrak{X}$ to $\mathfrak{X}$ which leaves $\mathfrak{X}_i$ invariant.

Choose a basis $x_1^{(i)}, \ldots, x_{n_i}^{(i)}, i = 1, \ldots, k$ for $\mathfrak{X}_i - \mathfrak{X}_{i-1}$ with $n_i = \dim(\mathfrak{X}_i) - \dim(\mathfrak{X}_{i-1})$ and $\mathfrak{X}_0 = \phi$ (null set).

Then $g \in G$ can be written as

$$g = \begin{pmatrix} g_{(11)} & g_{(12)} & \cdots & g_{(1k)} \\ 0 & g_{(22)} & \cdots & g_{(2k)} \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & g_{(kk)} \end{pmatrix} \tag{1.7}$$

where $g_{(ij)}$ is a block of n_i rows and n_j columns for $i \leq j$. If $n_i = 1$ for all $i, G = G_{UT}$.

1.3. Homomorphism, Isomorphism and Direct Product

Let G, H be groups. A mapping f of G into H is called a homomorphism if, for $g_1, g_2 \in G$,

$$f(g_1 g_2) = f(g_1) * f(g_2) \tag{1.8}$$

where * is the group operation in H. If e is the identity element in $G, f(e)$ is the identity element in H and $f(g^{-1}) = [f(g)]^{-1}$. In addition, if f is an one-to-one mapping, then f is called an isomorphisim.

The Cartesian product $G \times H$ of groups G and H with the operation $(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2); g_1, g_2 \in G, h_1, h_2 \in H$ is called the direct product of G and H.

A subgroup H of G is a normal subgroup of G if for all $h \in H$ and $g \in G$, $ghg^{-1} \in H$, or equivalently if $gHg^{-1} = H$.

Example 1.9. The group $G_T^+(p)$ of lower triangular matrices of order p with positive diagonal elements is a normal subgroup of $G_T(p)$. Any subgroup of an abelian group is normal.

Let G be a group and let H be a subgroup of G. The set G/H of all sets of the form $gH, g \in G$ is called the quotient group of G modulo H with group operation defined by, for $g_1, g_2 \in G, (g_1 H)(g_2 H)$ is a set of elements obtained by multiplying all elements of $g_1 H$ with all elements of $g_2 H$. The identity element is H and $(gH)^{-1} = g^{-1}H$.

1.4. Topological Transitive Groups

Let $\mathfrak{X}$ be a set and $\mathcal{A}$ a collection of subsets of $\mathfrak{X}$. $(\mathfrak{X}, \mathcal{A})$ is called a topological space if $\mathcal{A}$ satisfies the following axioms:

$\mathrm{TA}_1 : \mathfrak{X} \in \mathcal{A}$.

TA_2 : The union of an arbitrary subfamily of $\mathcal{A}$ belongs to $\mathcal{A}$.

TA_3 : The intersection of a finite subfamily of $\mathcal{A}$ belongs to $\mathcal{A}$.

The elements of $\mathcal{A}$ are called open sets. If, in addition the topological space $(\mathfrak{X}, \mathcal{A})$ satisfies: for $x, y \in \mathfrak{X}, x \neq y$, there exist two open sets A, B, belonging to $\mathcal{A}$ such that $A \cap B = \phi$ with $x \in A, y \in B$; then it is called a *Hausdorff space.*

A collection of open sets is a base for a topology if every open set in $\mathcal{A}$ is the union of a sub-collection of sets in the base.

If a group G has a topology τ defined on it such that under τ the mapping ab from $G \times G$ into G and a^{-1} from G into $G(a, b \in G)$ are continuous, then (G, τ) is a topological group.

A compact group is a topological group which is a compact space. A locally compact group is a topological group which is a locally compact space.

A compact space is a Hausdorff space $\mathfrak{X}$ such that every open covering of $\mathfrak{X}$ can be reduced to a finite subcovering. A locally compact space is a Hausdorff space such that every point has at least one compact neighborhood.

Let $\mathfrak{X}$ be a set. The group G operates from the left on $\mathfrak{X}$ if there exists a function on $G \times \mathfrak{X}$ into $\mathfrak{X}$ whose value at (g, x) is denoted by gx such that (i) $ex = x$ for all $x \in \mathfrak{X}$ and e, the identity element of G; (ii) $g_2(g_1 x) = g_2 g_1(x)$. This implies that $g \epsilon G$ is one to one on $\mathfrak{X}$ into $\mathfrak{X}$.

Let G operate from the left on $\mathfrak{X}$. G **operates transitively** on $\mathfrak{X}$ if for every $x_1, x_2 \in \mathfrak{X}$ there exists a $g \in G$ such that $gx_1 = x_2$.

Example 1.10. Let $\mathfrak{X}$ be a linear space. The full linear group operates transitively on $\mathfrak{X} - \{0\}$.

Example 1.11. The group $O(n)$ of $n \times n$ orthogonal matrices operates transitively on the set of all $n \times p(n \geqslant p)$ real matrices, x satisfying $x'x = I_p$.

Example 1.12. The linear group $G_\ell(p)$ acts transitively on the set of all $p \times p$ positive definite matrices s, transfering s to $gsg', g \in G_\ell(p)$.

1.5. Jacobians

Let $X_1, \ldots, X_n$ be a sequence of continuous random variables with *pdf* $f_{X_1,\ldots,X_n}(X_1, \ldots, X_n)$ and let $Y_i = g_i(X_1, \ldots, X_n), i = 1, \ldots, n$ be a set of continuous one-to-one transformations of $X_1, \ldots, X_n$. Assume that the function $g_1, \ldots, g_n$ have continuous partial derivatives with respect to $X_1 \ldots, X_n$. Denote the inverse functions by $X_i = h_i(Y_1, \ldots, Y_n), i = 1, \ldots, n$. Let J^{-1} be the determinant of the $n \times n$ matrix of partial derivatives

$$\begin{pmatrix} \frac{\partial X_1}{\partial Y_1} & \cdots & \frac{\partial X_1}{\partial Y_n} \\ \cdots & \ddots & \cdots \\ \frac{\partial X_n}{\partial Y_1} & \cdots & \frac{\partial X_n}{\partial Y_n} \end{pmatrix}, \tag{1.9}$$

J is called the Jacobian of the transformation $X_1, \ldots, X_n$ to $Y_1, \ldots, Y_n$. The *pdf* of $Y_1, \ldots, Y_n$ is given by

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = f_{X_1,\ldots,X_n}(h_1(y_1, \ldots, y_n), \ldots, h_n(y_1, \ldots, y_n)|J|$$

where $|J|$ is the absolute value of J. We now state some results on Jacobian without proof. We refer to Giri (1994), Olkin (1962) and Roy (1959) for further results and proofs.

Some Results on Jacobians

Let $X = (X_1, \ldots, X_p)', Y = (Y_1, \ldots, Y_p)' \in E^p$. The Jacobian of the transformation $X \to Y = gX$ $g \in G_l(p)$ is $(\det g)^{-1}$.

Let X, Y be $p \times n$ matrices. The Jacobian of the transformation $X \to Y = gx, g \in G_l(p)$ is $(\det g)^{-n}$.

Let X, Y be $p \times n$ matrices and let $A \in G_l(p), B \in G_\ell(n)$. The Jacobian of the transformation $X \to Y = AXB$ is $(\det A^{-1})^n$ $(\det B^{-1})^p$.

Let $g, h \in G_T(p)$. The Jacobain of the transformation $g \to hg$ is $\prod_{i=1}^p (h_{ii})^{-1}$ where $h = (h_{ij})$ and the Jacobian of the transformation $g \to gh$ is $\prod_{i=1}^p (h_{ii})^{i-p-1}$.

Let $g, h \in G_{UT}(p)$. The Jacobian of the transformation $g \to hg$ is $\prod_{i=1}^p (h_{ii})^{i-p-1}$ where $h = (h_{ij})$ and the Jacobian of the transformation $g \to gh$ is $\prod_{i=1}^b (h_{ii})^{-i}$

Let $G_{BT}(p)$ be the multiplicative group of $p \times p$ lower triangular nonsingular matrices in the block form, i.e. $g \in G_{BT}(p)$,

$$g = \begin{pmatrix} g_{(11)} & 0 & 0 & \cdots & 0 \\ g_{(21)} & g_{(22)} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ g_{(k1)} & g_{(k2)} & \cdot & \cdots & g_{(kk)} \end{pmatrix} \tag{1.10}$$

where $g_{(ii)}$ are submatrices of order $p_i \times p_i$ such that $\sum_1^k p_i = p$. For $g, h \in G_{BT}(p)$ the Jacobian of the transformation $g \to hg$ is $\prod_{i=1}^k [\det\ h_{(ii)}]^{-\sigma_i}$, where $\sigma_i = \sum_{j=1}^i p_j, \sigma_0 = 0$. The Jacobian of the transformation $g \to gh$ is $\prod_{i=1}^k [\det\ h_{(ii)}]^{\sigma_{i-1}-p}$.

Let $G_{BUT}(p)$ be the group of nonsingular upper triangular matrices of order p in the block form i.e. $g \in G_{BUT}$,

$$g = \begin{pmatrix} g_{(11)} & g_{(12)} & \cdots & g_{(1k)} \\ 0 & g_{(22)} & \cdots & g_{(2k)} \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & g_{(kk)} \end{pmatrix}$$

where $g_{(ii)}$ are submatrices of order p_i satisfying $\sum_1^k p_i = p$. For $g, h \in G_{BUT}$, the Jacobian of the transformation $g \to gh$ is $\prod_{i=1}^k [\det\ h_{(ii)}]^{-\sigma_i}$ and that of $g \to hg$ is $\prod_{i=1}^k [\det\ h_{(ii)}]^{\sigma_{i-1}-p}$.

Let S be a symmetric positive definite matrix of order p. The Jacobian of the transformation $S \to gSg', g \in G_l(p)$ is $[\det g]^{-(p+1)}$ and that of $S \to S^{-1}$ is $(\det S)^{2p}$.

Let S be a symmetric positive definite matrix of order p and let g be the unique lower triangular matrix such that $S = gg'$. The Jacobian of the transformation $S \to gSg'$ is $2^{-p}\prod_{i=1}^{p}(g_{ii})^{-(p+1-i)}$ with $g = (g_{ij})$. The Jacobian of the transformation $S \to gSg', g \in G_T^{+}(p)$ is $[\det g]^{-(p+1)}$.

Exercises

1. Show that for any nonsingular symmetric matrix A of order p and non null p vectors x, y

   (a) $(A + xy')^{-1} = -\dfrac{(A^{-1}x)(y'A^{-1})}{1 + y'A^{-1}x} + A^{-1}$;

   (b) $x'(A + xx')^{-1}x = \dfrac{x'A^{-1}x}{1 + x'A^{-1}x}$;

   (c) $|A + xx'| = |A|(1 + x'A^{-1}x)$.

2. Let A, B be $p \times q$ and $q \times p$ matrices respectively. Show that $|I_p + AB| = |I_q + BA|$.
3. Show that for any lower triangular matrix C the diagonal elements are its characteristic roots.
4. Let $\mathfrak{X}$ be the set of all $n \times p$ real matrices X satisfying $X'X = I_p$. Show that the group $O(n)$ of orthogonal matrices θ of order p, which transform $X \to \theta X$ acts transitively on $\mathfrak{X}$.
5. Let $\mathcal{S}$ be the set of all symmetric positive definite matrices s of order p. Show that $G_l(p)$ which transforms $s \to gsg', g \in G_l(p)$, acts transitively on $\mathcal{S}$.

References

1. N. Giri, *Multivariate Statistical Inference*, Academic Press, New York, 1977.
2. N. Giri, *Introduction to Probability and Statistics*, 2nd Edition (Expanded), Marcel Dekker, New York, 1993.
3. N. Giri, *Multivariate Statistical Analysis*, Marcel Dekker, New York, 1996.
4. I. Olkin, *Note on the Jacobian of certain transformations useful in multivariate analysis*, Biometrika, **40** (1962), 43–46.
5. S. N. Roy, *Some Aspects of Multivariate Analysis*, Wiley, New York, 1957.

Chapter 2

INVARIANCE

2.0. Introduction

Invariance is a mathematical term for symmetry. In practice many statistical problems involving testing of hypothesis, exhibit symmetries, which imposes additional restrictions for the choice of appropriate statistical tests, for example, the statistical tests must also exhibit the same kind of symmetry as is present in the problem. In this chapter we shall consider the principle of invariance of statistical testing problems only. For an additional reference the reader is referred to Lehmann (1959), Ferguson (1967) and Nachbin (1965).

2.1. Invariance of Distributions

Let $(\mathcal{X}, \mathcal{A})$ be a measure space and Ω the set of points θ, that is, $\Omega = \{\theta\}$. Consider a function P on Ω to the set of probability measures on $(\mathcal{X}, \mathcal{A})$ whose value at θ is P_θ. Let G be a group of transformations operating from the left on $\mathcal{X}$ such that $g \in G$

$$g : \mathcal{X} \overset{\text{onto}}{\to} \mathcal{X} \text{ is } (\mathcal{A},\, \mathcal{A}) \text{ measurable;} \tag{2.1}$$

and

$$\bar{g} : \Omega \overset{\text{onto}}{\to} \Omega \tag{2.2}$$

is such that if X has distribution P_θ, gX, $X \in \mathcal{X}$ has distribution $P_{\theta'}$ where $\theta' = \bar{g}\theta \in \Omega$. All transformations considered in connection with invariance will be taken for granted as one-to-one from $\mathcal{X}$ onto $\mathcal{X}$. An equivalent way of stating (2.2) is as follows:

$$P_\theta(gX \in A) = P_{\bar{g}\theta}(X \in \mathcal{A}), A \in \mathcal{A}, \tag{2.2a}$$

or

$$P_\theta(g^{-1}A) = P_{\bar{g}\theta}(A) \tag{2.2b}$$

This can also be written as

$$P_\theta(B) = P_{\bar{g}\theta}(gB), \quad B \in \mathcal{A}$$

If all P_θ are distinct, that is, $\theta_1 \neq \theta_2$ $P_{\theta_1} \neq P_{\theta_2}$, then g is a homomorphism. The condition (2.2a) is often known as the condition of invariance of the distribution.

Very often in statistical problems there exists a measure λ such that P_θ is absolutely continuous with respect to λ for all $\theta \in \Omega$ so that p_θ is the corresponding probability density function with respect to the measure λ. In other words, we can write

$$P_\theta(A) = \int_A p_\theta(x)d\lambda(x) \,. \tag{2.3}$$

Also in great many cases of interest, it is possible to choose the measure λ such that it is invariant under G, viz

$$\lambda(A) = \lambda(gA) \quad \text{for all } A \in \mathcal{A},\ g \in G \,. \tag{2.4}$$

Definition 2.1.1. A measure λ satisfying (2.4) is left invariant.

In such cases the condition of invariance of distribution under G reduces to

$$p_{\bar{g}\theta}(gx) = p_\theta(x) \quad \text{for all } x \in \mathcal{X},\ g \in G \,. \tag{2.5}$$

The general theory of invariant measures in a large class of topological groups was first given by Haar (1933). However, the results we need were all known by the end of the nineteenth century. In the terminology of Haar, $P_\theta(A)$ is called a positive integral of p_θ, not necessarily a probability density function. The basic result is that for a large class of topological groups there is a left invariant measure, positive on open sets and finite on compact sets and to within multiplication by a positive constant this measure is unique. Haar proved this result for locally compact topological groups. The definition of right invariant positive integrals, viz. $P_\theta(A) = P_{\bar{g}\theta}(Ag)$ is analagous to that for left invariant positive integrals.

Because of the poineering works of Haar such invariant measures are called invariant (left or right) Haar measures. A rigorous presentation of Haar measure will not be attempted here. The reader is referred to Nachbin (1965).

It has been shown that such invariant measures do exist, and from a left invariant Haar measure one can construct the right invariant Haar measure.

We need also the following measure-theoritic notions for further developments. Let G be a locally compact group and let B be the σ-algebra of compact subsets of G.

Definition 2.1.1. (Relatively left invariant measure). A measure ν on (G, B) is relatively left invariant with left multiplier $\chi(g)$ if ν satisfies

$$\nu(gB) = \chi(g)\nu(B), \quad \text{for all } B \in B. \tag{2.5a}$$

The multiplier $\chi(g)$ is a continuous homomorphisim from $G \to R^+$. If ν is relatively left invariant with left multiplier $\chi(g)$ then $\chi(g^{-1}) = 1/\chi(g)$ and $\chi(g^{-1})\nu(dg)$ is a left invariant measure.

Definition 2.1.2. A group G acts topologically on the left of χ if the mapping $T(g, x) : \chi \times G \to \chi$ is continuous.

Definition 2.1.3. (Proper G-space). Let the group G acts topologically on the left of χ and let h be a mapping on $G \times \chi \to \chi \times \chi$ given by

$$h(g, x) = (gx, x)$$

for $g \in G, x \in \chi$. The group G acts properly on χ if for every compact $C \subset \chi \times \chi$, $h^{-1}(C)$ is compact.

The space χ is called a proper G-space fG acts properly on χ. An equivalent concept of the proper G-space is the Cartan G-space.

Definition 2.1.4 (Cartan G-space). Let G acts topologically on χ. Then χ is called a Cartan G-space if for every $x \in \chi$ there exists a neighborhood V of x such that $(V, V) = \{g \in G | (gV) \cap V \neq \phi\}$ has a compact closure.

Wijsman (1967, 1978, 1985, 1986) demonstrate the verification of the condition of Cartan G-space and proper actions in a number of multivariate testing problems on covariance structures.

Example 2.1.1. Let G be a subgroup of the permutation group of a finite set χ. For any subset A of χ define $\lambda(A) =$ number of points in A. The measure

λ is an invariant measure under G and is unique upto a positive constant. It is known as counting measure.

Example 2.1.2. Let χ be the Euclidean n space and G the group of translations defined by $x_1 \in \chi, g_{x_1} \in G$ implying

$$g_{x_1}(x) = x + x_1 .$$

Clearly G acts transitively on χ. The n-dimensional Lebesgue measure λ is invariant under G. Again λ is unique upto a positive multiplicative constant.

Example 2.1.3. Consider the group $G_l(n)$ operating on itself from the left. Since an element g operating from the left on $G_l(n)$ is linear, it has a Jacobian with respect to this linear operation. Obviously, $G_l(n)$ is a vector space of dimension n^2. Let us make the convention to display a matrix of dimension $n \times n$ by column vectors of dimension $n \times 1$, by writing the first column horizontally. Then the second is likewise and so forth. In this manner a matrix of dimension $n \times n$ represents a point in R^{n^2}. Let $g, x, y \in G_l(n)$ and let $g = (g_{ij}), x = (x_{ij}), y = (y_{ij})$. If we display these matrices by their column vectors, the relationship $y = gx$ gives us the coordinates of the point

$$(y_{11}, \ldots, y_{1n}, \ldots, y_{n1}, \ldots, y_{nn})$$

as

$$y_{ij} = \sum_{k=1}^{n} g_{ik} x_{kj}$$

in terms of the coordinates of

$$(x_{11}, \ldots, x_{1n}, \ldots, x_{n1}, \ldots, x_{nn}) .$$

The Jacobian is thus the determinant of the $n^2 \times n^2$ matrix

$$\begin{pmatrix} g & 0 & \ldots & 0 \\ 0 & g & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & g \end{pmatrix} ,$$

where each 0 indicates null matrix of dimension $n \times n$. Hence the Jacobian of the transformation $y = gx$ is $(\det g)^n$. Let us now define for $y \in G_l(n)$

$$d\lambda(y) = \frac{\prod dy_{ij}}{(\det y)^n}$$

Then

$$\begin{aligned} d\lambda(gy) &= \frac{(\det g)^n \prod dy_{ij}}{(\det(gy))^n} \\ &= \frac{\prod dy_{ij}}{(\det y)^n} \\ &= d\lambda(y) \end{aligned}$$

Hence λ is an invariant measure on $G_l(n)$ under $G_l(n)$. It is also unique upto a positive multiplicative constant.

Example 2.1.4. Let G be a group of $n \times n$ nonsingular lower triangular matrices with positive diagonal elements and let G operate from the left on itself. We will verify that the Jacobian of the transformation

$$g \rightarrow hg$$

for $h, g \in G$, is $\prod_1^n (h_{ii})^i$. Thus an invariant measure on G under G which is unique upto a positive multiplicative constant, is

$$d\lambda(g) = \frac{\prod_{i \geq j} dg_{ij}}{\prod_{i=1}^n (g_{ii})^i}$$

To show that the Jacobian is $\prod_{i=1}^n (h_{ii})^i$, let $g = (g_{ij}), h = (h_{ij})$ with $g_{ij} = h_{ij} = 0$ for $i < j$. Then

$$hg = \left(\sum_{k=1}^n h_{ik} g_{kj} \right) = (C_{ij})$$

This defines a transformation

$$(g_{ij}) \rightarrow (C_{ij}) .$$

We can write the Jacobian of this transformation as the determinant of the matrix

$$\begin{pmatrix} \dfrac{\partial C_{11}}{\partial g_{11}} & \dfrac{\partial C_{11}}{\partial g_{21}} & \cdots & \dfrac{\partial C_{11}}{\partial g_{nn}} \\ \dfrac{\partial C_{21}}{\partial g_{11}} & \dfrac{\partial C_{21}}{\partial g_{21}} & \cdots & \dfrac{\partial C_{21}}{\partial g_{nn}} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{\partial C_{nn}}{\partial g_{11}} & \dfrac{\partial C_{nn}}{\partial g_{21}} & \cdots & \dfrac{\partial C_{nn}}{\partial g_{nn}} \end{pmatrix},$$

which is a $[n(n+1)]/2 \times [n(n+1)]/2$ lower triangular matrix with the diagonal elements

$$\frac{\partial C_{km}}{\partial g_{km}} = h_{kk}, k \geq m .$$

In other words h_{11} will appear once as a diagonal element, h_{22} will appear twice as a diagonal element and finally h_{nn} will appear n times as a diagonal element. Hence the Jacobian is $\prod_{i=1}^{n}(h_{ii})^{i}$.

Example 2.1.5. Let $\bar{S}$ be the space of all positive definite symmetric matrices S of dimension $n \times n$ and let $G_l(n)$ operate on S as

$$S \to gSg', \quad S \in \bar{S}, \quad g \in G_l(n) . \tag{2.6}$$

To show that the Jacobian of this transformation is $(\det g)^{n+1}$, let E_{ij} be the matrix obtained from the $n \times n$ identity matrix by interchanging the ith and the jth row; $M_i(C)$ be the matrix obtained from the $n \times n$ identity matrix by multiplying the ith row by any non-zero constant C and let A_{ij} be the matrix obtained from the $n \times n$ identity matrix by adding the jth row to the ith. We need only show this for the matrices $E_{ij}, M_i(C)$ and A_{ij}. This can be easily verified by the reader; for example, $\det(M_i(C)) = C$ and $M_i(C)SM_i(C)$ is obtained from $S = (S_{ij})$ by multiplying S_{ii} by C^2 and S_{ij} for $i \neq j$ by C so that the Jacobian is $C^{2+n-1} = C^{n+1}$.

Let us define the measure λ on $\bar{S}$ by

$$d\lambda(S) = \frac{\prod dS_{ij}}{(\det S)^{(n+1)/2}} .$$

Clearly,

$$d\lambda(S) = d\lambda(gSg') ,$$

so that λ is an invariant measure on S under the transformation (2.6).

Here also λ is unique upto a positive multiplicative constant. The group $G_l(p)$ acts transitively on $\bar{S}$.

Example 2.1.6. Let H be a subgroup of G. The group G acts transitively on the space $\chi = G/H$ (the quotient group of H) where group action satisfies

$$g_1(gh) = g_1 gH, \quad \text{for} \quad g_1 \in G, \quad g \in G .$$

When the group G acts transitively on χ, there is a one to one correspondence between χ and the quotient space G/H of any subgroup H of G.

2.1.1. *Transformation of variable in abstract integral*

Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and let $(\mathcal{Y}, \mathcal{C})$ be a measurable space. Suppose φ is a measurable function on $\mathcal{X}$ into $\mathcal{Y}$. Let us define the measure ν on $(\mathcal{Y}, \mathcal{C})$ by

$$\nu(C) = \mu(\varphi^{-1}(C)), \quad C \in \mathcal{C}. \tag{2.7}$$

Theorem 2.1.1. *If f is a real measurable function on $\mathcal{X}$ such that $f(x) = g(\varphi(x))$ and if f is integrable, then*

$$\begin{aligned} \int_x f(x) d\mu(x) &= \int_x g(\varphi(x)) d\mu(x) \\ &= \int_{\mathcal{Y}} g(y) d\nu(y) \end{aligned} \tag{2.8}$$

For a proof of this theorem, the reader is referred to Lehmann [(1959), p. 38].

Example 2.1.6. (Wishart distribution). Let $X_1, \ldots, X_n$ be n independent and identically distributed normal p-vectors with mean 0 and covariance matrix Σ (positive definite). Assume $n \geq p$ so that

$$S = \sum_{i=1}^{n} X_i X_i'$$

is positive definite almost everywhere. The joint probability density function of $X_1, \ldots, X_n$ is given by

$$p(x_1, \ldots, x_n) = (2\pi)^{-np/2} (\det \Sigma^{-1})^{n/2} \times \exp\left\{-\frac{1}{2} \text{tr } \Sigma^{-1} \sum_{i=1}^{n} x_i x_i'\right\}. \tag{2.9}$$

Using Theorem 2.1.1, we get for any measurable set A in the space of S

$$\begin{aligned} P(S \in A) &= (2\pi)^{-np/2} \int (\det \Sigma^{-1})^{n/2} \\ &\qquad s = \Sigma_{i=1}^{n} x_i x_i' \in A \exp\left\{-\frac{1}{2} \text{tr } \Sigma^{-1} s\right\} \prod_{i=1}^{n} dx_i, \qquad (2.10) \\ &= (2\pi)^{-np/2} \int_{s \in A} (\det \Sigma^{-1})^{n/2} \exp\left\{-\frac{1}{2} \text{ tr } \Sigma^{-1} s\right\} dm(s) \end{aligned}$$

where m is the measure corresponding to the measure ν of Theorem 2.1.1. Let us define the measure m^* by

$$dm^*(s) = \frac{dm(s)}{(\det s)^{n/2}}. \tag{2.11}$$

To find the distribution of S, which is popularly called Wishart distribution ($W(\Sigma, n)$) with parameter Σ and degrees of freedom n, it is sufficient to find dm^*. To do this let us first observe the following:

(i) Since Σ is positive definite symmetric, there exists $\alpha \in G_l(p)$ such that $\Sigma = \alpha\alpha'$.

(ii) Let

$$\tilde{S} = \alpha^{-1} S \alpha'^{-1} = \sum_{i=1}^{n} (\alpha^{-1} X_i)(\alpha^{-1} X_i)'.$$

As $\alpha^{-1}X_i$'s are independently normally distributed with mean vector 0 and covariance matrix I (identity matrix), S is distributed as $W(I, n)$. Thus

$$\begin{aligned} P_{\alpha\alpha'}(S \in A) &= (2\pi)^{-np/2} \int_{s \in A} (\det(\alpha\alpha')^{-1} s))^{n/2} \\ &\quad \times \exp\left\{-\frac{1}{2}\operatorname{tr}(\alpha\alpha')^{-1} s\right\} dm^*(s); \\ P_I(\alpha\tilde{S}\alpha' \in A) &= (2\pi)^{-np/2} \int_{s \in A} (\det((\alpha\alpha')^{-1}\, s))^{n/2} \\ &\quad \times \exp\left\{-\frac{1}{2}\operatorname{tr}(\alpha\alpha')^{-1} s\right\} \times dm^*(\alpha^{-1} s \alpha'^{-1}) \end{aligned}$$

Since $P_{\alpha\alpha'}(S \in A) = P_I(\alpha\tilde{S}\alpha' \in A)$ for all measurable sets A in the space of S, we must have for all $g \in G_l(n)$

$$dm^*(s) = dm^*(gsg'). \tag{2.12}$$

By example 2.1.5 we get

$$dm^*(s) = \frac{C \prod_{i \geq j} ds_{ij}}{(\det\, s)^{(p+1)/2}},$$

where C is a positive constant.

Thus the probability density function of the Wishart random variable is given by

$$W(\Sigma, n) = C(\det \Sigma^{-1})^{n/2}(\det s)^{(n-p-1)/2} \exp\left\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1}s\right\} \tag{2.13}$$

where C is the universal constant depending on Σ. To evaluate C let us first observe that C is a function of n and p. Let us write it as $C = C_{n,p}$. Since $C_{n,p}$ is a constant for all values of Σ, we can, in particular, take $\Sigma = I$ to evaluate C_{np}. Write

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix},$$

where S_{11} is $(p-1) \times (p-1)$. Make the transformations

$$\begin{aligned} S_{11} &= T_{11}, \\ S_{12} &= T_{12}, \\ S_{22} - S_{21}S_{11}^{-1}S_{12} &= T_{22} = Z(\text{say}), \end{aligned}$$

which has a Jacobian equal to unity.

Using the fact that

$$|S| = |S_{11}|\left|S_{22} - S_{21}S_{11}^{-1}S_{12}\right|,$$

and

$$\int C_{n,p}|s|^{(n-p-1)/2}\exp\left\{-\frac{1}{2}\operatorname{tr} s\right\} ds = 1 \quad \text{for all } p, n,$$

we get

$$\begin{aligned} &\int C_{n,p}|s|^{(n-p-1)/2}\exp\left\{-\frac{1}{2}\operatorname{tr} s\right\} ds \\ &= C_{n,p}\int |s_{11}|^{(n-p-1)/2}\exp\left\{-\frac{1}{2}\operatorname{tr} s_{11}\right\} ds_{11} \\ &\quad \times \int (z)^{(n-p-1)/2}\exp\left\{-\frac{1}{2}z\right\} dz \\ &\quad \times \int \exp\left\{-\frac{1}{2}s_{1}^{'}2\, s_{11}^{-1}s_{12}\right\} ds_{12} \end{aligned}$$

$$
\begin{aligned}
&= C_{n,p}(2\pi)^{(p-1)/2}2^{(n-p-1)/2}\Gamma\left(\frac{n-p+1}{2}\right) \\
&\quad \times \int |s_{11}|^{(n-p-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\, s_{11}\right\} ds_{11} \\
&= \frac{C_{n,p}}{C_{n,(p-1)}}(2\pi)^{(p-1)/2}2^{(n-p+1)/2}\Gamma\left(\frac{n-p+1}{2}\right) \\
&\quad \times \int C_{n,(p-1)}|s_{11}|^{(n-p)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\, s_{11}\right\} ds_{11}\,.
\end{aligned}
$$

Since $C_{n,1} = 2^{-n/2}/\Gamma(n/2)$, we have

$$
C_{n,1} = \frac{\pi^{-p(p-1)/2}2^{-n/2}}{\Gamma(N-p+1)/2} C_{n,(p-1)} \frac{\pi^{-p(p-1)/2}2^{-np/2}}{\prod_{i=0}^{p-1}\Gamma\left(\frac{n-i}{2}\right)}\,.
$$

2.2. Invariance of Testing Problems

We have seen in the previous section that for the invariance of the statistical distribution, transformation $\bar{g}$ of $g \in G$ must satisfy $\bar{g}\theta \in \Omega$ for $\theta \in \Omega$.

Definition 2.2.1. (Invariance of parametric space). The parametric space Ω remains invariant under a group of transformation $G : \mathcal{X} \overset{\text{onto}}{\to} \mathcal{X}$, if

(i) for $g \in G$, the induced mapping $\bar{g}$ on Ω satisfies $\bar{g}\theta \in \Omega$ for all $\theta \in \Omega$;
(ii) for any $\theta' \in \Omega$ there exists $\theta \in \Omega$ such that $\bar{g}\theta = \theta'$.

An equivalent way of writing (i) and (ii) is

$$
\bar{g}\Omega = \Omega. \tag{2.14}
$$

It may be remarked that if P_θ for different values of $\theta \in \Omega$ are distinct then $\bar{g}$ is one-to-one. The following theorem will assert that the set of all induced transformations $\bar{g}, g \in G$ also forms a group.

Theorem 2.2.1. *Let g_1, g_2 be two transformations which leave Ω invariant. The transformations $g_2 g_1$ and g_1^{-1} defined by*

$$
\begin{aligned}
g_2 g_1(x) &= g_2(g_1(x)) \\
g_1^{-1} g_1(x) &= x
\end{aligned} \tag{2.15}
$$

for all $x \in \mathcal{X}$ *leave* Ω *invariant and* $\overline{g_2 g_1} = \bar{g}_2 \bar{g}_1$, $(\bar{g}_1^{-1}) = (\bar{g}_1)^{-1}$

Proof. We know that if the distribution of X is $P_\theta, \theta \in \Omega$, the distribution of gX is $P_{\bar{g}\theta}, \bar{g}\theta \in \Omega$. Hence the distribution of $g_2 g_1(X)$ is $P_{\bar{g}_2 \bar{g}_1 \theta}$, as $\bar{g}_1 \theta \in \Omega$ and g_2 leaves Ω invariant, $\overline{g_2 g_1}\theta \in \Omega$. Thus $g_2 g_1$ leaves Ω invariant and, obviously,

$$\overline{g_2 g_1} = \bar{g}_2 \bar{g}_1 \, .$$

Similarly the reader may find it instructive to verify the other assertion. □

Let us now consider the problem of testing the hypothesis $H_0 : \theta \in \Omega_{H_O}$ against the alternatives $H_1 : \theta \in \Omega_{H_1}$, where Ω_{H_0} and Ω_{H_1} are two disjoint subsets of Ω. Let G be a group of transformations which operate from the left on $\mathcal{X}$ satisfying conditions (i) and (ii) above and (2.14). In what follows we will assume that $g \in G$ is measurable which ensures that whenever X is a random variable then gX is also a random variable.

Definition 2.2.2. (Invariance of statistical problem). The problem of testing $H_0 : \theta \in \Omega_{H_0}$ against the alternative $H_1 : \theta \in \Omega_{H_1}$ remains invariant under a group of transformations G operating from the left on $\mathcal{X}$ if

(i) for $g \in G, A \in \mathcal{A}$, $P_\theta(A) = P_{\bar{g}_\theta}(gA)$,

(ii) $\bar{g}\Omega_{H_0} = \Omega_{H_0}, \bar{g}\Omega_{H_1} = \Omega_{H_1}$ for all $g \in G$.

2.3. Invariance of Statistical Tests and Maximal Invariant

Let $(\mathcal{X}, \mathcal{A}, \lambda)$ be a measure space. Suppose $p_\theta, \theta \in \Omega$, is the probability density function of the random variable $X \in \mathcal{X}$ with respect to the measure λ. Consider a group G operating on $\mathcal{X}$ from the left and suppose that the measure λ is invariant with respect to G, that is,

$$\lambda(gA) = \lambda(A) \quad \text{for all } A \in \mathcal{A}, g \in G \, .$$

Let $(\mathcal{Y}, \mathcal{C})$ be a measurable space and suppose $T(X)$ is a measurable mapping from $\mathcal{X}$ into $\mathcal{X}$.

Definition 2.3.1. (Invariant function). $T(X)$ is an invariant function on $\mathcal{X}$ under G operating from the left on $\mathcal{X}$ if $T(X) = T(gX)$ for all $g \in G, X \in \mathcal{X}$.

Definition 2.3.2. (Maximal invariant). $T(X)$ is a maximal invariant function on $\mathcal{X}$ under G if $T(X)$ is invariant under G and $T(X) = T(Y)$ for all $X. Y \in \mathcal{X}$ implies that there exists $g \in G$ such that $Y = gX$.

If a problem of testing $H_0 : \theta \in \Omega_{H_0}$ against the alternatives $H_1; \theta \in \Omega_{H_1}$ remains invariant under the group transformations G, it is natural to restrict attention to statistical tests which are also invariant under G. Let $\varphi(X), X \in \mathcal{X}$ be a statistical test for testing H_0 against H_1, that is, $\varphi(x)$ be the probability of rejecting H_0 when x is the observed sample point. If φ is invariant under G then φ must satisfy

$$\varphi(x) = \varphi(gx), \quad x \in \mathcal{X}, g \in G \tag{2.16}$$

A useful characterization of invariant test $\varphi(X)$ in terms of the maximal invariant $T(X)$ on $\mathcal{X}$ is given in the following theorem.

Theorem 2.3.1. *Let $T(X)$ be a maximal invariant on $\mathcal{X}$ under G. A test $\varphi(X)$ is invariant under the group G if and only if there exists a function h for which $\varphi(X) = h(T(X))$.*

Proof. Let $\varphi(X) = h(T(X))$. Obviously,

$$\varphi(x) = \varphi(g(x)), \quad g \in G, \ x \in \mathcal{X}.$$

Conversely, if $\varphi(x)$ is invariant and if $T(x) = T(y); x, y \in \mathcal{X}$ then there exists a $g \in G$ such that $y = gx$ and therefore $\varphi(x) = \varphi(y)$. $\square$

In general if φ is invariant and measureable, then $\varphi(x) = h(T(x))$ but h may not be a Borel measurable function. However if the range of the maximal invariant $T(X)$ is Euclidean and T is Borel measurable then $\mathcal{A}_1 = \{A : A \in \mathcal{A} \text{ and } gA = A \text{ for all } g \in G\}$ is the smallest σ-field with respect to which T is measurable and φ is invariant if and only if $\varphi(x) = h(T(x))$ where h is a Borel measurable function. This result is essentially due to Blackwell (1956).

Let $\bar{G}$ be the group of induced transformations on Ω (induced by the group G). Let $\upsilon(\theta), \theta \in \Omega$, be a maximal invariant on Ω under $\bar{G}$.

Theorem 2.3.2. *The distribution of $T(X)$ (or any function of $T(X)$) depends on Ω only through $\nu(\theta)$.*

Proof. Suppose $\nu(\theta_1) = \nu(\theta_2), \theta_1, \theta_2 \in \Omega$. Then there exists a $\bar{g} \in \bar{G}$ such that $\theta_2 = \bar{g}\theta_1$. Now for any measurable set C

$$\begin{aligned} P_{\theta_1}(T(X) \in C) &= P_{\theta_1}(T(gX)) \in C) \\ &= P_{\bar{g}\theta_1}(T(X) \in C) \\ &= P_{\theta_2}(T(X)) \in C) \end{aligned} \qquad \square$$

2.4. Some Examples of Maximal Invariants

Example 2.4.1. Let $X = (X_1, \ldots, X_n)' \in \mathcal{X}$ and let G be the group of translations $g_C(X) = (X_1 + C, \ldots, X_n + C)', -\infty < C < \infty$.

A maximal invariant on $\mathcal{X}$ under G is $T(X) = (X_1, -X_n, \ldots, X_{n-1} - X_n)$. Obviously, it is invariant. Suppose $x = (x_1, \ldots, x_n)', y = (y_1, \ldots, y_n)'$ and let $T(x) = T(y)$. Writing $C = y_n - x_n$ we get $y_i = x_i + C$ for $i = 1, \ldots, n$.

Another maximal invariant is the redundant $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ where $\bar{X} = \frac{1}{n}\sum_1^n X_i$. It may be further observed that if $n = 1$ there is no maximal invariant. The whole space forms a single orbit in the sense that given any two points in the space there exists $C \in G$ transforming one point into the other. In other words G acts transitively on $\mathcal{X}$.

Example 2.4.2. Let $\mathcal{X}$ be a Euclidean n-space, and let G be the group of scale changes, that is, $g_a \in G$,

$$g_a(X) = (aX_1, \ldots, aX_n)', \quad a > 0.$$

A maximal invariant on $\mathcal{X}$ under G is $\bar{X}/S$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$.

Example 2.4.3. Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ where $\mathcal{X}_1$ is the p-dimensional Euclidean space and $\mathcal{X}_2$ is the set of all $p \times p$ symmetric positive definite matrices. Let $X \in \mathcal{X}_1$, $S \in \mathcal{X}_2$. Write

$$\begin{aligned} X &= (X_{(1)}, \ldots, X_{(k)})', & X_{[i]} &= (X_{(1)}, \ldots, X_{(i)})'; \\ S &= \begin{pmatrix} S_{(11)} & \cdots & S_{(1k)} \\ \vdots & & \vdots \\ S_{(k1)} & \cdots & S_{(kk)} \end{pmatrix} & S_{[ii]} &= \begin{pmatrix} S_{(11)} & \cdots & S_{(1i)} \\ \vdots & & \vdots \\ S_{(i1)} & \cdots & S_{(ii)} \end{pmatrix}, \end{aligned}$$

where $X_{(i)}$ is $d_i \times 1$ and $S_{(ii)}$ is $d_i \times d_i$ such that $\sum_{i=1}^k d_i = p$. Let G be the group of all nonsingular lower triangular $p \times p$ matrices in the block form, $g \in G$

$$g = \begin{pmatrix} g_{(11)} & 0 & \cdots & 0 \\ g_{(21)} & g_{(22)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ g_{(k1)} & g_{(k2)} & \cdots & g_{(kk)} \end{pmatrix},$$

operating as

$$(X, S) \to (gX, gSg').$$

Let

$$f(X, S) = (R_1^*, \ldots, R_k^*),$$

where

$$\sum_{j=1}^{i} R_j^* = X'_{[i]} S_{[ii]}^{-1} X_{[i]} \quad i = 1, \ldots, k \tag{2.17}$$

Obviously, $R_i^* > 0$. Let us show that $R_1^*, \ldots, R_k^*$ is a maximal invariant on $\mathcal{X}$ under G. We shall prove it for the case $k = 2$. The general case can be proved in a similar fashion.

Let us first observe the following:

(i) If $X \to gX$, $S \to gSg'$ then $X_{(1)} \to g_{(11)} X_{(1)}$, $S_{(11)} \to g_{(11)} S_{(11)} g'_{(11)}$. Thus $(R_1^*, R_1^* + R_2^*)$ is invariant under G.

(ii) $X'S^{-1}X = X'_{(1)} S_{(11)}^{-1} X_{(1)} + (X_{(2)} - S_{(21)} S_{(11)}^{-1} X_{(1)})'$

$$(S_{(22)} - S_{(21)} S_{(11)}^{-1} S_{(12)})^{-1} (X_{(2)} - S_{(21)} S_{(11)}^{-1} X_{(1)}).$$

Now, suppose that for $X, Y \in \mathcal{X}_1$; $S, T \in \mathcal{X}_2$

$$(X'_{(1)} S_{(11)}^{-1} X_{(1)}, X'S^{-1}X) = (Y'_{(1)} T_{(11)}^{-1} Y_{(1)}, Y'T^{-1}Y), \tag{2.18}$$

where Y, T are similarly partitioned as X, S. We must now show that there exists a $g \in G$ such that $X = gY, S = gTg'$. Let us first choose

$$g = \begin{pmatrix} g_{(11)} & 0 \\ g_{(21)} & g_{(22)} \end{pmatrix}$$

with

$$\begin{aligned} g_{(11)} &= S_{(11)}^{-1/2} \\ g_{(22)} &= (S_{(22)} - S_{(21)} S_{(11)}^{-1}) S_{(12)})^{-1/2} \\ g_{(12)} &= g_{(22)} S_{(21)} S_{(11)}^{-1}. \end{aligned}$$

Then $gSg' = I$. Similarly choose

$$h = \begin{pmatrix} h_{(11)} & 0 \\ h_{(21)} & h_{(22)} \end{pmatrix}$$

such that $hTh' = I$. Since

$$X'_{(1)} S^{-1}_{(11)} X_{(1)} = Y'_{(1)} T^{-1}_{(11)} Y_{(1)}$$

implies

$$(g_{(11)} X_{(1)})' (g_{(11)} X_{(1)}) = (h_{(11)} Y_{(1)})' (h_{(11)} Y_{(1)}),$$

we conclude that there exists an orthogonal matrix O_1 or order $d_1 \times d_1$ such that

$$g_{(11)} X_{(1)} = O_1 h_{(11)} Y_{(1)} .$$

Since

$$\begin{aligned} X' S^{-1} X &= Y' T^{-1} Y, \\ S^{-1} &= g' g, \\ T^{-1} &= h' h, \\ (g_{(11)} X_{(1)})' (g_{(11)} X_{(1)}) &= (h_{(11)} Y_{(1)})' (h_{(11)} Y_{(1)}), \end{aligned}$$

then

$$||g_{(21)} X_{(1)} + g_{(22)} X_{(2)}||^2 = ||h_{(21)} Y_{(1)} + h_{(22)} Y_{(2)}||^2$$

so that there exists an orthogonal matrix O_2 of order $d_2 \times d_2$ such that

$$(g_{(21)} X_{(1)} + g_{(22)} X_{(2)}) = O_2 (h_{(21)} Y_{(1)} + h_{(22)} Y_{(2)}) .$$

Now let

$$\Gamma = \begin{pmatrix} O_1 & 0 \\ 0 & O_2 \end{pmatrix}$$

Then, obviously,

$$\begin{aligned} X &= g^{-1} \Gamma h Y \\ &= \alpha Y, \end{aligned}$$

where $\alpha \in G$ and

$$gSg' = \Gamma h T h' \Gamma' ,$$

so that

$$S = \alpha T \alpha'$$

Hence $(R_1^*, R_1^* + R_2^*)$ and equivalently (R_1^*, R_2^*) is a maximal invariant on χ under G.

Note that if G is the group of $p \times p$ nonsingular upper triangular matrices in the block form, that is, if $g \in G$

$$g = \begin{pmatrix} g_{(11)} & g_{(12)} & \cdots & g_{(1k)} \\ 0 & g_{(22)} & \cdots & g_{(2k)} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & g_{(kk)} \end{pmatrix}$$

where $g_{(ii)}$ is $d_i \times d_i$, then a maximal invariant on $\mathcal{X}$ under G is

$$(T_1^*, \ldots, T_k^*)$$

defined by

$$\sum_{j=i}^{k} T_j^* = (X_{(i)}, \ldots, X_{(k)})' \begin{pmatrix} S_{(ii)} & \cdots & S_{(ik)} \\ \vdots & & \vdots \\ S_{(ki)} & \cdots & S_{(kk)} \end{pmatrix}^{-1} (X_{(i)}, \ldots, X_{(k)}) i = 1, \ldots, k$$

2.5. Distribution of Maximal Invariant

Let $(\mathcal{X}, \mathcal{A}, \lambda)$ be a measure space and suppose $p_\theta, \theta \in \Omega$, is a probability density function with respect to λ. Consider a group G operating from the left on $\mathcal{X}$ and suppose that λ is a left invariant measure,

$$\lambda(A) = \lambda(gA), \quad g \in G, \quad A \in \mathcal{A}.$$

Let $(\mathcal{Y}, \mathcal{C})$ be a measurable space and suppose that $T(X)$ is a measurable mapping such that T is a maximal invariant on $\mathcal{X}$ under G. Let λ^* be a measure on $(\mathcal{Y}, \mathcal{C})$ induced by λ, such that

$$\lambda^*(C) = \lambda(T^{-1}(C)), \quad C \in \mathcal{C}$$

Also let P_2 be the probability measure on $(\mathcal{Y}, \mathcal{C})$ defined by

$$P_2(C) = \int_{T^{-1}(c)} p_\theta(x) d\lambda(x),$$

so that P_2 has a density function p^* with respect λ^*. Hence

$$\int_C p^*(T)d\lambda^*(T) = \int_{T^{-1}(C)} p(x)d\lambda(x), C \in \mathcal{C}$$

Assumption. There exists an invariant probability measure μ on the group G under G.

Lemma 2.5.1. *The density function p^* of the maximal invariant T with respect to λ^* is given by*

$$p^*(T) = \int p(gx)d\mu(g)\,. \tag{2.19}$$

Proof. Let f be any measurable function on $(\mathcal{Y}, \mathcal{C})$. Then

$$\int f(T)p^*(T)d\lambda^*(T) = \int f(T(x))p(x)d\lambda(x).$$

However,

$$\begin{aligned}
&\int f(T(x)) \int_G p(g(x))d\mu(g)d\lambda(x) \\
&\quad = \int_G \int f(T(x))p(g(x))d\lambda(x)d\mu(g) \\
&\quad = \int_G \int f(T(x))p(g(x))d\lambda(g(x))d\mu(g) \\
&\quad = \int f(T(y))p(y)d\lambda(y)\,.
\end{aligned}$$

Then, since the function

$$\int_G p(g(x))d\mu(g)$$

is invariant, we have

$$p^*(T) = \int_G p(g(x))d\mu(g)\,. \qquad \square$$

2.5.1. *Existence of an invaraint probability measure on $O(p)$ (group of $p \times p$ orthogonal matrices).*

Let $X = (X_1, \dots, X_p)$ be a $p \times p$ random matrix where the column vectors X_i (p-vectors) are independent and are identically distributed normal (p-vectors) with mean vector 0 and covariance matrix I (identity). Apply the Gram–Schmidt orthogonalization process to the columns of X to obtain the random orthogonal matrix $Y = (Y_1, \dots, Y_p)$, where

$$Y_i = \frac{X_i - \sum_{j=1}^{i-1}(X_i'Y_j)Y_j}{\|X_i - \sum_{j=1}^{i-1}(X_i'Y_j)Y_j\|}, \quad i = 1, \dots, p.$$

Define $Y = T(X)$ and let $G = O(p)$. Obviously,

$$\begin{aligned} gY &= (gY_1, \dots, gY_p) \\ &= T(gX_1, \dots, gX_p) \\ &= T(X_1^*, \dots, X_p^*), \end{aligned}$$

where

$$X_i^* = gX_i, \quad i = 1, \dots, p.$$

Since X_i, X_i^* have the same distribution and $X_1^*, \dots, X_p^*$ are independently normally distributed with mean 0 and covariance matrix I, we obtain Y and gY have the same distribution for each $g \in O(p)$. Thus the probability measure of Y defined by

$$P(Y \in C) = P(X \in T^{-1}(C))$$

is invariant under $O(p)$.

2.6. Applications

Example 2.6.1. (Non-central Wishart distribution). Let $X = (X_1, \dots, X_n)$ (each X_i is a p-vector) be a $p \times n$ matrix ($n \geq p$) where each column X_i is normally distributed with mean vector μ_i and covariance matrix I and is independent of the remaining columns of X. The probability density function of X with respect to the pn-dimensional Lebesgue measure is

$$p(x, \mu) = \frac{1}{(2\pi)^{np/2}} \exp\left\{-\frac{1}{2}\operatorname{tr} xx' - \frac{1}{2}\operatorname{tr}\mu\mu' + \operatorname{tr} x\mu'\right\} \tag{2.19}$$

where

$$\mu = (\mu_1, \ldots, \mu_p)\,.$$

It is well known that

$$S = XX' = \sum_{i=1}^{n} X_i X_i'$$

is a maximal invariant under the transformation

$$X \to X\Gamma, \quad \Gamma \in O(n)\,,$$

where $\mathrm{O}(n)$ is the group of orthogonal matrices of dimension $n \times n$. Let $p(s, \mu)$ be the probability density function of S at the parameter point μ with respect to the induced measure $d\lambda^*(S)$ as defined in Lemma 2.5.1. To find the probability density function of S with respect to d we first find $d\lambda^*(s)$. From (2.13)

$$p(s,0)d\lambda^*(s) = C(\det s)^{(n-p-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\ s\right\} ds. \tag{2.20}$$

By (2.18)

$$\begin{aligned} p(s,0) &= \int_{0(n)} p(x\Gamma, 0)\mathrm{d}\mu(\Gamma) \\ &= \frac{1}{(2\pi)^{np/2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\ s\right\} \int_{O(n)} \mathrm{d}\mu(\Gamma) \\ &\quad \times \frac{1}{(2\pi)^{np/2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\ s\right\}, \end{aligned}$$

where μ is the invariant probability measure on $\mathrm{O}(n)$. Hence,

$$d\lambda^*(s) = (2\pi)^{np/2}C(\det s)^{(n-p-1)/2}ds\,.$$

Again by (2.18)

$$\begin{aligned} p(s,\mu) &= \exp\left\{-\frac{1}{2}\mathrm{tr}(s+\mu\mu')\right\} \int_{O(n)} \exp\left\{-\frac{1}{2}\mathrm{tr}\ x\Gamma\mu'\right\} d\mu(\Gamma) \\ &= \exp\left\{-\frac{1}{2}\mathrm{tr}(s+\mu\mu')\right\} h(x,\mu)\,, \end{aligned}$$

where

$$h(x,\mu) = \int_{O(n)} \exp\left\{-\frac{1}{2}\text{tr } x\Gamma\mu'\right\} du(\Gamma)$$

It is evident that

$$h(x,\mu) = h(x\Gamma,\mu\Psi),$$

where $\Gamma, \Psi \in 0(n)$. Thus we can write $h(x,\mu)$ as

$$h(x,\mu) = f(s,\mu\mu').$$

The probability density function of S with respect to the Lebesgue measure ds is given by

$$p(s,\mu)d\lambda^*(s) = p(s,\mu)(2\pi)^{np/2}C(\det s)^{(n-p-1)/2}ds\,.$$

This is called the probability density function of the noncentral Wishart distribution with non-centrality parameter $\mu\mu'$.

Example 2.6.2. (Non-central distribution of the characteristic roots). Let S of dimension $p \times p$ be distributed as central Wishart random variable with parameter Σ and degrees of freedom $n \geq p$. Let $R_1 > R_2 > \cdots > R_p > 0$ be the roots of characteristic equation $\det(S - \lambda I) = 0$, and let P denote a diagonal matrix with diagonal elements $R_1, \ldots, R_p$. Denote by $p(R,\Sigma)$ the probability density function of R at the parameter point Σ with respect to induce measure $d\lambda^*(R)$ of Lemma 2.5.1. We know [see, for example, Giri(1977)]

$$p(R,I)d\lambda^*(R) = C_1\left(\prod_{i=1}^{p} R_i\right)^{(n-p-1)/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{p} R_i\right\} \times \prod_{i<j}(R_i - R_j)dR$$

where dR stands for the Lebesgue measure and C_1 is the universal constant. It may be checked that R is a maximal invariant under the transformation

$$S \to \Gamma S\Gamma', \quad \Gamma \in O(p).$$

By (2.18)

$$p(R,I) = C_2\left(\prod_{i=1}^{p} R_i\right)^{(n-p-1)/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{p} R_i\right\} \int_{O(p)} d\mu(\Gamma).$$

So

$$d\lambda^*(R) = \frac{C_2}{C_1} \prod_{i<j} (R_i - R_j) dR.$$

Again by (2.18)

$$\begin{aligned} p(R, \Sigma) &= (\det \Sigma^{-1})^{n/2} \left(\prod_{i=1}^{p} R_i \right)^{(n-p-1)/2} \\ &\quad \times \int_{O(p)} \exp\left\{ -\frac{1}{2} \text{tr } \Sigma^{-1} \Gamma R \Gamma' \right\} d\mu(\Gamma) \\ &= (\det \theta^{-1})^{n/2} \left(\prod_{i=1}^{p} R_i \right)^{(n-p-1)/2} \\ &\quad \times \int_{O(p)} \exp\left\{ -\frac{1}{2} \text{tr } \theta^{-1} \Gamma R \Gamma' \right\} d\mu(\Gamma) \end{aligned}$$

where θ is a diagonal matrix whose diagonal elements are the characteristic roots of Σ. The non-central probability density function of R depends on Σ only through its characteristic roots.

2.7. The Distribution of a Maximal Invariant in the General Case

Invariant tests depend on the observations only through the maximal invariant. To find the optimum test we need to find the explicit form of the maximal invariant statistic and its distribution. For many multivariate testing problems it is not always convenient to find the explicit form of the maximal invaraint. Stein (1956) gave the following representation of the ratio of probability densities of a maximal invariant with respect to a group G of transformations g leaving the testing problem invariant.

Theorem 2.7.1. *Let G be a group operating from the left on a topological space $(\mathcal{X}, \mathcal{A})$ and λ a measure in χ which is left invariant under G (G is not necessarily transitive on $\mathcal{X}$). Assume that there are two given probability densities p_1, p_2 with respect to λ such that*

$$P_1(A) = \int_A p_1(x) d\lambda(x),$$

$$P_2(A) = \int_A p_2(x) d\lambda(x),$$

for $A \in \mathcal{A}$ and P_1 and P_2 are absolutely continuous. Suppose $T : \mathcal{X} \to \mathcal{X}$ is a maximal invariant and p_i^* is the distribution of $T(X)$ when X has distribution P_i. Then under certain conditions

$$\frac{dP_2^*(T)}{dP_2^*(T)} = \frac{\int_G p_2(gx)d\mu(g)}{\int_G p_1(gx)d\mu(g)}, \tag{2.21}$$

where μ is a left invariant Haar measure on G. An often used alternative form of (2.21) is given by

$$\frac{dP_2^*(T)}{dP_1^*(T)} = \frac{\int_G f_2(gx)\chi(g)d\mu(g)}{\int_G f_1(gx)\chi(g)d\mu(g)} \tag{2.21a}$$

where f_i is the probability density with respect to a relatively invariant measure with multiplier $\mathcal{X}(g)$.

Stein (1956) gave the statements of Theorem 2.7.1. without stating explicitly the conditions under which (2.21) holds. However this representation was used by Giri (1961, 1964, 1965, 1971) and Schwartz (1967). Schwartz (1967) also gave a set of conditions (rather complicated) which must be satisfied for Stein's representation to be valid. Wijsman (1967, 1969) gave a sufficient condition for (2.21) using the concept of Cartan G-space. Koehn (1970) gave a generalization of the results of Wijsman (1967). Bonder (1976) gave some conditions for (2.21) through topological arguments. Anderson (1982) has obtained some results for the validity of (2.21) in terms of the proper action of the group. Wijsman (1985) studied the properness of several groups of transformations used in several multivariate testing problems. Subsequently Wijsman (1986) investigated the global cross-section technique for factorization of measures and applied it to the representation of the ratio of densities of a maximal invariant.

2.8. An Important Multivariate Distribution

Let $X_1, \ldots, X_N$ be N independently, identically distributed p-dimensional normal random variables with mean $\xi = (\xi, \ldots, \xi_p)'$ and covariance matrix Σ (positive definite). Write

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i, \quad S = \sum^{N}(X_i - \bar{X})(X_i - \bar{X})'.$$

It is well-known that $\bar{X}, S$ are independent in distribution and $\bar{X}$ has p-dimensional normal distribution with mean ξ and covariance matrix $(1/N)$ Σ and S has central Wishart distribution with parameter Σ and degrees of freedom $n = N - 1$. Throughout this section we will assume that $N > p$ so that S is positive definite almost everywhere. Write for any p-vector b,

$$b = (b_{(1)}, \ldots, b_{(k)})', \quad b_{[i]} = (b_{(1)}, \ldots, b_{(i)})',$$

where $b_{(i)}$ is the $d_i \times 1$ subvector of b such that $\sum_1^k d_i = p$ and for any $p \times p$ matrix A

$$A = \begin{pmatrix} A_{(11)} & \cdots & A_{(1k)} \\ \vdots & & \vdots \\ A_{(k1)} & \cdots & A_{(kk)} \end{pmatrix}, \qquad A_{[ii]} = \begin{pmatrix} A_{(11)} & \cdots & A_{(1i)} \\ \vdots & & \vdots \\ A_{(i1)} & \cdots & A_{(ii)} \end{pmatrix},$$

where $A_{(ii)}$ is a $d_i \times d_i$ submatrix of A. Let us define $R_1, \ldots, R_k$ and $\delta_1, \ldots, \delta_k$ by

$$\sum_{j=1}^{i} R_j = N\bar{X}_{[i]}'(S_{[ii]} + N\bar{X}_{[i]}\bar{X}_{[i]})^{-1}\bar{X}_{[i]}$$

$$\sum_{j=1}^{i} \delta_j = N\xi_{[i]}'\Sigma_{[ii]}^{-1}\xi_{[i]}, \quad i = 1, 2, \ldots, k.$$

Theorem 2.8.1. *The joint probability density function of $R_1, \ldots, R_k$ is given by (for $R_i > 0, \sum_{j=1}^k R_j < 1$)*

$$\begin{aligned} f(R_i, \ldots, R_k) = & \frac{\Gamma(N/2)}{\prod_{i=1}^{k}\Gamma\left(\frac{d_i}{2}\right)\Gamma\left(\frac{N-p}{2}\right)} \\ & \times \prod_{i=1}^{k} R^{(d_i/2)-1}\left(1 - \sum_{i=1}^{k} R_i\right)^{[(N-p)/2]-1} \\ & \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{k}\delta_i + \frac{1}{2}R_j\sum_{i>j}\delta_i\right\} \\ & \times \prod_{i=1}^{k}\varphi\left(\frac{N-\delta_{j-1}}{2}, \frac{d_i}{2}; \frac{R_i\delta_i}{2}\right) \end{aligned} \tag{2.22}$$

where $\sigma_i = \sum_{j=1}^{i} d_j$ *with* $\sigma_0 = 0$ *and* $\varphi(a, b : x)$ *is the confluent hypergeometric series given by*

$$1 + \frac{a}{b}x + \frac{a(a+1)}{b(b+1)}\frac{x^2}{2!} + \frac{a(a+1)(a+2)}{b(b+1)(b+2)}\frac{x^3}{3!} + \cdots.$$

Furthermore, the marginal probability density function of $R_1, \ldots, R_j, j < k$ can be obtained from (2.22) by replacing k by j and p by $\sum_{i=1}^{j} d_i$.

Proof. We will first prove the theorem for the case $k = 2$ and then use this result for the case $k = 3$. The proof for the general case will follow from these cases. For the case $k = 2$ consider the random variables

$$\begin{aligned} W &= S_{(22)} - S_{(21)}S_{(11)}^{-1}S_{(12)}\,, \\ L &= S_{(21)}S_{(11)}^{-1/2} \\ V &= S_{(11)}\,. \end{aligned} \tag{2.23}$$

It is well-known [see, for example, Stein (1956), or Giri(1977), p. 154] that

(i) V is distributed as Wishart $W(N-1, \Sigma_{(11)})$;
(ii) W is distributed as Wishart $W(N-1-d_1, \Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)})$;
(iii) the conditional distribution of L given V is multivariate normal with mean $\Sigma_{(21)}\Sigma_{(11)}^{-1}V^{-1/2}$ and covariance matrix

$$(\Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)})\bar{\otimes} I_{d_1}\,,$$

where $\bar{\otimes}$ is the tensor product of two matrices $\Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}$ and the $d_1 \times d_1$ identity matrix I_{d_1}.

(iv) W is independent of (L, V).

Let us now define W_1, W_2 by

$$\begin{aligned} W_1 &= N\bar{X}_{(1)}'S_{(11)}^{-1}\bar{X}_{(1)}\,, \\ W_2(1+W_1) &= N(\bar{X}_{(2)} - S_{(21)}S_{(11)}^{-1}\bar{X}_{(1)})'(S_{(22)} - S_{(21)}S_{(11)}^{-1}S_{(12)})^{-1} \\ &\quad \times (\bar{X}_{(2)} - S_{(21)}S_{(11)}^{-1}\bar{X}_{(1)})\,. \end{aligned} \tag{2.24}$$

Since $\bar{X}$ is independent of S, the conditional distribution of $\sqrt{N}\bar{X}_{(2)}$ given $S_{(11)}$ and $\bar{X}_{(1)}$ is normal with mean $\sqrt{N}(\xi_{(2)} + \Sigma_{(21)}\Sigma_{(11)}^{-1}\xi_{(1)})$ and covariance matrix

$\Sigma_{(22)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}$. It thereupon follows that the conditional distribution of $\sqrt{N}\bar{X}_{(2)}$ given $S_{(11)}$ and $\sqrt{N}\bar{X}_{(1)}$ is independent of the conditional distribution of

$$\sqrt{N}S_{(21)}S_{(11)}^{-1}\bar{X}_{(1)}\,,$$

given $S_{(11)}$ and $\bar{X}_{(1)}$. Furthermore, the conditional distribution of $\sqrt{N}S_{(12)}$ $S_{(11)}^{-1}\bar{X}_{(1)}$ given $S_{(11)}$ and $\bar{X}_{(1)}$ is normal with mean $\sqrt{N}\Sigma_{(21)}\Sigma_{(11)}^{-1}\xi_{(1)}$ and covariance matrix

$$N\bar{X}_{(1)}'S_{(11)}^{-1}\bar{X}_{(1)}\left(\Sigma_{(21)}\Sigma_{(11)}^{-1}\Sigma_{(12)}\right).$$

Hence the conditional distribution of

$$\sqrt{N}\left(\bar{X}_{(2)} - S_{(21)}S_{(11)}^{-1}\bar{X}_{(1)}\right)(1+W_1)^{-1/2}$$

given $S_{(11)}$ and $\bar{X}_{(1)}$ is normal with mean

$$\sqrt{N}\left(\xi_{(2)} - \Sigma_{(21)}\Sigma_{(11)}^{-1}\xi_{(1)}\right)(1+W_1)^{-1/2}$$

and covariance matrix

$$\Sigma_{(22)} - \Sigma_{(21)}\,\Sigma_{(11)}^{-1}\,\Sigma_{(12)}\,.$$

From Giri (1977) it follows that the conditional distribution of W_2 given $\bar{X}_{(1)}$ and $S_{(11)}$ is

$$p(W_2|\bar{X}_{(1)}, S_{(11)}) = \chi^2_{d_2}\left(\delta_2(1+W_1)^{-1}\right)/\chi^2_{N-d_1-d_2}\,, \tag{2.25}$$

where $\chi^2_n(\delta)$ denotes the noncentral chisquare random variable with noncentrality parameter δ and with n degrees of freedom and χ^2_n denotes the central chisquare random variable with n degrees of fredom. Since this conditional distribution depends on $\bar{X}_{(1)}$ and $S_{(11)}$ only through W_1, we get the joint probability density of W_1, W_2 as

$$p(W_2, W_1) = \chi^2_{d_2}(\delta_2(1+W_1)^{-1})/\chi^2_{N-d_1-d_2}, \tag{2.26}$$

and furthermore it is well known that the marginal probability density function of W_1 is

$$p(W_1) = \chi^2_{d_1}(\delta_1)/\chi^2_{N-d_1}\,. \tag{2.27}$$

Hence, we have

$$\begin{aligned} p(W_1, W_2) = {} & \exp\left\{-\frac{1}{2}\delta_2(1+W_1)^{-1}\right\} \\ & \times \sum_{\beta=0}^{\infty} \frac{\left(\frac{1}{2}\delta_2(1+W_1)^{-1}\right)^{\beta}}{\beta!} \frac{(W_2)^{\frac{1}{2}d_2+\beta-1}}{(1+W_2)^{\frac{1}{2}(N-d_1)+\beta}} \\ & \times \frac{\Gamma\left(\frac{n-d_1}{2}+\beta\right)}{\Gamma\left(\frac{d_2}{2}+\beta\right)\Gamma\left(\frac{N-p}{2}\right)} \exp\left\{-\frac{1}{2}\delta_1\right\} \sum_{j=0}^{\infty} \frac{\left(\frac{\delta_1}{2}\right)^j}{j!} \\ & \times \frac{(W_1)^{\frac{1}{2}d_1+j-1}}{(1+W_1)^{\frac{1}{2}N+j}} \frac{\Gamma(\frac{1}{2}N+j)}{\Gamma(\frac{1}{2}d_1+j)\Gamma(N/2)} \end{aligned} \tag{2.28}$$

From Lemma 2.8.1. below,

$$\begin{aligned} W_1 &= R_1(1-R_1)^{-1} \\ W_2 &= \{(R_1+R_2)(1-R_1-R_2)^{-1} - R_1(1-R_1)^{-1}\} \\ &\quad \times (1+R_1(1-R_1)^{-1})^{-1} \\ &= R_2(1-R_1-R_2)^{-1}. \end{aligned} \tag{2.29}$$

From (2.28) and (2.29) the probability density function of R_1 and R_2 is given by

$$\begin{aligned} p(R_1, R_2) = {} & \Gamma\left(\frac{1}{2}N\right)\left(\Gamma\frac{1}{2}(N-p)\Gamma\left(\frac{1}{2}d_1\right)\Gamma\left(\frac{1}{2}d_2\right)\right)^{-1} \\ & \times R_1^{\frac{1}{2}d_1-1} R_2^{\frac{1}{2}d_2-1}(1-R_1-R_2)^{\frac{1}{2}(N-p)-1} \\ & \times \exp\left\{-\frac{1}{2}(\delta_1+\delta_2)+\frac{1}{2}\delta_2 R_1\right\} \\ & \times \prod_{i=1}^{2} \varphi\left(\frac{1}{2}(N-\sigma_{i-1}), \frac{d_i}{2}; \frac{1}{2}R_i\delta_i\right) \end{aligned} \tag{2.30}$$

We now consider the case $k = 3$. Write

$$W_3 = \left(N\bar{X}'S^{-1}\bar{X} - N\bar{X}'_{[2]}S^{-1}_{[22]}\bar{X}'_{[2]}\right)\left(1 + N\bar{X}'_{[2]}S^{-1}_{[22]}\bar{X}_{[2]}\right)^{-1}.$$

It is easy to conclude that

$$S_{(33)} - S_{[32]}S^{-1}_{[22]}S_{[23]}$$

is distributed as Wishart

$$W\left(N-1-d_1-d_2, \Sigma_{(33)}-\Sigma_{[32]}\Sigma_{[22]}^{-1}\Sigma_{[23]}\right),$$

and is independent of $S_{(22)}$ and $S_{[32]}$. Furthermore, the conditional distribution of $\sqrt{N}\bar{X}_{(3)}$ given $S_{[22]}, X_{[2]}$ is normal with mean

$$\sqrt{N}\left(\xi_{(3)}-\Sigma_{[32]}\Sigma_{[22]}^{-1}(\bar{X}_{[2]}-\xi_{[2]})\right)$$

and covariance matrix

$$\Sigma_{(33)}-\Sigma_{[32]}\Sigma_{[22]}^{-1}\Sigma_{[23]},$$

and is independent of the conditional distribution of

$$\sqrt{N}S_{[32]}S_{[22]}^{-1}\bar{X}_{[2]}$$

given $S_{[22]}$ and $\bar{X}_{[2]}$, which is normal with mean

$$\sqrt{N}\Sigma_{[32]}\Sigma_{[22]}^{-1}\xi_{[2]}$$

and covariance matrix

$$N\bar{X}_{[2]}'S_{[22]}^{-1}\bar{X}_{[2]}\left(\Sigma_{[33]}-\Sigma_{[32]}\Sigma_{[22]}^{-1}\Sigma_{[23]}\right).$$

Thus the conditional distribution of W_3 given $S_{[22]}$ and $\bar{X}_{[2]}$ or, equivalently given (W_1, W_2), is

$$\begin{aligned} p(W_3|W_2, W_1) = X_{d_2}^2(\delta_3(1+W_1)(W_1+W_2+W_1W_2)^{-1}) \\ /\chi_{N-d_1-d_2-d_3}^2 . \end{aligned} \tag{2.31}$$

Hence the joint probability density function of W_1, W_2, W_3 is given by

$$\begin{aligned} p(W_1, W_2, W_3) &= \frac{\chi_{d_3}^2\left(\delta_3(1+W_1)(W_1+W_2+W_1W_2)^{-1}\right)}{\chi_{N-d_1-d_2-d_3}^2} \\ &\quad \times \frac{\chi_{d_2}^2\left(\delta_2(1+W_1)^{-1}\right)}{\chi_{N-d_1-d_2}^2} \cdot \frac{\chi_{d_1}^2(\delta_1)}{\chi_{N-d_1}^2} \end{aligned} \tag{2.32}$$

From Lemma 2.8.1. below,

$$\begin{aligned} W_1 &= R_1(1-R_1)^{-1}, \\ W_2 &= R_2(1-R_1-R_2)^{-1}, \\ W_3 &= ((R_1+R_2+R_3)(1-R_1-R_2-R_3)^{-1}-(R_1+R_2) \\ &\quad \times (1-R_1-R_2)^{-1})(1-R_1-R_2) \\ &= R_3(1-R_1-R_2-R_3)^{-1}. \end{aligned} \tag{2.33}$$

From (2.32)–(2.33) the joint probability density function of R_1, R_2, R_3 is given by

$$p(R_1, R_2, R_3) = \Gamma\left(\frac{1}{2}N\right)\left(\Gamma\left(\frac{1}{2}(N-p)\right)\prod_{i=1}^{3}\left(\frac{1}{2}d_i\right)\right)^{-1}$$
$$\times \prod_{i=1}^{3} R_i^{\frac{1}{2}d_i - 1}\left(1 - \sum_{j=1}^{3} R_i\right)^{\frac{1}{2}(N-p)-1}$$
$$\times \exp\left\{-\frac{1}{2}\sum_{1}^{3}\delta_i + \frac{1}{2}\sum_{j=1}^{3} R_j \sum_{i>j}\delta_i\right\}$$
$$\times \prod_{i=1}^{3}\varphi\left(\frac{1}{2}(N-\sigma_{i-1}), \frac{d_i}{2}; \frac{R_i\delta_i}{2}\right). \tag{2.34}$$

Proceeding exactly in the above fashion we get (2.22) for general k. Since the marginal distribution of $X_{[i]}$ is normal with mean $\xi_{[i]}$ and covariance matrix $\Sigma_{[ii]}$ it follows that given the joint probability density function of $R_1, \ldots, R_k$, the marginal distribution of $R_1, \ldots, R_j, 1 \le j \le p$ can be obtained from (2.22) by replacing k by j. □

The following lemma will show that there is one-to-one correspondence between $(R_1, \ldots, R_k)$ and $(R_1^*, \ldots, R_k^*)$ as defined earlier in this chapter.

Lemma 2.8.1.

$$N\bar{X}'(S + N\bar{X}\bar{X}')^{-1}\bar{X} = \frac{N\bar{X}'S^{-1}\bar{X}}{1 + N\bar{X}'S^{-1}\bar{X}}.$$

Proof. Let

$$(S + N\bar{X}\bar{X}')^{-1} = S^{-1} + A.$$

Hence,

$$I + N\bar{X}\bar{X}'S^{-1} + (S + N\bar{X}\bar{X}')A = I.$$

So we get

$$(S + N\bar{X}\bar{X}')^{-1} = S^{-1} - N(S + N\bar{X}\bar{X}')^{-1}\bar{X}\bar{X}'S^{-1}.$$

Now

$$N\bar{X}'(S + N\bar{X}\bar{X}')^{-1}\bar{X} = N\bar{X}'S^{-1}\bar{X} - N\bar{X}'(S + N\bar{X}\bar{X}')^{-1}\bar{X}(\bar{X}'S^{-1}\bar{X}).$$

Therefore,

$$N\bar{X}'(S + N\bar{X}\bar{X}')^{-1}\bar{X}) = \frac{N\bar{X}'S^{-1}\bar{X}}{1 + N\bar{X}'S^{-1}\bar{X}} \qquad \square$$

Note that

$$\sum_{j=1}^{i} R_j = \sum_{j=1}^{i} R_j^* / \left(1 + \sum_{j=1}^{i} R_j^*\right) . \tag{2.34}$$

2.9. Almost Invariance, Sufficiency and Invariance

Definition 2.9.1. Let $(\mathcal{X}, \mathcal{A}, P)$ be a measure space. A test $\psi(X), X \in \mathcal{X}$ is said to be equivalent to an invariant test with respect to the group of transformations G, if there exists an invariant test $\varphi(x)$ with respect to G such that

$$\psi(x) = \varphi(x) ,$$

for all $x \in \mathcal{X}$ except possibly for a subset N of $\mathcal{X}$ of probability measure zero.

Definition 2.9.2. Almost invariance. A test $\varphi(x)$, $x \in \mathcal{X}$, is said to be almost invariant with respect to the group of transformations G, if

$$\varphi(x) = \varphi(gx) \, \text{for all} \, x \in \mathcal{X} - \mathcal{N}_g ,$$

where N_g is a subset of χ depending on g of probability measure zero.

It is tempting to conjecture that if $\varphi(x)$ is almost invariant then it is equivalent to an invariant test $\psi(x)$. Obviously, any $\varphi(x)$ which is equivaleant to an invariant test is almost invariant. For, take $N_g = N \cup g^{-1}N$. If $x \notin N_g$, then $x \notin N$, and $gx \notin N$ so that

$$\varphi(x) = \varphi(gx) = \psi(gx) = \psi(x) .$$

Conversely, if G is countable or finite then given an almost invariant test $\varphi(x)$ we define

$$N = \bigcup_{g \in G} N_g ,$$

where $N_g = \{x \in \chi : \varphi(x) \neq \varphi(gx)\}$ and N has probability measure zero.

The uncountable G presents difficulties. Such examples are not rare where an almost invariant test can be different from an invariant test.

Let G be a group of transformations operating on $\mathcal{X}$ and let, $\mathcal{A}, \mathcal{B}$, be the σ-field of subsets of $\mathcal{X}$ and G respectively such that for any $A \in \mathcal{A}$ the set of pairs (x, g) for which $gx \in A$ is measurable $\mathcal{A} \times \mathcal{B}$. Suppose further that there exists a σ-finite measure ν on G such that for all $B \in \mathcal{B}, \nu(B) = 0$ implies $\nu(Bg) = 0$ for all $g \in G$. Then any measurable almost invariant test function with respect to G is equivalent to an invariant test under G.

For a proof of this fact the reader is referred to Lehmann (1959). The requirement that for all $g \in G$ and $B \in \mathcal{B}, \nu(B) = 0$ imply $\nu(Bg) = 0$ is satisfied in particular when $\nu(B) = \nu(Bg)$ for all $g \in G$ and $B \in \mathcal{B}$. Such a right invariant measure exists for a large number of groups.

Let $A_1 = \{A : A \in \mathcal{A} \text{ and } gA = A \text{ for all } g \in G\}$ and let $\mathcal{A}_s$ be the sufficient σ-field of $\mathcal{X}$. Let G be a group, leaving a testing problem invariant, be such that $g\mathcal{A}_s = \mathcal{A}_s$ for each $g \in G$. If we first reduce the problem by sufficiency, getting the measure space $(\mathcal{X}, \mathcal{A}_s, P)$, and subsequently by invariance then we arrive at the measure space $(\mathcal{X}, \mathcal{A}_{sI}, P)$ where $\mathcal{A}_{sI}$ is the class of all invariant $\mathcal{A}_s$ measurable sets. The following theorem which is essentially due to Stein and Cox provides an answer to our questions under certain conditions. The proof is given by Hall, Wijsman and Ghosh (1965).

Theorem 2.9.1. *Let $(\mathcal{X}, \mathcal{A}, P)$ be a measure space and let G be a group of measurable transformations leaving P invariant. Let A_s be a sufficient subfield for P on $(\mathcal{X}, \mathcal{A})$ such that $g\mathcal{A}_s = \mathcal{A}_s$ for each $g \in G$. Suppose that for each almost invariant $\mathcal{A}_s$ measurable function $\varphi = \psi$ except possibly on a set N of probability measure zero. Then $\mathcal{A}_s$ is sufficient for P on $(\mathcal{X}, \mathcal{A})$.*

Thus under conditions of the theorem if we reduce by sufficiency and invariance then the order in which it is done is immaterial.

2.10. Invariance, Type D and E Regions.

The notion of a type D or type E region is due to Issacson (1951). Kiefer (1958) showed that the F-test of the univariate general linear hypothesis possesses this property. Suppose, for a parametric space $\Omega = \{(\theta, \eta) : \theta \in \Omega', \eta \in H\}$ with associated distributions, with Ω' a Euclidean set, that every test function ϕ has a power function $\beta_\phi(\theta, \eta)$ which, for each η, is twice continuously differentiable in the components of θ at $\theta = 0$, an interior point of Ω'. Let Q_α be the class of locally strictly unbiased level α tests of $H_0 : \theta = 0$ against the alternatives $H_1 : \theta \neq 0$. Our assumption on β_ϕ implies that all tests in Q_α are similar and that

$$(\delta \beta_\phi / \delta \theta_i)\,|_{\theta=0} = 0$$

for all ϕ in Q_α.

Let $(\beta_\phi(\eta))$ be the matrix of second derivates of $\beta_\phi(\theta, \eta)$ with respect to the components of θ (which is also popularly known as the Gaussian curvature) at $\theta = 0$. Define

$$\Delta_\phi(\eta) = \det(\beta_\phi(\eta)).$$

Assumption. $\Delta_{\phi'}(\eta) > 0$ for all $\eta \in H$ and for at least one $\phi' \in Q_\alpha$.

Definition 2.10.1. (Type E and Type D Tests). A test ϕ^* in Q_α is said to be of type E if

$$\Delta_{\phi^*}(\eta) = \max_{\phi \in Q_\alpha} \Delta_\phi(\eta)$$

for all $\eta \in H$. If H is a single point, ϕ^* is said to be of type D.

Lehmann (1959a) showed that, in finding regions which are of type D, invariance could be invoked in the manner of Hunt–Stein theorem. (Theorem 5.0.1); and that this could also be done for type E regions (if they exist) provided that one works with a group which operates on H as the identity.

Suppose that our problem is invariant under a group of transformations for G which Hunt–Stein theorem holds and which acts trivially on Ω', such that for $g \in G$

$$g(\theta, \eta) = (\theta, g\eta)$$

If ϕg is the test function defined by

$$\phi g(x) = \phi(gx),$$

then a trivial computation shows that

$$\Delta_{\phi g}(\eta) = \Delta_\phi(g\eta)$$

and hence

$$\bar{\Delta}(\eta) = \bar{\Delta}(g\eta)$$

where

$$\bar{\Delta}(\eta) = \max_{\phi \in Q_\alpha} \Delta_\phi(\eta)\,.$$

Also, if ϕ is better than ϕ' in the sense of either type D or type E, then ϕg is clearly better than $\phi' g$.

Exercises

1. Let $X_1, \ldots, X_n$ be independently and identically distributed normal random variables with mean μ and variance (unknown) σ^2. Using Theorem 2.7.1. find the UMP invariant test of $H_0 : \mu = 0$ against the alternative $H_1 : \mu \neq 0$ with respect to the group of transformations which transform each $X_i \to cX_i, c \neq 0$.
2. Let $\{P_\theta, \theta \in \Omega\}$ be a family of distributions on $(\mathcal{X}, \mathcal{A})$ such that P_θ has a density $p(\cdot/\theta)$ with respect to some σ-finite measure μ. For testing $H_0 : \theta \in \Omega_0$ against $H_1 : \theta \in \Omega_1$ with $\Omega_0 \cap \Omega_1 = \phi$ (null set), the likelihood ratio test rejects H_0 whenever $\lambda(x)$ is less than a constant, where,
$$\lambda(x) = \frac{\sup_{\theta \in \Omega_0} p(x|\theta)}{\sup_{\theta \in \Omega_0 \cup \Omega_1} p(x|\theta)}.$$
Let the problem of testing H_0 against H_1 be invariant under a group G of transformations g transforming $x \to g(x)$, $g \in G, x \in \mathcal{X}$. Assume that
$$p(x|\theta) = p(gx|\bar{g}\theta)\mathcal{X}(g)$$
for some multiplier $\mathcal{X}$. Show that $\lambda(x)$ is an invariant function.
3. Prove (2.8) and (2.25).
4. Prove Theorem 2.9.1.

References

1. S. A. Anderson, *Distribution of maximal invariants using quotient measures*, Ann. Statist. **10**, pp. 955–961 (1982).
2. D. Blackwell, *On a class of probability spaces*, Proc. Berkeley Symp. Math. Statist. Probability. 3rd, Univ. of California Press, Berkeley, California, 1956.
3. J. V. Bonder, *Borel cross-sections and maximal invariants*, Ann. Statist. **4**, pp. 866–877 (1976).
4. T. Ferguson, *Mathematical Statisties*, Academic Press, 1969.
5. N. Giri, *On the likelihood ratio test of a normal multivariate testing problems*; Ph.D. Thesis, Stanford University, California, 1961.
6. N. Giri, *On the likelihood ratio test of a normal multivariate testing problems*, Ann. Math. Statist. **35**, pp. 181–189 (1964).
7. N. Giri, *On the likelihood ratio test of a normal multivariate testing problem-II*, Ann. Math. Statist. **36**, pp. 1061–1065 (1965).
8. N. Giri, *On the distribution of a multivariate statistic*, Sankhya, **33**, pp. 207–210 (1971).

9. N. Giri, *Multivariate Statistical Inference*, Academic Press, N.Y., 1977.
10. S. L. Isaacson, *On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters*, Ann. Math. Statist. 22, pp. 217–234 (1951).
11. A. Haar, *Der Massbegriff in der theorie der kontinwierchen gruppen*, Annals of Mathematics, **34** (1933), pp. 147–169.
12. W. J. Hall, R. A. Wijsman, and J. K. Ghosh, *The relationship between sufficiency and invariance with application in sequential analysis*, Ann. Math. Statist. **36**, pp. 575–614 (1965).
13. P. R. Halmos, *Finite Dimensional Vector Space*, D. Van Nostrand Company, Inc. Princeton, 1958.
14. J. Kiefer, *On the nonrandomized optimality and randomized nonoptimality of symmetrical designs*, Ann. Math. Statist. **29**, pp. 675–699 (1958).
15. U. Koehn, *Global cross sections and the densities of maximal invariants*, Ann. Math. Statist. **41**, pp. 2045–2056 (1970).
16. E. L. Lehmann, *Testing of Statistical Hypotheseses*, John Wiley, N.Y., 1959.
17. E. L. Lehmann, *Optimum invariant tests*, Ann. Math. Statist. **30**, pp. 881–884 (1959a).
18. L. Loomis, *An Introduction to Abstract Hermonic Analysis*, D. Van Nostrand Company, Inc., Princeton, 1948.
19. L. Nachbin, *The Haar Integral.*, D. Van Nostrand Company, Inc., Princeton, 1965.
20. R. Schwartz, *Locally mimimax tests*, Ann. Math. Statist., **38**, pp 340–359, 1967.
21. C. Stein, *Multivariate Analysis-I*, Technical Report 42, Department of Statistics, Standford University, 1956.
22. R. A. Wijsman, *Cross-Section of orbits and their application to densities of maximal invariants*, Proc. Fifth. Berk. Symp. Math. Statist. Prob. 1. pp. 389–400 (1967) University of California, Berkeley.
23. R. A. Wijsman, *General proof of termination with probability one of invariant sequential probability ratio test based on multivariate observation*, Ann. Math. Statist. **38**, pp. 8–24 (1969).
24. R. A. Wijsman, *Distribution of maximal invariants, using global and local cross-sections*, Preprint, Univ. of Illinois at Urbana-Champaign, 1978.
25. R. A. Wijsman, *Proper action in steps, with application to density ratios of maximal invariants*, Ann. Statist. 13, pp. 395–402 (1985).
26. R. A. Wijsman, *Global cross sections as a tool for factorization of measures and distribution of maximal invariants*, Sankhya A, 48, pp. 1–42 (1986).

Chapter 3

EQUIVARIANT ESTIMATION IN CURVED MODELS

Let (χ, A) be a measure space and Ω be the set of points θ. Denote by $\{P_\theta, \theta \in \Omega\}$, the set of probability distributions on (χ, A). Let G be a group of transformations operating from the left on χ such that $g \in G$, $g : \chi \to \chi$ is one to one onto (bijective). Let $\bar{G}$ be the corresponding group of induced transformations $\bar{g}$ on Ω.

Assume

(a) For $\theta_i \in \Omega$, $i = 1, 2; \theta_1 \neq \theta_2, P_{\theta_1} \neq P_{\theta_2}$,
(b) $P_\theta(A) = P_{\bar{g}\theta}(gA)$, $A \in A, g \in G, \bar{g} \in \bar{G}$.

Let $\lambda(\theta)$ be a maximal invariant on Ω under $\bar{G}$, and let

$$\Omega^* = \{\theta | \theta \in \Omega, \text{with } \lambda(\theta) = \lambda_0\} \tag{3.1}$$

where λ_0 is known. We assume χ to be the space of minimal sufficient statistic for θ.

Definition 3.1. (Equivariant Estimator) A point estimator $\hat{\theta}(X), X \in \chi$, a mapping from $\chi \to \Omega$, is equivariant if $\hat{\theta}(gX) = g\hat{\theta}(X), g \in G$.

Note 1. For notational convenience we are taking G to the group of transformations on

$$\hat{\theta}(X).$$

Note 2. The unique maximum likelihood estimator is equivariant.

Let $T(X), X \in \chi$, be a maximal invariant on χ under G. Since the distribution of $T(X)$ depends on $\theta \in \Omega$ only through $\lambda(\theta)$, given $\lambda(\theta) = \lambda_0 T(X)$ is an ancillary statistic. An ancillary statistic is defined to be a part of the minimal sufficient statistic X whose marginal distribution is parameter free. In this chapter we consider the approach of finding the best equivariant estimator in models admitting an ancillary statistic. Such models are assumed to be generated as an orbit under the induced group $\bar{G}$ on Ω. The ancillary statistic is realized as the maximal invariant on χ under G. A model which admits an ancillary statistic is referred to as a curved model. Models of this nature are not uncommon in statistics. Fisher (1925) considered the problem of estimating the mean of a Normal population with variance σ^2 when the coefficient of variation is known. The motivation behind this was based on the empirically observed fact that a standard deviation σ often becomes large proportionally to a corresponding mean μ so that the coefficient of variation σ/μ remains constant. This fact is often present in mutually correlated multivariate data. In the case of multivariate data, no well-accepted measure of variation between a mean vector μ and a covariance matrix Σ is available. Kariya, Giri and Perron(1988) suggested the following multivariate version of the coefficent of variation. (i) $\lambda = \mu'\Sigma^{-1}\mu$ (ii) $\nu = \Sigma^{-1/2}\mu$ where $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$ with $\Sigma^{1/2}$ as the unique lower triangular matrix with positive diagonal elements.

In recent years Cox and Hinkley (1977), Efron (1978), Amari (1982 a, b) among other have reconsidered the problem of estimating μ when the coefficent of variation is known in the context of curved models.

3.1. Best Equivariant Estimation of μ with λ Known

Let $X_1, \ldots, X_n$ $(n > p)$ be independently and identically distributed N_p (μ, Σ). We want to estimate μ under the loss function

$$L(\mu, d) = (d - \mu)'\Sigma^{-1}(d - \mu) \tag{3.2}$$

when $\lambda = \mu'\Sigma^{-1}\mu = \lambda_0$ (known). Let

$$n\bar{X} = \sum_{i=1}^{n} X_i, S = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'.$$

Then $(\bar{X}, S)$ is minimal sufficient for (μ, Σ), $\sqrt{n}\bar{X}$ is distributed independently of S as $N_p(\sqrt{n}\mu, \Sigma)$ and S is distributed as Wishart $Wp(n-1, \Sigma)$ with $n-1$ degrees of freedom and parameter Σ. Under the loss function (3.2) this problem remains invariant under the group of transformations $G_l(p)$ of $p \times p$ nonsingular matrices g transforming $(\bar{X}, S) \to (g\bar{X}, gSg')$. The corresponding group $\bar{G}$ of induced transformations $\bar{g}$ on the parametric space Ω transforms $\theta = (\mu, \Sigma) \to \bar{g}\theta = (g\mu, g\Sigma g')$. A maximal invariant invariant on the space of $(\sqrt{n}\bar{X}, S)$ is

$$T^2 = n(n-1)\bar{X}'S^{-1}\bar{X} \tag{3.3}$$

and the corresponding maximal invariant on the parametric space Ω is $\lambda = \mu'\Sigma^{-1}\mu$. Since the distribution of T^2 depends on the parameters only through λ, given $\lambda = \lambda_0\, T^2$ is an ancillary statistic. Since the loss function $L(\mu, d)$ is invariant under the group $G_l(p)$ of transformations g transforming $(\bar{X}, S) \to (g\bar{X}, gSg')$ and $(\mu, \Sigma) \to (g\mu, g\Sigma g')$ we get for any equivariant estimator $\bar{\mu}$ of $\mu(= \bar{\mu}(X))$

$$\begin{aligned} R(\theta, \bar{\mu}) &= E_\theta(\bar{\mu} - \mu)'\Sigma^{-1}(\bar{\mu} - \mu)) \\ &= E_\theta(g\bar{\mu} - g\mu)'(g\Sigma g')^{-1}(g\bar{\mu} - g\mu)) \\ &= E_\theta(\bar{\mu}(gX) - g\mu)'(g\Sigma g')^{-1}(\bar{\mu}(gX) - g\mu) \\ &= E\bar{g}_\theta(\bar{\mu}(X) - g\mu)'(g\Sigma g')^{-1}(\bar{\mu}(gX) - g\mu) \\ &= R(\bar{g}\theta, \bar{\mu}), \end{aligned} \tag{3.4}$$

for $g \in G_l(p), \bar{g}$ is the induced transformation on Σ corresponding to g on χ. Since $\bar{G}_l(p)$ acts transitively on Σ we conclude from (3.4) that the risk $R(\theta, \hat{\mu})$ of any equivariant estimator $\hat{\mu}$ is a constant for all $\theta \in \Omega$.

Taking $\lambda_0 = 1$, which we can do without any loss of generality and using the fact that $R(\theta, \hat{\mu})$ is a constant we can choose $\mu = e = (1, 0, \ldots, 0)', \Sigma = I$.

To find the best equivariant estimator which minimizes $R(\theta, \bar{\mu})$ among all equivariant estimators $\bar{\mu}$ satisfying

$$\bar{\mu}(g\bar{X}, gSg') = g\bar{\mu}(\bar{X}, S)$$

we need to characterize $\bar{\mu}$. Let $G_T(p)$ be the subgroup of $G_l(p($ containing all $p \times p$ lower triangular matrices with positive diagonal elements. Since S is positive definite with probability one we can write, $S = WW', W \in G_T(p)$. Let

$$V = W^{-1}Y, Q = \frac{V}{\| V \|}$$

where $Y = \sqrt{n}\bar{X}, \| V \|^2 = V'V = T^2/(n-1)$.

Theorem 3.1. *If $\bar{\mu}$ is an equivariant estimator of μ under $G_l(p)$ then*

$$\bar{\mu}(Y, S) = k(U)WQ \tag{3.5}$$

where k is a measurable of $U = T^2/(n-1)$.

Proof. Since $\bar{\mu}$ is equivariant under $G_l(p)$ we get for $g \in G_l(p)$

$$g\bar{\mu}(Y, S) = \bar{\mu}(gY, gSg') . \tag{3.6}$$

Replace Y by $W^{-1}Y$, g by W and S by I in (3.6) to obtain

$$\bar{\mu}(Y, S) = W\bar{\mu}(V, I) . \tag{3.7}$$

Let O be an $p \times p$ orthogonal matrix with $Q = V/ \parallel V \parallel$ as its first column. Then

$$\bar{\mu}(V, I) = \bar{\mu}(OO'V, OO') = O\bar{\mu}(\sqrt{U}e, I) .$$

Since the columns of O except the first one are arbitrary as far as they are orthogonal to Q it is easy to claim that the components of $\bar{\mu}(\sqrt{U}e, I)$ except the first component $\bar{\mu}_1(\sqrt{U}\ e, I)$ are zero. Hence

$$\bar{\mu}(V, I) = Q\bar{\mu}_1(\sqrt{U}\ e, I) . \qquad \square$$

The following theorem gives the best equivariant estimator (*BEE*) of μ.

Theorem 3.2. *Under the loss function* (3.2) *the unique BEE of μ is $\bar{\mu} = \hat{k}$ $(U) WQ$ where*

$$\hat{k}(U) = E(Q'W'e|U)/E(Q'W'WQ|U) . \tag{3.8}$$

Proof. By Theorem 3.1 the risk function of an equivariant estimator $\bar{\mu}$, given $\lambda_0 = 1$, is

$$\begin{aligned} R(\theta, \bar{\mu}) &= R((e, I), \bar{\mu}) \\ &= E(k(U)WQ - e)'(k(U)WQ - e) . \end{aligned}$$

Using the fact that U is ancillary, a unique *BEE* is obtained as $\bar{\mu} = \hat{k}(u)WQ$ where $\hat{k}(u)$ minimizes the conditional risk given U

$$E[((k(U)WQ - e)(k(U)WQ - e)')|U] .$$

This implies $\hat{k}(U)$ is given by (3.8). $\square$

3.1.1. *Maximum likelihood estimators*

The maximum likelihood estimators $\hat{\mu}$ and $\hat{\Sigma}$ of μ and Σ respectively under the restriction $\lambda_0 = 1$ are obtained by maximizing

$$-\frac{n}{2}\log(\det \Sigma) - \frac{1}{2} \operatorname{tr} S\Sigma^{-1} - \frac{n}{2}(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu) - \frac{1}{2}\gamma(\mu'\Sigma^{-1}\mu - 1) \quad (3.9)$$

with respect to μ, Σ where γ is the Lagrange multiplier. Maximizing (3.9) we obtain

$$\hat{\mu} = \frac{n\bar{X}}{n+\gamma}, \; \hat{\Sigma} = \frac{1}{n}S + \gamma\bar{X}\bar{X}'(\gamma + n)^{-1}. \quad (3.10)$$

Using the condition $\hat{\mu}'\hat{\Sigma}^{-1}\hat{\mu} = 1$ we get

$$\begin{aligned} \hat{\mu} &= \left(\frac{U - \sqrt{U(4+5U)}}{2U}\right)\bar{X}, \\ \hat{\Sigma} &= \frac{1}{n}S + \frac{U + \sqrt{U(4+5U)}}{2U}\bar{X}\bar{X}'. \end{aligned} \quad (3.11)$$

The maximum likelihood estimator (mle) $\hat{\mu}$ is clearly equivariant and hence it is dominated by $\bar{\mu}$ of Theorem 3.2 for any p. In the univariate case the mle is

$$\frac{1}{2}\bar{X} - \left[\frac{1}{n}S + \frac{5}{4}\bar{X}^2\right]^{1/2} \quad (3.12)$$

Thus (3.12) is the natural extension of (3.11). Hinkley (1977) investigated some properties of this model associated with the Fisher information. Amari (1982 a, b) proposed through a geometric approach what he called the dual mle which is also equivariant.

3.2. A Particular Case

$$\Sigma = \left(\frac{\mu'\mu}{C^2}\right)I.$$

As a particular case of λ constant we consider now the case $\Sigma = \left(\frac{\mu'\mu}{C^2}\right)I$ when C^2 is known. Let $Y = \sqrt{n}\bar{X}, W = \operatorname{tr} S$. Then (Y, W) is a sufficient statistic for this problem, and $\frac{C^2W}{\mu'\mu}$ is distributed independently of Y as $\chi^2_{(n-1)p}$. We are interested here to estimate μ with respect to the loss function

$$L(\mu, d) = \frac{(d-\mu)'(d-\mu)}{\mu'\mu}. \quad (3.13)$$

This problem remains invariant with respect to the group $G = R_+ \times O(p), R_+$ being the multiplicative group of positive reals and $O(p)$ being the multiplicative group of $p \times p$ orthogonal matrices transforming

$$x_i \to b\Gamma x_i\,,\ i = 1\,, \dots\,, n$$
$$d \to b\Gamma d\,,$$
$$(\mu, \left(\frac{\mu'\mu}{C^2}\right) I) \to (b\Gamma\mu, b^2 \left(\frac{\mu'\mu}{C^2}\right) I)\,,$$

where $(b, \Gamma) \in G$ with $b \in R_+, \Gamma \in O(p)$. The transformation induced by G on (Y, W) is given by

$$(Y, W) \to (b\Gamma Y, b^2 W)\,.$$

The theorem below gives the representation of an equivariant estimator under G.

Theorem 3.3. *An estimator $d(Y, W)$ is equivariant if and only if there exists a measurable function $h : R_+ \to R$ such that*

$$d(Y, W) = h\left(\frac{Y'Y}{W}\right) Y$$

for all $(Y, W) \in R^p \times R_+$.

Proof. If h is a measurable function from $R_+ \to R$ and $d(Y, W) = h\left(\frac{Y'Y}{W}\right) Y$ then clearly d is equivariant under G. Conversely if d is equivariant under G, then

$$d(Y, W) = b\Gamma d\left(\frac{\Gamma' Y}{b}, \frac{W}{b^2}\right) \tag{3.14}$$

for all $\Gamma \in O(p), Y \in R^p, b > 0, W > 0$. We may assume without any loss of generality that $Y'Y > 0$.

Let Y, W be fixed and

$$d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

where d_1 is the first component of the p-vector d. Let $A \in O(p)$ be a fixed $p \times p$ matrix such that

$$Y'A = (\ \| Y \|, 0, \dots, 0)\,.$$

Let A be partitioned as $A = (A_1, A_2)$ where $A_1 = \|Y\|^{-1}Y$. Now choose $\Gamma = (A_1, A_2B)$ with $B \in O(p-1)$ and $b = \|Y\|$. From (3.14) we get

$$d(Y,W) = d_1\left((1,0\ldots0).\frac{W}{Y'Y}\right)Y\,.$$
$$+\|Y\|A_2Bd_2\left((1,0...0),\frac{W}{Y'Y}\right). \qquad (3.15)$$

Since (3.15) holds for any choice of $B \in O(p-1)$ we must have then

$$d(Y,W) = d_1\left((1,0\ldots0).\frac{W}{Y'Y}\right)Y\,. \qquad \square$$

It may be easily verified that a maximal invariant under G in the space of sufficient statistic $V = W^{-1}Y'Y$ and a corresponding maximal invariant in the parametric space is

$$\mu'\left(\frac{(\mu'\mu)I}{C^2}\right)^{-1}\mu = C^2\,.$$

As the group acts transitively on the parametric space the risk function (using (3.13))

$$R(\mu,d) = E_\mu(L(\mu,d))$$

of any equivariant estimator d is constant. Hence we can take $\mu = \mu_0 = (C,0\ldots0)'$. Thus the risk of any equivariant estimator d can be written as

$$\begin{aligned} R(\mu_0,d) &= E\mu_0\left(L\left(\mu_0,h\left(\frac{Y'Y}{W}\right)Y\right)\right) \\ &= E\mu_0\left(E\left(L\left(\mu_0,h\left(\frac{Y'Y}{W}\right)Y\right)\right)V=v\right). \end{aligned}$$

To find a BEE we need to find the function h_0 satisfying

$$E\mu_0(L(\mu_0,h_0(V)Y|V=v) \le E\mu_0(L(\mu_0,h(v)y)|V=v)$$

for all $h: R_+ \to R$ measurable functions and for all values v of V. Since

$$\begin{aligned} &E\mu_0(L(\mu_0,h(v)Y)|V=v) \\ &\quad = h^2(v)E\mu_0(Y'Y|V=v) - 2h(v)E\mu_0(Y_1|V=v)+1 \end{aligned}$$

where $Y = (Y_1,\ldots,Y_p)'$, we get

$$h_0(v) = \frac{E\mu_0(Y_1|V=v)}{E\mu_0(Y'Y|V=v)}\,. \qquad (3.16)$$

Theorem 3.4. *The BEE* $d_0(x_1, \ldots, X_n; C) = d_0(Y, W)$ *is given by*

$$d_0(Y,W) = \frac{nC^2}{q}\left[\frac{\sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}np+i+1\right)}{\Gamma\left(\frac{1}{2}p+i+1\right)i!}\left(\frac{nC^2}{2}\right)^i}{\sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}np+j+1\right)}{\Gamma\left(\frac{1}{2}p+j\right)j!}\left(\frac{nC^t}{2}\right)^j}\right]\bar{X} \tag{3.17}$$

where $t = v(1+v)^{-1}$.

Proof. The joint probability density function of Y and W under the assumption $\mu = \mu_0$ is

$$f_{(Y,W)}^{y,w} = \begin{cases} \dfrac{\exp\{-(C^2/2)(y'y - 2\sqrt{n}y_1 + n + w)\}w^{\frac{1}{2}(n-1)p-1}}{2^{np/2}(C^2)^{-np/2}(\Gamma(\frac{1}{2}))^p\Gamma((n-1)p/2)} & \text{, if } w > 0 \\ 0, \text{ otherwise}. \end{cases}$$

Hence the joint probability density function of Y, V is given by

$$f_{Y,V}(y,v) = \begin{cases} \dfrac{\exp\{-(C^2/2[((1+v)/v)y'y + n]\}}{2^{np/2}(C^2)^{-np/2}(\Gamma(\frac{1}{2}))^p\Gamma((n-1)p/2)} \\ \quad \times \exp\{\sqrt{n}C^2 y_1\}[(y'y)^{(n-1)p/2}v^{-(n-1)p/2-1} & \text{, if } v > 0 \\ 0, \text{ otherwise}. \end{cases}$$

Using (3.16) we get (3.17).

Corollary. If $m = p(n-1)/2$ is an integer, the BEE is given by

$$d_0(Y,W) = \frac{nC^2}{2}h_0(v) = g(t)\bar{X}\,, \tag{3.18}$$

with $g(t) = \mu(t)/w(t)$ where

$$u(t) = \sum_{i=0}^{m+1} \frac{\binom{i}{m+1}\binom{m+1}{i}}{\Gamma\left(\frac{1}{2}p+i\right)}\left(\frac{nC^2}{2}t\right)^i,$$

$$w(t) = \sum_{j=0}^{m+1} \frac{\binom{m+1}{j}\left(\frac{nC^2}{2}\right)^j t^{j+1}}{\Gamma\left(\frac{1}{2}p+i\right)}.$$

Proof. Let Y_k be distributed as χ_k^2. Then

$$E(Y_k^\alpha) = 2^\alpha \frac{\Gamma(k/2+\alpha)}{\Gamma(k/2)} \text{ if } \alpha > \frac{-k}{2}.$$

Hence with m as integer

$$\begin{aligned} h_0(v) &= \sqrt{n}C^2 \frac{\sum_{i=0}^{\infty} E(Y_{p+2i+2}^m)\exp\{-nC^2t/2\}(\frac{nC^2t}{2})^i \frac{1}{i!}}{\sum_{j=0}^{\infty} E(Y_{p+2j}^{m+1})\exp\{-nC^2t/2\}(\frac{nC^2t}{2})^j \frac{1}{j!}} \\ &= \sqrt{n}C^2 E(V_1^m)/E(V_2^{m+1}), \end{aligned} \tag{3.19}$$

where V_1 is distributed as noncentral $\chi_{p+2}^2(nC^2t)$, and V_2 is distributed as noncentral $\chi_p^2(nC^2t)$. Hence letting $V = \chi_v^2(\delta^2)$ and taking r as an integer we get

$$E(V^r) = 2^r \sum_{k=0}^{r} \frac{\Gamma(v/2+r)}{\Gamma(v/2+k)} \binom{r}{k} \left(\frac{\sigma^2}{r}\right)^k . \tag{3.20}$$

From (3.19) and (3.20) we get (3.18). □

Note 1. $g(t)$ is a continuous function of t, and

$$\begin{aligned} &\lim_{t\to 0^+} g(t) = nC^2/p\,, \\ &\lim_{t\to 1} g(t) = g(1) < 1\,, \end{aligned}$$

and $g(t) > 0$ for all t. Thus when $Y'Y$ is large the BEE is less than $\bar{X}$.

Note 2. We can also write $d_0 = (1 - \frac{\tau(v)}{v})\bar{X}$ where $\tau(v)/v = 1 - g(t)$. This form is very popular in the literature. Perron and Giri (1990) have shown that $g(t)$ is a strictly decreasing function of t and $\tau(v)$ is strictly increasing in v. The result that $g(t)$ is strictly decreasing in t tells what one may intuitively do if he has an idea of the true value of C and observe many large values concentrated. Normally one is suspicious of their effects on the sample mean and they have the tendency to shrink the sample mean towards the origin. That is what our estimator does. The result that $\tau(v)$ is strictly increasing in v relates the BBE of the mean for C known with the class of minimax estimators of the mean for C unknown. Efron and Morris (1977) have shown that a necessary condition for an equivariant estimator of the form $g(t)\bar{X}$ to be minimax is $g(t) \to 1$ as $t \to 1$, so our estimator fails to be minimax if we do not know the value of C. On the other hand Efron and Morris (1977) have shown that an estimator of the form $d = (1 - \frac{\tau(v)}{v})\bar{X}$ is minimax if (i) τ is an increasing function, (ii) $0 \leq \tau(v) \leq (p-2)/(n-1) + 2$ for all $v \in (0, \infty)$. Thus our estimator satisfies (i) but fails to satisfy (ii). So a truncated version of our estimator could be a compromise solution between the best, when one knows the values of C and the worst, one can do by using an incorrect value of C.

3.2.1. *An application*

The following interesting application of this model is given by Kent, Briden, and Mardia (1983). The natural remanent magnetization (NRM) in rocks is known to have, in general, originated in one or more relatively short time intervals during rock-forming or metamorphic events during which NRM is frozen in by falling temperature, grain growth, etc. The NRM acquired during each such event is a single vector magnetization parallel to the then-prevailing geometric field and is called a component of NRM. By thermal, alternating fields or chemical demagnetization in stages these components can be identified. Resistance to these treatments is known as "stability of remanence". At each stage of the demagnetization treatment one measures the remanent magnetization as a vector in 3-dimensional space. These observations are represented by vectors $X_1, \ldots, X_n$ in R^3. They considered the model given by $X_i = \alpha_i + \beta_i + e_i$ where α_i denotes the true magnetization at the ith step, β_i represents the model error, and e_i represents the measurement error. They assumed that β_i and e_i are independent, β_i is distributed as $N_3(0, \tau^2(\alpha_i)I)$, and e_i is distributed as $N_3(0, \sigma^2(\alpha_i)I)$. The α_i are assumed to possess some specific structures, like collinearity etc., which one attemps to determine. Sometimes the magnitude of model error is harder to ascertain and one reasonably assumes $\tau^2(\alpha) = 0$. In practice $\sigma^2(\alpha)$ is allowed to depend on α_i and plausible model for $\sigma^2(\alpha)$ which fits many data reasonably well is $\sigma^2(\alpha) = a(\alpha'\alpha) + b$ with $a > 0, b > 0$. When $\alpha'\alpha$ large, b is essentially 0 and a is unknown.

3.2.2. *Maximum likelihood estimator*

The likelihood function of $x_1, \ldots x_n$ with C known is given by

$$L(x_1, \ldots, x_n|\mu)$$
$$= \left(\frac{2n}{C^2}\right)^{-np/2} (\mu'\mu)^{-np/2} \exp\left\{-\frac{C^2}{2\mu'\mu}(w + y'y - 2\sqrt{n}y'\mu + n\mu'\mu)\right\}.$$

Thus the mle $\hat{\mu}$ of μ (if it exists) is given by

$$[(np/C^2)(\hat{\mu}'\hat{\mu}) - w - y'y + 2\sqrt{n}y'\hat{\mu}]\hat{\mu} = \sqrt{n}\hat{\mu}'\hat{\mu}y\,. \tag{3.21}$$

If the Eq. (3.21) in $\hat{\mu}$ has a solution it must be colinear with y. Let $\hat{\mu} = ky$ be a solution. From (3.21) we obtain

$$k[(np/C^2)y'y)k^2 + \sqrt{n}y'yk - (y'y + w)] = 0\,.$$

Two nonzero solutions of k are

$$k_1 = \frac{-1 - (1 + \frac{4p}{C^2}(\frac{1+v}{v}))^{1/2}}{2\sqrt{n}p/C^2}, \quad k_2 = \frac{-1 + (1 + \frac{4p}{C^2}(\frac{1+v}{v}))^{1/2}}{2\sqrt{n}p/C^2}\,.$$

To find the value of k which maximizes the likelihood we compute the matrix of mixed derivatives

$$\frac{\partial^2(-\log L)}{\partial\mu'\partial\mu}\Big|_{\mu=ky} = \frac{C^2}{k^4(y'y)^2}\left[\sqrt{n}k(y'y)I + \frac{2np}{C^2}k^2yy'\right]$$

and assert that this matrix should be positive definite. The characteristic roots of this matrix are given by

$$\lambda_1 = \frac{\sqrt{n}C^2}{k^3y'y}, \quad \lambda_2 = \frac{\sqrt{n}C^2 + 2npk}{k^2y'y}\,.$$

If $k = k_1$, then $\lambda_1 < 0$ and $\lambda_2 < 0$. But if $k = k_2$, then $\lambda_1 > 0$, $\lambda_2 > 0$. Hence the mle $\hat{\mu} = d_1(x_1, \ldots, x_n, C)$ is given by (corresponding to positive characteristic roots)

$$d_1(x_1, \ldots x_n, C) = \left[\frac{(1 + 4p/C^2t)^{1/2} - 1}{2p}\right] C^2\bar{X}\,. \tag{3.22}$$

Since the maximum likelihood estimator is equivariant and it differs from the BEE d_0 the mle d_1 is inadmissible. The risk function of d_0 depends on C. Perron and Giri (1990) computed the relative efficiency of d_0 when compared with d_1, the James-Stein estimator (d_2), the positive part of the James-Stein estimator (d_3), and the sample mean $\bar{X}(d_4)$ for different values of C, n, and p. They have concluded that when the sample size n increases for a given p and C the relative efficiency of d_0 when compared with $d_i, i = 1, .., 4$ does not change significantly. This phenomenon changes markedly when C varies. When C is small, d_0 is markedly superior to others. On the other hand, when C is large all five estimators are more or less similar. These conclusions are not exact as the risk of d_0, d_1 are evaluated by simulation. Nevertheless, it give us sufficient indication that for small values of C the use of BEE is clearly advantageous.

3.3. Best Equivariant Estimation in Curved Covariance Models

Let $X_1, \ldots, X_n\, (n > p > 2)$ be independently and identically distributed p-dimensional normal vectors with mean vector μ and positive definite convariance matrix Σ.

Let Σ and $S = \sum_{\alpha=1}^{n} (X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ be partitioned as

$$\Sigma = \begin{array}{c} \begin{pmatrix} 1 & p-1 \\ \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{1}_{p-1} \end{array}, S = \begin{pmatrix} 1 & p-1 \\ S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}^{1}_{p-1} .$$

We are interested to find the BEE of $\beta = \Sigma_{22}^{-1}\Sigma_{21}$ on the basis of n sample observations $x_1, \ldots, x_n$ when one knows the value of the multiple correlation coefficent $\rho^2 = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. If the value of ρ^2 is significant, one would naturally be interested to estimate β for the prediction purpose and also to estimate Σ_{22} to ascertain the variability of the prediction variables.

Let $G_l(p)$ be the full linear group of $p \times p$ nonsingular matrices and let $O(p)$ be the multiplicative group of $p \times p$ orthogonal matrices. Let H_1 be the subgroup of $G_l(p)$ defined by

$$H_1 = \{h \in G_l(p) : h = \begin{pmatrix} l & p-1 \\ h_{11} & 0 \\ 0 & h_{22} \end{pmatrix}^{1}_{p-1}$$

and let H_2 be the additive group in R^p. Define $G = H_1 \oplus H_2$ the direct sum of H_1 and H_2. The transformation $g = (h_1, h_2) \in G$ transforms H_2. The transformation $g = (h_1, h_2) \in G$ transforms

$$\begin{aligned} X_i &\to h_1 X_i + h_2, \quad i = 1, \ldots, n, \\ (\mu, \Sigma) &\to (h_1\mu + h_2, h_1 \Sigma h_1'). \end{aligned}$$

The corresponding transformation on the sufficient statistic is given by $(\bar{X}, S) \to (h_1\bar{X} + h_2, h_1 S h_1')$. A maximal invariant in the space of $(\bar{X}, S)$ under G is

$$R^2 = S_{11}^{-1} S_{12} S_{22}^{-1} S_{21}$$

and a corresponding maximal invariant in the parametric space of (μ, Σ) is ρ^2.

3.3.1. *Characterization of equivariant estimators of* Σ

Let S_p be the space of all $p\times p$ positive definite matrices and let $G_T^+(p)$ be the group of $p\times p$ lower triangular matrices with positive diagonal elements. An equivariant estimator $d(\bar{X}, S)$ of Σ with respect to the group of transformations G is a measurable function $d(\bar{X}, S)$ on $S_p\times R^p$ to S_p satisfying

$$d(h\bar{X}+\xi, hSh') = hd(\bar{X}, S)h'$$

for all $S \in S_p, h \in H_1$, and $\bar{X}, \xi \in R^p$. From this definition it is easy to remark that if d is equivariant with respect to G then $d(\bar{X}, S) = d(0, S)$ for all $\bar{X} \in R^p, S \in S_p$. Thus without any loss of generality we can assume that d is a mapping from S_p to S_p. Furthermore, if u is a function mapping S_p into another space Y (say) then d^* is an equivariant estimator of $u(\Sigma)$ if and only if $d^* = u \cdot d$ for some equivariant estimator d of Σ.

Let

$$\Theta_\rho = \{(\mu, \Sigma) : \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \rho^2\}$$

and let $\bar{G}$ be the group of induced transformations corresponding to G on O_p.

Lemma 3.1. *$\bar{G}$ acts transitively on Θ_p.*

Proof. It is sufficient to show that there exists *a* $g = (h, \xi) \in G$ with $h \in H_1, \xi \in R^p$ such that

$$(h\mu+\xi,\ h\Sigma h') = \left(0, \begin{array}{cc} \begin{pmatrix} 1 & 1 & p-2 \\ 1 & p & 0 \\ p & 1 & 0 \\ 0 & 0 & I \end{pmatrix} & \begin{array}{c} 1 \\ 1 \\ p-2 \end{array} \end{array}\right). \tag{3.23}$$

If $\rho = 0$, i.e., $\Sigma_{12} = 0$, we take $h_{11} = \Sigma_{11}^{-1/2}$, $\xi = -h\mu$ to obtain (3.23). If $\rho \neq 0$, choose $h_{11} = \Sigma_{11}^{-1/2}$, $h_{22} = \Gamma\ \Sigma_{22}^{-1/2}$, where Γ is a $(p-1)\times(p-1)$ orthogonal matrix such that

$$\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Gamma = (\rho,\ 0,\ \ldots,\ 0),$$

and $\xi = -h\mu$ to obtain (3.23).

The following theorem gives a characterization of the equivariant estimator $d(S)$ of Σ.

Theorem 3.5. *An estimator d of Σ is equivariant if and only if it admits the decomposition*

$$d(s) = \begin{pmatrix} a_{11}(R) & a_{12}(R)R^{-1}S_{12} \\ a_{12}(R)R^{-1}S_{12} & R^{-2}a_{22}(R)S_{21}S_{11}^{-1}S_{12} + C(R)(S_{22} - S_{21}S_{11}^{-1}S_{12}) \end{pmatrix} \tag{3.24}$$

where $C(R) > 0$ and

$$A(R) = \begin{pmatrix} a_{11}(R) & a_{12}(R) \\ a_{21}(R) & a_{22}(R) \end{pmatrix}$$

is a 2×2 positive definite matrix. Furthermore

$$d_{11}^{-1}d_{12}d_{22}^{-1}d_{21} = \rho^2 = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

if and only if $a_{11}^{-1}a_{12}a_{22}^{-1}a_{21} = \rho^2$.

Note. The d_{ij} are submatrices of d as partitioned in (3.24) and $a_{ij} = a_{ij}(R)$.

Proof. The sufficiency part of the proof is computational. It consists in verifying $d(hsh') = hd(S)h'$ for all $h \in H, S \in Sp$ and $d_{11}^{-1}d_{12}d_{22}^{-1}d_{21} = a_{11}^{-1}a_{12}a_{22}^{-1}a_{21}$. It can be obtained in a straightforward way from the computations presented in the necessary part.

To prove the necessary part we observe that

$$P = \begin{pmatrix} 1 & R \\ R & 1 \end{pmatrix}, \ R > 0,$$

and d satisfies

$$d\begin{pmatrix} P & 0 \\ 0 & I_{p-2} \end{pmatrix} = \begin{pmatrix} I_2 & 0 \\ 0 & \Gamma \end{pmatrix} d\begin{pmatrix} P & 0 \\ 0 & I_{p-2} \end{pmatrix}\begin{pmatrix} I_2 & 0 \\ 0 & \Gamma' \end{pmatrix}$$

for all $\Gamma \in O(p-2)$. This implies that

$$d\begin{pmatrix} P & 0 \\ 0 & I_{p-2} \end{pmatrix} = \begin{pmatrix} A(R) & 0 \\ 0 & C(R)I_{p-2} \end{pmatrix}$$

with $C(R) > 0$.

In general, S has a unique decomposition of the form

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix}\begin{pmatrix} 1 & U' \\ U & I_{p-1} \end{pmatrix}\begin{pmatrix} T_1' & 0 \\ 0 & T_2' \end{pmatrix}$$

with $T_1 \in G_T^+(1), T_2 \in G_T^+(p-1)$, and $U = T_2^{-1} S_{21} T_1^{-1}$. Without any loss of generality we may assume that $U \neq 0$. Corresponding to U there exists a $B \in O(p-1)$ such that $U'B = (R,\ 0,\ \ldots 0)$ with $R = \| U \| = (S_{11}^{-1} S_{12} S_{22}^{-1} S_{21})^{\frac{1}{2}} > 0$. For $p > 2$, B is not uniquely determined but its first column is $R^{-1}U$. Using such a B we have the decomposition

$$\begin{pmatrix} 1 & U' \\ U & I_{p-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} P & 0 \\ 0 & I_{p-2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix}$$

and

$$\begin{aligned} d(S) &= \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} A(R) & 0 \\ 0 & I_{p-2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix} \begin{pmatrix} T_1' & 0 \\ 0 & T_2' \end{pmatrix} \\ &= \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} \begin{pmatrix} a_{11}(R) & R^{-1} a_{12}(R) U' \\ R^{-1} a_{21}(R) U & R^{-2} a_{22}(R) UU' + C(R)(I_{p-1} - UU') \end{pmatrix} \begin{pmatrix} T_1' & 0 \\ 0 & T_2' \end{pmatrix} \\ &= \begin{pmatrix} a_{11}(R) S_{11} & R^{-1} a_{12}(R) S_{12} \\ R^{-1} a_{21}(R) S_{21} & R^{-2} a_{22}(R) S_{21} S_{11}^{-1} S_{12} + C(R)(S_{22} - S_{21} S_{11}^{-1} S_{12}) \end{pmatrix} \end{aligned}$$

which proves the necessary part of the theorem. □

3.3.2. *Characterization of equivariant estimators of β*

The following theorem gives a characterization of the equivariant estimator of β.

Theorem 3.6. *If d^* is an equivariant estimator of β then $d^*(S)$ has the form*

$$d^*(S) = R^{-1} a(R) S_{22}^{-1} S_{21} \,,$$

where $a : R_+ \to R$.

Proof. Define $u : S^p \to R^{p-1}$ by

$$u(\Sigma) = \beta = \Sigma_{22}^{-1} \Sigma_{21} \,.$$

If d^* is equivariant then

$$\begin{aligned} d^*(S) &= (R^{-2} a_{22}(R) S_{21} S_{11}^{-1} S_{12} + C(R)(S_{22} - S_{21} S_{11}^{-1} S_{12}))^{-1} S_{21} R^{-1} a_{21}(R) \\ &= (T_2 (R^{-2} a_{22}(R) UU' + C(R)((I_{p-1} - UU') T_2' S_{21} R^{-1} a_{21}(R) \\ &= R^{-1} (a_{22}(R) + (1 - R^2)(C(R))^{-1} a_{21}(R) S_{22}^{-1} S_{21} \\ &= R^{-1} a(R) S_{22}^{-1} S_{21} \,. \end{aligned}$$

□

The risk function of an equivariant estimator d^* of β is given by

$$\begin{aligned}
R(\beta, d^*) &= E_\Sigma(L(\beta, d^*)) \\
&= E_\Sigma\{S_{11}^{-1}(R^{-1}a(R)S_{12}S_{22}^{-1} - \beta)'S_{22}(R^{-1}a(R)S_{22}^{-1}S_{21} - \beta)\} \\
&= E_\Sigma\{a^2(R) - 2R^{-1}a(R)S_{11}^{-1}S_{12}\beta + S_{11}^{-1}\beta' S_{22}\beta\}\,. \qquad (3.25)
\end{aligned}$$

Theorem. 3.7. The best equivariant estimator of β given ρ, under the loss function L is given by

$$R^{-1}a^*(R)S_{22}^{-1}S_{21} \qquad (3.26)$$

where

$$a^*(R) = r\rho^2 \frac{\sum_{i=0}^{\infty}\Gamma\left(\frac{n+1}{2}+i\right)\Gamma\left(\frac{n-1}{2}+i\right)\left(r^2\rho^2\right)^i/1!\Gamma\left(\frac{p+1}{2}+i\right)}{\sum_{\infty}^{\infty}\Gamma^2\left(\frac{n-1}{2}+m\right)(\rho^2r^2)^m/m!\Gamma\left(\frac{p-1}{2}+m\right)}\,. \qquad (3.27)$$

Proof. From (3.25), the minimum of $R(\beta, d^*)$ is attained when $a(R) = a^*(R) = E_\Sigma(S_{11}^{-1}S_{12}\beta R^{-1}|R)$. Since the problem is invariant and d^* is equivariant we may assume, without any loss of generality, that

$$\Sigma = \Sigma_\rho = \begin{pmatrix} C(\rho) & 0 \\ 0 & I_{p-2} \end{pmatrix}$$

with $C(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. To evaluate $a^*(R)$ we write

$$\begin{aligned}
&S_{22} = TT',\ T \in G_T^+(p-1),\ T = (t_{ij}) \\
&S_{21} = RTW,\ 0 < R < 1,\ W \in R^{p-1}, \\
&S_{11} = W'W\,.
\end{aligned}$$

The joint probability density function of (R, W, T) is (Giri (1977))

$$\begin{aligned}
f_{R,W,T}(r, w, t) =& k^{-1}r^{p-2}(1-r^2)^{(n-p)/2-1}(w'w)^{(n-p)/2\}} \\
&\times\ \exp\{-\frac{1}{2}\sum_{i=2}^{p-1}\sum_{j=1}^{i} t_{ij}^2 \\
&\times \exp\{-\frac{1}{2(1-\rho^2)}(w'w + t_{11}^2 - 2r\rho t_{11}w_1)\}\prod_{i=1}^{p-1}(t_{ii})^{n-i-1}\,. \qquad (3.28)
\end{aligned}$$

A straightforward computation gives (3.26). □

Lemma 3.2. *Let $\beta > 0$, $0 < \gamma < 1, m$ be an integer. Then*

$$\sum_{l=0}^{\infty} \frac{\Gamma(\alpha+m+l)}{\Gamma(\alpha+l)} \frac{\Gamma(\beta+l)}{1!} \gamma^l$$
$$= (1-\gamma)^{-\beta} \sum_{j=0}^{m} \binom{m}{j} \frac{\Gamma(\alpha+m)}{\Gamma(\alpha+j)} \Gamma(\beta+j) \left(\frac{\gamma}{1+\gamma}\right)^j .$$

Proof.

$$\sum_{i=0}^{\infty} \frac{\Gamma(\alpha+m+l)}{\Gamma(\alpha+l)} \frac{\Gamma(\beta+l)}{l!} \gamma^l$$
$$= \frac{d^m}{dt^m} t^{\alpha+m-1} (1-\gamma t)^{-\beta} \Gamma(\beta) \Big|_{t=l}$$
$$= (1-\gamma)^{-\beta} \Gamma(\beta) \frac{d^m}{du^m} (1+u)^{\alpha+m-1} \left(1 - \frac{\gamma}{1-\gamma} u\right)^{-\beta} \Big|_{u=0}$$
$$= (1-\gamma)^{-\beta} \sum_{j=0}^{m} \binom{m}{j} \frac{\Gamma(\alpha+m)}{\Gamma(\alpha+j)} \Gamma(\beta+j) \left(\frac{\gamma}{1-\gamma}\right)^j . \qquad \square$$

Theorem 3.8. *If $m = \frac{1}{2}(n-p)$ is an integer then*

$$a^*(r) = \frac{1}{2}(n-1) r \rho^2$$
$$\times \frac{\sum_{i=0}^{m} \binom{m}{i} \left[\Gamma\left(\frac{n-1}{2}+i\right) / \Gamma\left(\frac{p+1}{2}+i\right)\right] \left(\frac{r^2\rho^2}{1-r^2\rho^2}\right)^i}{\sum_{j=0}^{m} \binom{m}{j} \left[\Gamma\left(\frac{n-1}{2}+j\right) / \Gamma\left(\frac{p-1}{2}+j\right)\right] \left(\frac{r^2\rho^2}{1-r^2\rho^2}\right)^j} .$$

Proof. Follows trivially from Lemma (3.2). $\square$

Note. If ρ^2 is such that terms of order $(rp)^2$ and higher can be neglected, the best equivariant estimator of β is approximately equal to $\rho^2(n-1) \cdot (p-1)^{-1} S_{22}^{-1} S_{21}$. The mle of β is $S_{22}^{-1} S_{21}$.

Exercises

1. Let $X_\alpha = (X_{\alpha 1}, \ldots, X_{\alpha p})', \alpha = 1, \ldots, N(>p)$ be independently and identically distributed p-variate normal random vectors with common mean μ and positive definite common covariance matrix Σ and let

$$\bar{X} = \frac{1}{N}\sum_{\alpha=1}^{N} X_\alpha, \quad S = \sum_{\alpha=1}^{N}(X_\alpha - \bar{X})(X_\alpha - \bar{X})'.$$

show that if $\delta^2 = N\mu'\Sigma^{-1}\mu$ is known, then $N\bar{X}'(S + N\bar{X}\bar{X}')^{-1}\bar{X}$ is an ancillary statistic.

2. Let $X_\alpha = (X_{\alpha 1}, \ldots, X_{\alpha p})', \alpha = 1, \ldots, N(\geqslant p)$ be independently and identically distributed p-variate normal random vectors with mean 0 and positive definite covariance matrix Σ and let

$$\Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}$$

with $\Sigma_{(11)} : 1\times 1$. Given the value of $\rho^2 = \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}/\Sigma_{11}$ find the ancillary statistic.

3. (Basu (1955)). If T is a boundedly complete sufficient statistic for the family of distributions $\{P_\theta, \theta \in \Omega\}$, then any ancillary statistic is independent of T.
4. Find the conditions under which the maximum likelihood estimator is equivariant.

References

1. S. Amari, *Differential geometry of curved exponential families – curvatures and information loss*, Ann. Statist. **10**, 357–385, (1982a).
2. S. Amari, *Geometrical theory of asymptotic ancillary and conditional inference*, Biometrika, **69**, 1–17 (1982b).
3. D. Basu, *On statistics independent of sufficient statistic*, Sankhya **20**, 223–226 (1955).
4. D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Champman and Hill, London, (1977).
5. B. Efron, *The geometry of exponential families*, Ann. Statist., **6**, 262–376, (1978).
6. B. Efron and C. Morris, *Stein's estimation rule and its competitors, An empirical approach*, J. Amer. Statist. Assoc. **68**, 117–130, (1973).
7. B. Efron and C. Morris, *Stein's paradox in statistics*, Sci. Ann., **239**, 119–127, (1977).

8. R. A. Fisher, *Theory of statistical estimation*, Proc. Cambridge Phil. Soc. **22**, 700–725, (1925).
9. N. Giri, *Multivariate Statistical Inference*, Academic Press, N.Y., 1977.
10. D. V. Hinkley, *Conditional inference about a normal mean with known coefficent of variation*, Biometrika, **64**, 105–108, (1977).
11. T. Kariya, N. Giri, and F. Perron, *Equivariant estimator of a mean vector* μ *of* $N_p(\mu, \Sigma)$ *with* $\mu'\Sigma^{-1}\mu = 1$ *or* $\Sigma^{-1/2}\mu = c$ *or* $\Sigma = \sigma^2(\mu'\mu)I$, *J. Multivariate Anal.* **27**, 270–283 (1988).
12. J. Kent, J. Briden, and K. Mardia, *Linear and planer structure in ordered multivariate data as applied to progressive demagnetization of palaemagnetic remanence*, Geophys. J. Roy. Astronom. Soc. **75**, 593–662, (1983).
13. F. Perron and N. Giri, *On the best equivariant estimator of mean of a multivariate normal population*, J. Multivariate Analysis, **32**, 1–16 (1990).
14. F. Perron and N. Giri, *Best equivariant estimation in Curved Covariance Models*, **40**, 46–55 (1992).

Chapter 4

SOME BEST INVARIANT TESTS IN MULTINORMALS

4.0. Introduction

In Chapter 3 we have dealt with some applications of invariance in statistical estimation. We discuss in this chapter various testing problems concerning means of multivariate normal distributions. Testing problems concerning discriminant coefficents, as they are somewhat related to mean problems will also be cousidered. This chapter will also include testing problems concerning multiple correlation cocfficient and a related problem concerning multiple correlation with partial informations. We will be concerned with invariant tests only and we will take a different approach to derive tests for these problems. Rather than deriving the likelihood ratio tests and studying their optimum properties we will look for a group under which the testing problem remain invariant and then find tests based on the maximal invariant under the group. A justification of this approach is as follows: If a testing problem is invariant under a group, the likelihood ratio test, under a mild regularity condition, depends on the observations only through the maximal invariant in the sample space under the group (Lehmann, 1959 p. 252). We find then the optimum invariant test using the above approach, the likelihood ratio test can be no better, since it is an invariant test.

4.1. Tests of Mean Vector

Let $X = (X_1, \ldots, X_p)$ be normally distributed with mean $\xi = E(X) = (\xi_1, \ldots, \xi_p)^1$ and positive definite covariance matrix $\Sigma = E(X - \mu)\ (X - \mu)'$. Its *pdf* is given by

$$f_X(x) = (2\pi)^{-p/2}(\det\Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}(x-\xi)(x-\xi)^{'}\right\}. \tag{4.1}$$

In what follows we denote $\mathfrak{X}$ a p-dimensional linear vector space, and $\mathfrak{X}'$, the dual space of $\mathfrak{X}$. The uniqueness of nonmal distribution follows from the following two facts:

(a) The distribution of X is completely determined by the family of distributions of $\theta' X$, $\theta \in \mathfrak{X}'$.
(b) The mean and the variance of a normal variable completely determines its distribution.

For relevant results of univariate and multivariate normal distribution we refer to Giri (1996). We shall denote a p-variate normal with *pdf* (4.1) as $N_p(\xi, \Sigma)$.

On the basis of observations $x^\alpha = (x_{\alpha 1}, \ldots, x_{\alpha p})^{'}$, $\alpha = 1, \ldots, N$ from $N_p(\xi, \Sigma)$ we are interested in testing the following statistical problems.

Problem 1. To test the null hypothesis H_0: $\xi = 0$ against the alternatives H_1: $\xi \neq 0$ when ξ, Σ are unknown. This is known as Hotelling's T^2 problem.

Problem 2. To test the null hypothesis H_0: $\xi_1 = \cdots = \xi_{p_1} =$ against the alternatives $H_1 : \xi \neq 0$ when ξ, Σ are unknown and $p > p_1$.

Problem 3. To test the null hypothesis $H_0 : \xi_1 = \cdots = \xi_{p_1}$ against the alternatives $H_1 : \xi_1 = \cdots = \xi_{p_1} = 0$ when μ, Σ are unknown and $p_1 + p_2 \leq p$.

Let $\bar{X} = \frac{1}{N}\Sigma_1^N X^\alpha$, $S = \Sigma_1^N (X^\alpha - \bar{X})(X^\alpha - \bar{X})^{'}$. $(\bar{X}, S)$ is sufficient (minimal) for (ξ, Σ) and $\sqrt{N}\bar{X}$ is distributed independently of S as $N_p(\sqrt{n}\xi, \Sigma)$ and S has Wishart distribution $W_p(N-1, \Sigma)$ with $n = N-1$ degrees of freedom and parameter Σ. The *pdf* of S is given by

$$W_p(n, \Sigma) = \begin{cases} K(\det\Sigma)^{-n/2}(\det s)^{\frac{n-p-1}{2}} \exp\{-\frac{1}{2}\mathrm{tr}\Sigma^{-1}s\} \\ \text{if } s \text{ is positive definite,} \\ 0 \text{ otherwise.} \end{cases} \tag{4.2}$$

Problem 1 remains invariant under the general linear group $G_l(\mathrm{p})$ transfering $(\bar{X}, S) \to (g\bar{X}, gSg')$, $g \in G_l(p)$. From Example 2.4.3 with $d_1 = p$

$$\begin{aligned} R &= N\bar{X}'(S + n\bar{X}\bar{X}')^{-1}\bar{X} \\ &= \frac{N\bar{X}'S^{-1}\bar{X}}{1 + N\bar{X}'S^{-1}\bar{X}} \end{aligned} \tag{4.3}$$

is a maximal invariant in the space of $(\bar{X}, S)$ under $G_l(p)$ and $0 \leq R \leq 1$.

The likelihood ratio test of H_0 depends on the statistic (Giri (1996))

$$T^2 = N(N-1)\bar{X}'S^{-1}\bar{X} \tag{4.4}$$

which is known as Hotelling's T^2 statistic. Since R is a one-to-one function of T^2 we will use the test based on R. From Theorem 2.8.1 the *pdf* of R depends only on $\delta = N\xi'\Sigma^{-1}\xi$ and is given by

$$\begin{aligned} f_\delta(r) = & \frac{\Gamma(N/2)}{\Gamma(p/2)\Gamma((N-p)/2)} r^{(p/2)-1}(1-r)^{\frac{1}{2}(N-p)-1} \\ & \times \exp\left\{-\frac{1}{2}\delta\right\}\phi\left(\frac{1}{2}(N-p), \frac{p}{2}; \frac{1}{2}r\delta\right). \end{aligned} \tag{4.5}$$

It may be verified that δ_1 is a maximal invariant under the induced group $\bar{G}_l(p) = G_l(p)$ in the parametric space of (ξ, Σ). Since Σ is positive definite $\delta_1 = 0$ if and only if $\mu = 0$ and $\delta > 0$ for $\mu \neq 0$. For invariant tests with respect to $G_l(p)$ Problem 1 reduces to testing H_0: $\delta = 0$ against the alternatives H_1: $\delta > 0$. From (4.5) the *pdf* of R under H_0 is given by

$$f_0(r_1) = \begin{cases} \dfrac{\Gamma(\frac{N}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{N-p}{2})} r^{\frac{p}{2}-1}(1-r)^{\frac{1}{2}(N-p)-1} \\ \text{if } \ 0 < r < 1, \\ 0 \ \ \text{otherwise}, \end{cases} \tag{4.6}$$

which is a central beta with parameters $(\frac{p}{2}, \frac{N-p}{2})$. From (4.5) the distribution of R has monotone likelihood ratio in R (see Lehmann, 1959). Thus the test which rejects H_0 for large values of R is uniformly most powerful invariant ($UMPI$) under the full linear group for testing H_0 against H_1.

Since $R \geqslant$ constant, implies that $T^2 \geqslant$ constant we get the following theorem.

Theorem 4.1.1. *For Problem 1, Hotelling's T^2 test which rejects H_0 whenever $T^2 \geqslant C$, the constant C depends on the level α of the test, is $UMPI$ for testing H_0 against H_1.*

The group $G_l(p)$ induces on the space of $(\bar{X}, S)$ the transformations

$$(\bar{X}, S) \to (g\bar{X}, gSg'), \ \ g \in G_l(p).$$

It is easy to conclude that any statistical test $\phi(\bar{X}, S)$, a function of $(\bar{X}, S)$ whose power function depends on (μ, Σ) only through δ, is almost invariant under $G_l(p)$. Since

$$\begin{aligned}E_{\mu,\Sigma}\phi(\bar{X},S) &= E_{g^{-1}\mu,g^{-1}\Sigma g^{-1'}}\phi(\bar{X},S)\\ &= E_{\mu,\Sigma}\phi(g\bar{X},gSg'),\end{aligned}$$

thus

$$E_{\mu,\Sigma}(\phi(\bar{X},S)-\phi(g\bar{X},gSg'))\equiv 0$$

for all (μ,Σ). Using the fact that the joint probability density function of $(\bar{X},S)$ is complete, we conclude that

$$\phi(\bar{X},S)=\phi(g\bar{X},gSg')$$

for all $g\in G_l(p)$, except possibly for a set of measure zero. As there exists a left invariant measure (Example 2.1.3) on $G_l'(p)$, which is also right invariant, any almost invariant function is invariant (Lehmann, 1959). Hence we obtain the following Theorem for Problem 1.

Theorem 4.1.2. *Among all tests of H_0: $\xi=0$ with power function depending on δ, Hotelling's T^2 test is UMP:*

Problem 2. Let T_1 be the group of translations such that $t_1,\in T_1$ translates the last $p-p_1$ components of each X^α, $\alpha=1,\dots,N$ and let G_1 be the subgroup of $G_l(p)$ such that $g\in G_1$ has the following form

$$g=\begin{pmatrix} g_{(11)} & 0\\ g_{(21)} & g_{(22)}\end{pmatrix}$$

where $g_{(11)}$ is the $p_1\times p_1$ upper left-hand corner submatrix of g. This problem is invariant under the affine group (G_1,T_1) such that, for $g\in G_1, t_1\in T_1$, (g,t_1) transform

$$X^\alpha\to gX^\alpha+t_1,\quad \alpha=1,\dots,N\,.$$

A maximal invariant in the space of $(\bar{X},S)$ under (G_1,T_1) is R_1, where

$$\bar{R}_1=N\bar{X}'_{(1)}(S_{(11)}+N\bar{X}_{(1)}\bar{X}'_{(1)})^{-1}\bar{X}_{(1)} \tag{4.7}$$

as defined in Example 2.4.3 with $d_1=p_1$. A corresponding maximal invariant in the parametric space of (ξ,Σ) under (G_1,T_1) is $\bar{\delta}_1=N\xi'_{(1)}\Sigma^{-1}_{(11)}\xi_{(1)}$. For invariant test this problem reduces to testing $H_0:\bar{\delta}_1=0$ against the alternatives $H_1:\bar{\delta}_1>0$. From Theorem 2.8.1 the *pdf* of $\bar{R}_1$ is given by

$$\begin{aligned}f_{\bar{\delta}_1}(\bar{r}_1) &= \frac{\Gamma(\frac{1}{2}N)}{\Gamma(\frac{1}{2}p_1)\Gamma(\frac{1}{2}(N-p_1))}\bar{r}_1^{\frac{1}{2}p_1-1}(1-\bar{r}_1)^{\frac{1}{2}(N-p_1)-1}\\ &\quad\times\exp\left\{-\frac{1}{2}\bar{\delta}_1\right\}\phi\left(\frac{N}{2},\frac{p_1}{2};\frac{1}{2}\bar{r}_1\bar{\delta}_1\right).\end{aligned} \tag{4.8}$$

From (4.8) it follows that the distribution of $\bar{R}_1$ possesses a monotone likelihood ratio in $\bar{R}_1$. Hence we get the following Theorem.

Theorem 4.1.3. *For problem 2 the test which rejects H_0 for large values of $\bar{R}_1$ is UMPI under (G_1, T_1) for testing H_0 against H_1.*

Note: As in Problem 1 this UMPI test is also the likelihood ratio test for Problem 2. Using the arguments of Theorem 4.1.2 we can prove from Theorem 4.1.3 that among all tests of H_0 with power function depending on $\bar{\delta}_1$, the test which rejects H_0 for large values of $\bar{R}_1$ is UMP for Problem 2.

Problem 3. Let T_2 be the translation group which translates the last $p - p_1 - p_2$ components of each X^α, $\alpha = 1, \ldots, N$ and let G_2 be the subgroup of $G_l(p)$ such that $g \in G_2$ has the form

$$g = \begin{pmatrix} g_{(11)} & 0 & 0 \\ g_{(21)} & g_{(22)} & 0 \\ g_{(31)} & g_{(32)} & g_{(33)} \end{pmatrix}$$

where $g_{(11)}$ is $p_1 \times p_1$ and $g_{(22)}$ is $p_2 \times p_2$. This problem is invariant under the affine group (G_2, T_2) transforming

$$X^\alpha \to gX^\alpha + t_2, \;\; \alpha = 1, \ldots, N$$

where $g \in G_2$ and $t_2 \in T_2$. A maximal invariant in the space of $(\bar{X}, S)$ under (G_2, T_2) is $(\bar{R}_1, \bar{R}_2)$, where

$$\begin{aligned} \bar{R}_1 &= N\bar{X}'_{(1)}(S_{(11)} + N\bar{X}_{(1)}\bar{X}'_{(1)})^{-1}\bar{X}_{(1)}\,, \\ \bar{R}_1 + \bar{R}_2 &= N\bar{X}'_{[2]}(S_{[22]} + N\bar{X}_{[2]}\bar{X}'_{[2]})^{-1}\bar{X}_{[2]} \end{aligned}$$

as defined in Example 2.4.3 with $d_1 = p_1$, $d_2 = p_2$.

A corresponding maximal invariant in the parametric space of (ξ, Σ) is $(\bar{\delta}_1, \bar{\delta}_2)$, defined by,

$$\begin{aligned} \bar{\delta}_1 &= N\xi'_{(1)}\Sigma^{-1}_{(11)}\xi_{(1)}\,, \\ \bar{\delta}_1 + \bar{\delta}_2 &= N\xi'_{[2]}\Sigma^{-1}_{[22]}\xi_{[2]}\,. \end{aligned}$$

Under the hypothesis $H_0, \bar{\delta}_1 = 0$, $\bar{\delta}_2 = 0$ and under the alternatives $H_1, \bar{\delta}_1 = 0$, $\bar{\delta}_2 > 0$. From Theorem 2.3.1 the joint *pdf* of $\bar{R}_1$, $\bar{R}_2$ is given by

$$
\begin{aligned}
f(\bar{r}_1, \bar{r}_2) = & \frac{\Gamma(\frac{1}{2}N)}{\Gamma(\frac{1}{2}p_1)\Gamma(\frac{1}{2}p_2)\Gamma(\frac{1}{2}(p-p_1-p_2))} \\
& \times \bar{r}_1^{\frac{1}{2}p_1-1}\bar{r}_2^{\frac{1}{2}p_2-1}(1-\bar{r}_1-\bar{r}_2)^{\frac{1}{2}(N-p_1-p_2)-1} \\
& \times \exp\left\{-\frac{1}{2}(\bar{\delta}_1+\bar{\delta}_2)+\frac{1}{2}\bar{r}_1\bar{\delta}_2\right\} \\
& \times \phi\left(\frac{1}{2}N, \frac{1}{2}p_1; \frac{1}{2}\bar{r}_1\bar{\delta}_1\right)\, \phi\left(\frac{1}{2}(N-p_1), \frac{1}{2}p_2; \bar{r}_2\bar{\delta}_2\right) .
\end{aligned}
\tag{4.9}
$$

From Giri (1977) the likelihood ratio test of this problem rejects H_0 whenever

$$
Z = \frac{1-\bar{R}_1-\bar{R}_2}{1-\bar{R}_1} \geqslant c\,, \tag{4.10}
$$

where the constant c depends on the level α of the test and under H_0, Z is distributed as central beta with parameters $(\frac{1}{2}(N-p_1-p_2), \frac{1}{2}p_2)$.

From (4.9) it follows that the likelihood ratio test is not UMPI. However for fixed p, the likelihood ratio test is approximately optimum as the sample size N is large (Wald, 1943). Thus if p is not large, it seems likely that the sample size commonly occurring in practice will be fairly large enough for this result to hold. However, if the dimension p is large, it might be that the sample size N must be extremely large for this result to apply.

We shall now show that the likelihood ratio test is not locally best invariant (LBI) as $\bar{\delta}_2 \to 0$. The LBI test rejects H_0 whenever

$$
\bar{R}_1 + \frac{N-p_1}{p_2}\bar{R}_2 \geqslant c \tag{4.11}
$$

where the constant c depends on the level α of the test. Hotelling's T^2 test which rejects H_0 whenever $\bar{R}_1+\bar{R}_2 \geqslant c$, the constant c depends on the level α of the test, does not coincide with the LBI test, and hence it is locally worse for Problem 3.

Definition 4.1.1. (LBI test). For testing $H_0 : \theta \in \Omega_{H_0}$, an invariant test ϕ^* of level α is LBI if there exists an open neighborhood $\tilde{\Omega}_1$ of Ω_{H_0} such that

$$
E_\theta(\phi^*) \geqslant E_\theta(\phi), \qquad \theta \in \tilde{\Omega}_1 - \Omega_{H_0}\,, \tag{4.12}
$$

where ϕ is any other invariant test of level α.

Theorem 4.1.4. *For testing* $H_0 : \bar{\delta}_1 = 0, \bar{\delta}_2 = 0$ *against* $H_1 : \bar{\delta}_1 = 0, \bar{\delta}_2 > 0$, *the test which rejects* H_0 *whenever* $\bar{R}_1 + \frac{N-p_1}{p_2}\bar{R}_2 \geqslant c$, *where the constant* c *depends on the level* α *of the test, is LBI as* $\bar{\delta}_2 \to 0$.

Proof. Since $(\bar{R}_1, \bar{R}_2)$ is a maximal invariant under the affine group (G_2, T_2), the ratio of densities of $(\bar{R}_1, \bar{R}_2)$ under H_1, to that under H_0, is given by,

$$\frac{f(\bar{r}_1, \bar{r}_2|\bar{\delta}_2)}{f(\bar{r}_1, \bar{r}_2|0)} = 1 + \frac{\bar{\delta}_2}{2}\left(-1 + \bar{r}_1 + \frac{N-p_1}{p_2}\bar{r}_2\right) + o(\delta_2)$$

as $\bar{\delta}_2 \to 0$. Hence the power of any invariant test ϕ is given by

$$E_{\delta_2}(\Phi) = \alpha + E_{\delta_{2=0}}\Phi\bar{\delta}_2\left(-1 + \bar{R}_1 + \frac{N-p_1}{p_2}\bar{R}_2\right) + o(\bar{\delta}_2)$$

as $\bar{\delta}_2 \to 0$, which is maximized by taking $\Phi = 1$ whenever $\bar{R}_1 + \frac{N-p_1}{p_2}\bar{R}_2 \geqslant c$. □

Asymtotically Best Invariant (ABI) Test

Let Φ be an invariant test for testing $H_0 : \bar{\delta}_1 = 0$, $\bar{\delta}_2 = 0$ against $H_1 : \bar{\delta}_1 = 0$, $\bar{\delta}_2 = \lambda$ with associated critical region $R = \{(\bar{r}_1, \bar{r}_2) : U(\bar{r}_1, \bar{r}_2) \geqslant c_\alpha\}$ satisfying $P_{H_0}(R) = \alpha$;

$$\begin{aligned} P_{H_1}(R) &= 1 - \exp\{-H(\lambda)(1 + o(1))\}\,; \\ \frac{f(\bar{r}_1, \bar{r}_2|H_1)}{f(\bar{r}_1, \bar{r}_2|H_0)} &= \exp\{H(\lambda)[G(\lambda) + R(\lambda)U(\bar{r}_1, \bar{r}_2)]\} \\ &\quad + B(\bar{r}_1, \bar{r}_2; \lambda)\,; \end{aligned} \tag{4.12}$$

where $B(\bar{r}_1, \bar{r}_2; \lambda) = o(H(\lambda))$ as $\lambda \to \infty$ and $0 < c_1 < R(\lambda) < c_2 < \infty$. Then Φ is an ABI test for testing H_0 against H_1 as $\lambda \to \infty$.

Theorem 4.1.5. *For testing $H_0 : \bar{\delta}_1 = \bar{\delta}_2 = 0$ against $H_1 : \bar{\delta}_1 = 0, \bar{\delta}_2 = \lambda > 0$, Hotelling's test which rejects H_0 whenever $\bar{R}_1 + \bar{R}_2 \geqslant c$, the constant c depending on the level α of the test, is ABI as $\lambda \to \infty$.*

Proof. Since

$$\begin{aligned} \Phi(a, b, x) &= 1 + \frac{a}{b}x + \frac{a(a+1)}{b(b+1)}\frac{x^2}{2!} + \cdots \\ &= \exp\{x(1 + o(1)\} \quad \text{as } x \to \infty\,, \end{aligned}$$

we get from (4.9)

$$\frac{f(\bar{r}_1, \bar{r}_2|H_1)}{f(\bar{r}_1, \bar{r}_2|H_0)} = \exp\{(1 - \bar{r}_1 - \bar{r}_2)(1 + B(\bar{r}_1, \bar{r}_2; \lambda))\}\,; \tag{4.13}$$

where $B(\bar{r}_1, \bar{r}_2; \lambda) = o(1)$ as $\lambda \to \infty$. It follows from (4.2) that the pdf of $U = \bar{R}_1 + \bar{R}_2$ satisfies

$$\begin{aligned} P(U(R_1, R_2) &= \bar{R}_1 + \bar{R}_2 \leqslant c_\alpha) \\ &= \exp\left\{\frac{\lambda}{2}(c_\alpha - 1)(1 + o(1))\right\}. \end{aligned} \tag{4.14}$$

Thus, from (4.12) with $H(\lambda) = \frac{1}{2}(c_\alpha - 1)$, we conclude that Hotelling's T^2 test is ABI. □

4.2. The Classification Problem (Two Populations)

Given two different p-dimensional populations P_1, P_2 characterized by the probability density functions p_1, p_2 respectively; and an observation $x = (x_1, \ldots, x_p)'$ the classification problem deals with classifying it to one of the two populations. We assume for certainty that it belongs to one of the two populations. Denote by a_i, $i = 1, 2$ that x comes from P_i and let us denote, for simplicity, the states of nature P_i simply by i. Let $L(j, a_i)$ be the loss incurred by taking the action a_i when j is the true state of nature having property that $L(i, a_i) = 0$, $L(i, a_j) = c_j > 0$, $j \neq i = 1, 2$.

Let $(\pi, 1-\pi)$ be the *a priori* probability distribution on the states of nature. The posterior probability distribution $(\xi, 1 - \xi)$, given the observation x, is given by

$$\xi = \frac{\pi p_1(x)}{\pi p_1(x) + (1 - \pi)p_2(x)}. \tag{4.15}$$

The posterior expected loss of taking action a_1 is given by

$$\frac{(1 - \pi)c_1 p_2(x)}{\pi p_1(x) + (1 - \pi)p_2(x)} \tag{4.16}$$

and for taking action a_2 is given by

$$\frac{\pi c_2 p_1(x)}{\pi p_1(x) + (1 - \pi)p_2(x)}. \tag{4.17}$$

From (4.16)–(4.17) it follows that we take action a_2 if

$$\frac{\pi c_2 p_1(x)}{\pi p_1(x) + (1 - \pi)p_2(x)} < \frac{(1 - \pi)c_1 p_2(x)}{\pi p_1(x) + (1 - \pi)p_2(x)} \tag{4.18}$$

and action a_1 if

$$\frac{\pi c_2 p_1(x)}{\pi p_1(x) + (1 - \pi)p_2(x)} > \frac{(1 - \pi)c_1 p_2(x)}{\pi p_1(x) + (1 - \pi)p_2(x)} \tag{4.19}$$

and randomize when equality holds.

(4.18) and (4.19) can be simplified as,

$$\begin{aligned}
&\text{take action } a_2 \quad \text{if} \quad \frac{p_2(x)}{p_1(x)} > \frac{\pi c_2}{(1-\pi)e_1} = K \text{ (say)}, \\
&\text{take action } a_1 \quad \text{if} \quad \frac{p_2(x)}{p_1(x)} < K
\end{aligned} \tag{4.20}$$

and randomize when $\frac{p_2(x)}{p_1(x)} = K$.

Let us now specialize the case of two p-variate normal populations with different means but the same covariance matrix. Assume

$$p_1 : N_p(\xi, \Sigma),$$
$$p_2 : N_p(\mu, \Sigma),$$

where $\mu = (\mu_1, \ldots, \mu_p)'$, $\xi = (\xi_1, \ldots, \xi_p)' \in R^p$ and Σ is positive definite. Hence

$$\begin{aligned}
\frac{p_2(x)}{p_1(x)} &= \exp\left\{\frac{1}{2}\left[(x-\xi)'\Sigma^{-1}(x-\xi) - (x-\mu)'\Sigma^{-1}(x-\mu)\right]\right\} \\
&= \exp\left\{x'\Sigma^{-1}(\mu-\xi) + \frac{1}{2}(\xi'\Sigma^{-1}\xi - \mu'\Sigma^{-1}\mu)\right\}.
\end{aligned}$$

Using (4.20) we take action a_2 if

$$x'\Sigma^{-1}(\mu-\xi) \geqslant \frac{1}{2}(\mu'\Sigma^{-1}\mu - \xi'\Sigma^{-1}\xi) + \log K$$

and take action a_1, otherwise. The linear functional $\Gamma = \Sigma^{-1}(\mu - \xi)$ is called Fisher's discriminant function. In practice Σ, ξ, μ are usually unknown, and we consider estimation and testing problems for $\Gamma = (\Gamma_1, \ldots, \Gamma_p)'$.

Let X^α, $\alpha = 1, \ldots, N_1$ be a random sample of size N_1 from $N_p(\xi, \Sigma)$ and let Y^α, $\alpha = 1, \ldots, N_2$ be a sample of size N_2 from $N_p(\mu, \Sigma)$. Invariance and sifficiency considerations lead us to consider the statistic

$$\begin{aligned}
X &= \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1}(\bar{X} - \bar{Y}), \\
S &= \sum_{\alpha=1}^{N_1}(X^\alpha - \bar{X})(X^\alpha - \bar{X})' + \sum_{\alpha=1}^{N_2}(Y^\alpha - \bar{Y})(Y^\alpha - \bar{Y})',
\end{aligned}$$

where $N_1\bar{X} = \sum_{\alpha=1}^{N_1} X^\alpha$, $N_2\bar{Y} = \sum_{\alpha=1}^{N_2} Y^\alpha$ for the statistical inferences of Γ. We consider two testing problems about Γ.

Problem 4. To test the hypothesis $H_0 : \Gamma_{p_1+1} = \cdots = \Gamma_p = 0$ against the alternatives $H_1 : \Gamma \neq 0$.

Problem 5. To test the hypothesis $H_0 : \Gamma_{p_1+1} = \cdots = \Gamma_p = 0$ against the alternatives $H_1 : \Gamma_{p_1+p_2+1} = \cdots = \Gamma_p = 0$.

Without any loss of generality we restate our setup in the following canonical form, where $X = (X_1, \ldots, X_p)'$ is distributed as $N_p(\eta, \Sigma)$ and S is distributed, independently of X, as Wishart $W_p(n, \Sigma)$ and $\Gamma = \Sigma^{-1}\eta$. Since Σ is positive definite, $\Gamma = 0$ if and only if $\eta = 0$. Thus the problem of testing the hypothesis $\Gamma = 0$ against $\Gamma \neq 0$ is equivalent to Problem 1 considered earlier in this chapter.

Problem 4. It remains invariant under the group G of $p \times p$ nonsingular matrices g of the form

$$g = \begin{pmatrix} g_{(11)} & 0 \\ g_{(21)} & g_{(22)} \end{pmatrix} \tag{4.21}$$

where $g_{(11)}$ is of order p_1, operating as

$$(\bar{X}, S; \Gamma, \Sigma) \to (g\bar{X}, gSg'; (g')^{-1}\Gamma, g\Sigma g')$$

where $\bar{X}, \frac{S}{N}$ are the mean and the sample covariance matrix based on N observations on X. By Example 2.4.3 a maximal invariant in the space of the sufficient statistic $(\bar{X}, S)$ under G is given by $(\bar{R}_1, \bar{R}_2)$, with $d_1 = p_1$, $d_2 = p_2$ and $p = p_1 + p_2$. The joint *pdf* of $(\bar{R}_1, \bar{R}_2)$ is given by (4.4) where the expressions for $\bar{\delta}_1, \bar{\delta}_2$ in terms of Γ, Σ are given by

$$\begin{aligned} \bar{\delta}_1 + \bar{\delta}_2 &= N\Gamma'\Sigma\Gamma\,, \\ \bar{\delta}_2 &= N\Gamma'_{(2)}(\Sigma^{22})^{-1}\Gamma_{(2)}\,, \end{aligned} \tag{4.22}$$

with $\Sigma^{22} = (\Sigma_{(22)} - \Sigma_{(21)}\Sigma^{-1}_{(11)}\Sigma_{(12)})^{-1}$. Under the hypothesis $H_0\,\bar{\delta}_1 > 0$, $\bar{\delta}_2 = 0$ and under the alternatives $H_i\bar{\delta}_i > 0$, $i = 1, 2$. From (4.4) it follows that the probability density function of R_1 under H_0 is given by

$$\begin{aligned} f(\bar{r}_1|\bar{\delta}_1) = {} & \frac{\Gamma(\frac{1}{2}N)}{\Gamma(\frac{1}{2}p_1)\Gamma(\frac{1}{2}(N-p_1))} \\ & \times (\bar{r}_1)^{\frac{1}{2}p_1-1}(1-\bar{r}_1)^{\frac{1}{2}(N-p_1)-1} \\ & \times \exp\left\{-\frac{1}{2}\bar{\delta}_1\right\}\phi\left(\frac{1}{2}N, \frac{1}{2}p_1; \frac{1}{2}\bar{r}_1\bar{\delta}_1\right), \end{aligned} \tag{4.23}$$

and $\bar{R}_1$ is sufficient for $\bar{\delta}_1$.

Giri (1964) has shown that the likelihood ratio test for this problem rejects H_0 whenever

$$z = \frac{1 - \bar{r}_1 - \bar{r}_2}{1 - \bar{r}_1} \leqslant c\,, \tag{4.24}$$

where the constant c depends on the level α of the test and under $H_0 Z$ has a central beta distribution with parameter $(\frac{1}{2}(N-p), \frac{1}{2}p_2)$ and is independent of $\bar{R}_1$.

To prove that the likelihood ratio test is uniformly most powerful similar we first check that the family of *pdf*

$$\{f(\bar{r}_1|\bar{\delta}_1) : \bar{\delta}_1 \geqslant 0\} \tag{4.25}$$

is boundedly complete.

Definition 4.2.1. (Boundedly complete). A family of distributions $\{P_\delta(x) : \delta \in \Omega\}$ of a random variable X or the corresponding family of pdfs $\{p_\delta(x) : \delta \in \Omega\}$ is boundedly complete if

$$\begin{aligned} E_\delta h(X) &= \int h(x) dP_\delta(x) \\ &= \int h(x) p_\delta(x) dx \\ &= 0 \end{aligned}$$

for all $\delta \in \Omega$ and for any real valued bounded function $h(X)$, implies that $h(X) = 0$ almost everywhere with respect to each of the measures P_δ.

We also say X is a complete statistic or X is complete.

Lemma 4.2.1. *The family of pdfs* $\{f(\bar{r}_1|\bar{\delta}_1) : \bar{\delta}_1 \geqslant 0\}$ *is boundedly complete.*

Proof. For any real-valued bounded function $h(\bar{R}_1)$ of $\bar{R}_1$ we get (using (4.23))

$$\begin{aligned} E_{\bar{\delta}_1}(h(\bar{R}_1)) &= \int h(\bar{r}_1) f(\bar{r}_1|\bar{\delta}_1) d\bar{r}_1 \\ &= \exp\left\{-\frac{1}{2}\bar{\delta}_1\right\} \\ &\quad \times \sum_j \frac{\left(\frac{1}{2}\bar{\delta}_1\right)^j aj}{j!} \quad \int_0^1 h^*(\bar{r}_1)(\bar{r}_1)^{\frac{1}{2}p_1+j-1}(1-\bar{r}_1)^{\frac{1}{2}(N-p_1)-1} d\bar{z}_1 \,; \\ &= \exp\left\{-\frac{1}{2}\bar{\delta}_1\right\} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\bar{\delta}_1\right)^j}{j!} a_j \int_0^1 h^*(\bar{r}_1)(\bar{r}_1)^j d\bar{r}_1 \end{aligned}$$

where

$$h^*(\bar{r}_1) = h(\bar{r}_1)\bar{r}_1^{\frac{1}{2}p_1 - 1}(1-\bar{r}_1)^{\frac{1}{2}(N-p_1)-1},$$
$$a_j = \frac{B(\frac{1}{2}(p-p_1), \frac{1}{2}(N-p_1)+j)}{B(\frac{1}{2}(N-p_1), \frac{1}{2}p_1+j)B(\frac{1}{2}(N-p), \frac{1}{2}(p-p_1))}$$

and $B(p,q)$ is the beta function. Hence $E_{\bar{\delta}_1}h(\bar{R}_1) = 0$ implies that

$$\sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\bar{\delta}_1\right)^j a_j}{j!} \int_0^1 h^*(\bar{r}_1)(\bar{r}_1)^j d\bar{r}_1 = 0. \tag{4.26}$$

Since the right-hand side of (4.26) is a polynomial in $\bar{\delta}_1$, (4.26) implies that all its coefficents must be zero. In other words

$$\int_0^1 h^*(\bar{r}_1)\bar{r}_1^j d\bar{r}_1 = 0, \quad j = 0, 1, 2, \ldots. \tag{4.27}$$

Let

$$h^*(\bar{R}_1) = h^{*+}(\bar{R}_1) - h^{*-}(\bar{R}_1)$$

where h^{*+} and h^{*-} denote the positive and negative parts of h^*. Hence we get from (4.27)

$$\int_0^1 h^{*+}(\bar{r}_1)\bar{r}_1^j d\bar{r}_1 = \int_0^1 h^{*-}(\bar{r}_1)\bar{r}_1^j d\bar{r}_1$$

for $j = 0, 1, \ldots$; which implies that

$$h^{*+}(\bar{r}_1) = h^{*-}(\bar{r}_1)$$

for all $\bar{r}_1$ except possibly on a set of measure 0. Hence we conclude that $h^*(\bar{R}_1) = 0$ almost everywhere $\{P_{\bar{\delta}_1}(\bar{R}_1) : \bar{\delta}_1 \geqslant 0\}$ which, in turn, implies that $h(\bar{R}_1) = 0$ almost everywhere. □

Since $\bar{R}_1$ is complete, it follows; from Lehmann (1959), p. 34; that any invariant test $\phi(\bar{R}_1, \bar{R}_2)$ of level α has Neyman Structure with respect to $\bar{R}_1$, i.e.,

$$E_{H_0}(\phi(\bar{R}_1, \bar{R}_2)|\bar{R}_1) = \alpha.$$

To find the uniformly most powerful test among all similar invariant tests we need the ratio $\tilde{R}$

$$\tilde{R} = \frac{f_{H_1}(\bar{r}_1, \bar{r}_2|\bar{R}_1 = \bar{r}_1)}{f_{H_0}(\bar{r}_1, \bar{r}_2|\bar{R}_1 = \bar{r}_1)}$$

where $f_{H_i}(\bar{r}_1, \bar{r}_2 | \bar{R}_1 = \bar{r}_1)$ denotes the conditional pdf of $(\bar{R}_1, \bar{R}_2)$ given $\bar{R}_1 = \bar{r}_1$ under $H_i, i = 0, 1$. From (4.9) we get

$$\tilde{R} = \exp\left\{-\frac{1}{2}\bar{\delta}_2(1-\bar{r}_1)\right\}\phi\left(\frac{1}{2}(N-p_1), \frac{1}{2}(p-p_1); \frac{1}{2}\bar{r}_2\bar{\delta}_2\right). \tag{4.28}$$

From (4.28), using the fact that Z is independent of $\bar{R}_1$ when H_0 is true, we get the following theorem:

Theorem 4.2.1. *For Problem 4, the likelihood ratio test given in* (4.24) *is uniformly most powerful invariant similar.*

Problem 5. It remains invariant under the group G of $p \times p$ nonsingular matrices g of the form

$$g = \begin{pmatrix} g_{(11)} & 0 & 0 \\ g_{(21)} & g_{(22)} & 0 \\ g_{(31)} & g_{(32)} & g_{(33)} \end{pmatrix}$$

with $g_{(11)} : p_1 \times p_2$, $g_{(22)} : p_2 \times p_2$ submatrices of g. A maximal invariant in the space of $(\bar{X}, S)$ (as in Problem 4) under G is $(\bar{R}_1, \bar{R}_2, \bar{R}_3)$ as defined in Example 2.4.3 with $d_1 = p_1$, $d_2 = p_2$ and $d_3 = p - p_1 - p_2$. By Theorem 2.8.1 the joint probability density function of $\bar{R}_1, \bar{R}_2, \bar{R}_3$ is given by

$$\begin{aligned} f(\bar{r}_1, \bar{r}_2, \bar{r}_3) = & \frac{\Gamma(\frac{1}{2}N)}{\Gamma(\frac{1}{2}p_1)\Gamma(\frac{1}{2}p_2)\Gamma(\frac{1}{2}(p-p_1-p_2))} \\ & \times \bar{r}_1^{\frac{1}{2}p_1-1}\bar{r}_2^{\frac{1}{2}p_2-1}\bar{r}_3^{\frac{1}{2}(p-p_1-p_2)-1} \\ & \times (1-\bar{r}_1-\bar{r}_2-\bar{r}_3)^{\frac{1}{2}(N-p)-1} \\ & \times \exp\left\{-\frac{1}{2}\sum_{i=1}^{3}\bar{\delta}_i + \frac{1}{2}\sum_{j=1}^{3}\bar{r}_j\sum_{i>j}\bar{\delta}_i\right\} \\ & \prod_{i=1}^{3}\phi\left(\frac{1}{2}(N-\sigma_{i-1}), \frac{1}{2}p_i; \frac{1}{2}\bar{r}_i\bar{\delta}_i\right), \end{aligned} \tag{4.29}$$

where

$$\sigma_i = \sum_{j=1}^{i} p_j,\ \sigma_0 = 0,\ p_3 = p - p_1 - p_2\,.$$

and

$$\begin{aligned}
\bar{\delta}_1 &= N(\Sigma_{(11)}\Gamma_{(1)} + \Sigma_{(12)}\Gamma_{(2)} + \Sigma_{(13)}\Gamma_{(3)})'\Sigma_{(11)}^{-1} \\
&\quad \times (\Sigma_{(11)}\Gamma_{(1)} + \Sigma_{(12)}\Gamma_{(2)} + \Sigma_{(13)}\Gamma_{(3)}), \\
\bar{\delta}_1 + \bar{\delta}_2 &= \begin{pmatrix} \Sigma_{(11)}\Gamma_{(1)} + \Sigma_{(12)}\Gamma_{(2)} + \Sigma_{(13)}\Gamma_{(3)} \\ \Sigma_{(21)}\Gamma_{(1)} + \Sigma_{(22)}\Gamma_{(2)} + \Sigma_{(23)}\Gamma_{(3)} \end{pmatrix}' \Sigma_{[22]}^{-1} \\
&\quad \times \begin{pmatrix} \Sigma_{(11)}\Gamma_{(1)} + \Sigma_{(12)}\Gamma_{(2)} + \Sigma_{(13)}\Gamma_{(3)} \\ \Sigma_{(21)}\Gamma_{(1)} + \Sigma_{(22)}\Gamma_{(2)} + \Sigma_{(23)}\Gamma_{(3)} \end{pmatrix}, \\
\bar{\delta}_3 &= N\Gamma_{(3)}'(\Sigma_{(33)} - \begin{pmatrix} \Sigma_{(13)} \\ \Sigma_{(23)} \end{pmatrix}' \Sigma_{[22]}^{-1} \begin{pmatrix} \Sigma_{(13)} \\ \Sigma_{(23)} \end{pmatrix})\Gamma_{(3)}.
\end{aligned}$$

Under H_0, $\bar{\delta}_1 > 0$, $\bar{\delta}_2 = 0$, $\bar{\delta}_3 = 0$, and under H_1, $\bar{\delta}_1 > 0$, $\bar{\delta}_2 > 0$, $\bar{\delta}_3 = 0$. From (4.29) it follows that $\bar{R}_1$ is sufficient for $\bar{\delta}_1$ under H_0. From Giri (1964) the likelihood ratio test of this problem rejects H_0 whenever

$$Z = \frac{1 - \bar{R}_1 - \bar{R}_2}{1 - \bar{R}_1} \leqslant c \tag{4.30}$$

where the constant c depends on the level α of the test and under H_0 Z is distributed independently of $\bar{R}_1$ as central beta with parameter $(\frac{1}{2}(N - p_1 - p_2), \frac{1}{2}p_2)$.

Let $\phi(\bar{R}_1, \bar{R}_2, \bar{R}_3)$ be any invariant level α test of H_0 against H_1. From Lemma 4.2.1 $\bar{R}_1$ is boundedly complete. Thus ϕ has Neyman Structure with respect to $\bar{R}_1$. Since the conditional distribution of $(\bar{R}_1, \bar{R}_2, \bar{R}_3)$ given $\bar{R}_1$ does not depend on $\bar{\delta}_1$, the condition that the level α test ϕ has Neyman structure, that is,

$$E_{H_0}(\phi(\bar{R}_1, \bar{R}_2, \bar{R}_3)|\bar{R}_1) = \alpha$$

reduces the problem to that of testing the simple hypothesis $\bar{\delta}_2 = 0$ against the alternatives $\bar{\delta}_2 > 0$ on each surface $\bar{R}_1 = \bar{r}_1$. In this conditional situation the most powerful level α invariant test of $\bar{\delta}_2 = 0$ against the simple alternative $\bar{\delta}_2 = \bar{\delta}_2^0 > 0$ is given by,

$$\text{reject } H_0 \qquad \text{whenever} \quad \phi\left(\frac{1}{2}(N - p_1), \frac{1}{2}p_2; \frac{1}{2}\bar{r}_2\bar{\delta}_2^0\right) \geqslant c, \tag{4.31}$$

where the constant c depends on α. Since

$$\bar{R}_2 = (1 - \bar{R}_1)(1 - Z)$$

and Z is independent of $\bar{R}_1$, we get the following theorem.

Theorem 4.2.2. *For Problem 5 the likelihood ratio test given in* (4.30) *is uniformly most powerful invariant similar.*

4.3. Test of Multiple Correlation

Let X^α, $\alpha = 1, \ldots, N$ be N independent normal p-vectors with common mean μ and common positive definite covariance matrix Σ and let $N\bar{X} = \Sigma_{\alpha=1}^N X^\alpha$, $S = \Sigma_{\alpha=1}^N (X^\alpha - \bar{X})(X^\alpha - \bar{X})'$. As usual we assume $N > p$ so that S is positive definite with probability one. Partition Σ and S as

$$\Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}, \; S = \begin{pmatrix} S_{(11)} & S_{(12)} \\ S_{(21)} & S_{(22)} \end{pmatrix}$$

where $\Sigma_{(22)}$, $S_{(22)}$ are of order $p-1$. Let

$$\rho^2 = \frac{\Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}}{\Sigma_{(11)}},$$

the positive square root of ρ^2 is called the population multiple correlation coefficent.

Problem 6. To test $H_0 : \rho^2 = 0$ against the alternative $H_1 : \rho^2 > 0$. This problem is invariant under the affine group (G, T), operating as

$$(\bar{X}, S; \mu, \Sigma) \to (g\bar{X} + t, gSg'; g\mu + t, g\Sigma g')$$

where $g \in G$ is a $p \times p$ nonsingular matrix of the form

$$g = \begin{pmatrix} g_{(11)} & 0 \\ 0 & g_{(22)} \end{pmatrix}$$

where $g_{(22)}$ is of order $p-1$ and T is the group of translations t of components of each X^α. A maximal invariant in the space of $(\bar{X}, S)$ under the affine group (G, T) is

$$R^2 = \frac{S_{(22)}S_{(22)}^{-1}S_{(21)}}{S_{(11)}}$$

which is popularly called the square of sample multiple correlation coefficent.

Distribution of $\mathbf{R^2}$. To find the distribution of R^2 we first observe that

$$\frac{R^2}{1-R^2} = \frac{S_{(12)}S_{(22)}^{-1}S_{(21)}}{S_{(11)} - S_{(12)}S_{(22)}^{-1}S_{(21)}}.$$

From Giri (1977)

$$\frac{S_{(11)} - S_{(12)}S_{(22)}^{-1}S_{(21)}}{\Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}} \tag{4.31}$$

is distributed as χ^2_{N-p} with $N-p$ degrees of freedom and is independent of $S_{(12)}$, $S_{(22)}$. Thus $R^2/(1-R^2)$ is distributed as

$$\frac{1}{\chi^2_{N-p}} \cdot \frac{S_{(12)}S_{(22)}^{-1}S_{(22)}}{\Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}} .$$

But the conditional distribution of

$$\frac{S_{(12)}S_{(22)}^{-1/2}}{\sqrt{\Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}}}$$

given $S_{(22)}$, is $(p-1)$-variate normal with mean

$$\frac{\Sigma_{(22)}\Sigma_{(22)}^{-1}S_{(22)}^{1/2}}{\sqrt{\Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}}}$$

and covariance I. Hence $R^2/(1-R^2)$ is distributed as

$$\frac{1}{\chi^2_{N-p}} \cdot \chi^2_{p-1}\Big(\frac{\Sigma_{(12)}\Sigma_{(22)}^{-1}S_{(22)}\Sigma_{(22)}^{-1}\Sigma_{(21)}}{\Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}}\Big), \tag{4.32}$$

where $\chi^2_k(\lambda)$ denotes a noncentral chisquare random variable with noncentrality parameter λ and k degrees of freedom. Also

$$\frac{\Sigma_{(12)}\Sigma_{(22)}^{-1}S_{(22)}\Sigma_{(22)}^{-1}\Sigma_{(12)}}{\Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}}$$

is distributed as chisquare χ^2_{N-1}. Since

$$\frac{\Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}}{\Sigma_{(11)} - \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}} = \frac{\rho^2}{1-\rho^2} ,$$

we conclude that $R^2/(1-R^2)$ is distributed as

$$\frac{1}{\chi^2_{N-p}} \cdot \chi^2_{p-1}\Big(\frac{\rho^2}{1-\rho^2}\chi^2_{N-1}\Big) .$$

Using the fact that a noncentral chisquare random variable $\chi^2_m(\lambda)$ can be represented as χ^2_{m+2K}, where K is a Poisson random variable with parameter $\frac{\lambda}{2}$, we conclude that

$$\chi^2_{p-1}\Big(\frac{\rho^2}{1-\rho^2}\chi^2_{N-1}\Big)$$

is distributed as χ^2_{p-1+2K}, where the conditional distribution of K given χ^2_{N-1} is Poisson with parameter $\frac{1}{2}[\rho^2/(1-\rho^2)]\chi^2_{N-1}$. Let $\lambda/2 = \frac{1}{2}\rho^2/(1-\rho^2)$. Then

$$
\begin{aligned}
P(K=k) &= \frac{1}{2^{\frac{1}{2}(N-1)}\Gamma(\frac{1}{2}(N-1))k!} \\
&\quad \times \int_0^\infty \exp\left\{-\frac{1}{2}(\lambda x + x)\right\}\left(\frac{1}{2}\lambda x\right)^k x^{\frac{1}{2}(N-1)-1}dx \\
&= \frac{\Gamma(\frac{1}{2}(N-1)+k)\lambda^k}{k!\Gamma(\frac{1}{2}(N-1))(1-\lambda)^{\frac{1}{2}(N-1)+k}} \\
&= \frac{\Gamma(\frac{1}{2}(N-1)+k)}{k!\Gamma(\frac{1}{2}(N-1))}(\rho^2)^k(1-\rho^2)^{\frac{1}{2}(N-1)} \qquad (4.33)
\end{aligned}
$$

with $k = 0, 1, \ldots$. Thus

$$
\frac{R^2}{1-R^2} \quad \text{is distributed as} \quad \frac{\chi^2_{p-1+2K}}{\chi^2_{N-p}} \qquad (4.34)
$$

where K is a negative binomial with pdf given in (4.33). Simple calculations yield

$$
\frac{(N-p)R^2}{(p-1)(1-R^2)}
$$

has central F distribution with parameter $(p-1, N-p)$ when $\rho^2 = 0$. From (4.33)–(4.34) the noncentral distribution of R^2 is given by

$$
\begin{aligned}
f_{R^2}(r^2) &= \frac{(1-r^2)^{\frac{1}{2}(N-p-2)}(r^2)^{\frac{1}{2}(p-3)}(1-\rho^2)^{\frac{1}{2}(N-1)}}{\Gamma(\frac{1}{2}(N-1))\Gamma(\frac{1}{2}(N-p))} \\
&\quad \times \sum_{i=0}^{\infty}\frac{(\rho^2)^i(r^2)^i\Gamma^2(\frac{1}{2}(N-1)+i)}{i!\Gamma(\frac{1}{2}(p-1)+i)} . \qquad (4.35)
\end{aligned}
$$

It may be checked that a corresponding maximal invariant in the parametric space is ρ^2.

Theorem 4.3.1. *For problem 6 the test which rejects H_0 whenever $R^2 \geqslant c$, the constant c depends on the level α of the test, is uniformly most powerful invariant.*

Proof. From (4.36)

$$f_{\rho^2}(r^2)/f_0(r^2) = (1-\rho^2)^{\frac{1}{2}(N-1)} \times \sum_{i=0}^{\infty} \frac{(\rho^2 r^2)^i \Gamma^2(\frac{1}{2}(N-1)+i)\Gamma(\frac{1}{2}(p-1))}{i!\Gamma(\frac{1}{2}(p-1)+i)\Gamma^2(\frac{1}{2}(N-1))},$$

which for a given value of ρ^2 is an increasing function of r^2. Using Neyman-Pearson Lemma we get the theorem. □

As in Theorem 4.1.2 we can show that among all tests of $H_0 : \rho^2 = 0$ against $H_1 : \rho^2 > 0$ with power function depending only on ρ^2 the R^2-test is UMP.

4.4. Test of Multiple Correlation with Partial Information

Let X be a normally distributed p-dimensional random column vector with mean μ and positive definite covariance matrix Σ, and let X^α, $\alpha = 1, \ldots, N$ $(N > p)$ be a random sample of size N from this distribution. We partition X as $X = (X_1, X'_{(1)}, X'_{(2)})'$ where X_1 is one-dimensional, $X_{(1)}$ is p_1-dimensional and $X_{(2)}$ is p_2-dimensional and $1 + p_1 + p_2 = p$. Let ρ_1 and ρ denote the multiple correlatron coefficents of X_1 with $X_{(2)}$ and with $(X'_{(2)}X'_{(3)})'$, respectively. Denote by $\rho_2^2 = \rho^2 - \rho_1^2$. We consider here the following two testing problems:

Problem 7. To test H_{10}: $\rho^2 = 0$ against the alternatives $H_{1\lambda}$: $\rho_2^2 = 0$, $\rho_1^2 = \lambda > 0$.

Problem 8. To test H_{20}: $\rho^2 = 0$ against the alternatives $H_{2\lambda}$: $\rho_1^2 = 0$, $\rho_2^2 = \lambda > 0$.

Let $N\bar{X} = \Sigma_{\alpha=1}^N X^\alpha$, $S = \Sigma_{\alpha=1}^N (X^\alpha - \bar{X})(X^\alpha - \bar{X})'$, $b_{[i]}$ denote the i-vector consisting of the first i components of a vector b and $c_{[i]}$ denote the $i \times i$ upper left submatrix of a matrix c. Partitions S and Σ as

$$S = \begin{pmatrix} S_{11} & S_{(12)} & S_{(13)} \\ S_{(21)} & S_{(22)} & S_{(23)} \\ S_{(31)} & S_{(32)} & S_{(33)} \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{(12)} & \Sigma_{(13)} \\ \Sigma_{(21)} & \Sigma_{(22)} & \Sigma_{(23)} \\ \Sigma_{(31)} & \Sigma_{(32)} & \Sigma_{(33)} \end{pmatrix}$$

where $S_{(22)}$ and $\Sigma_{(22)}$ are each of dimension $p_1 \times p_1$; $S_{(33)}$ and $\Sigma_{(33)}$ are each of dimension $p_2 \times p_2$. Define

$$\begin{aligned} \rho_1^2 &= \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}/\Sigma_{11}, \\ \rho^2 = \rho_1^2 + \rho_2^2 &= (\Sigma_{(12)}\Sigma_{(13)}) \begin{pmatrix} \Sigma_{(22)} & \Sigma_{(23)} \\ \Sigma_{(32)} & \Sigma_{(33)} \end{pmatrix}^{-1} (\Sigma_{(12)}\Sigma_{(13)})'/\Sigma_{11}, \end{aligned} \tag{4.36a}$$

$$\bar{R}_1 = S_{(12)}S_{(22)}^{-1}S_{(21)}/S_{11},$$
$$\bar{R}_1 + \bar{R}_2 = (S_{(12)}S_{(13)})\begin{pmatrix} S_{(22)} & S_{(23)} \\ S_{(32)} & S_{(33)} \end{pmatrix}^{-1}(S_{(12)}S_{(13)})'/S_{11}.$$

It is easy to verify that the translation group transforming

$$(\bar{X}, S; \mu, \Sigma) \to (\bar{X} + b, S; \mu + b, \Sigma)$$

leaves the present problem invariant and, along with the full linear group G of $p \times p$ nonsingular matrices g

$$g = \begin{pmatrix} g_{11} & 0 & 0 \\ 0 & g_{(22)} & 0 \\ 0 & g_{(32)} & g_{(33)} \end{pmatrix} \tag{4.36}$$

where $g_{11} : 1 \times 1$, $g_{(22)} : p_1 \times p_1$, $g_{(33)}$: $p_2 \times p_2$, generates a group which leaves the present problem invariant. The action of these transformations is to reduce the problem to that where $\mu = 0$ and $S = \Sigma_{\alpha=1}^{N} X^{\alpha}X^{\alpha'}$ is sufficient for Σ, where N has been reduced by one from what it was originally. We now treat later formulation considering X^{α}, $\alpha = 1, \ldots, N \geqslant p \geqslant 2$, to have a common zero mean vector. We therefore consider the group G of transformations g operating as

$$(S, \Sigma) \to (gSg', g\Sigma g')$$

for the invariance of the problem. A maximal invariant in the sample space under G is $(\bar{R}_1, \bar{R}_2)$ as defined in (4.11). Since S is positive definite with probability 1 (as $N \geqslant p \geqslant 2$ by assumption): $\bar{R}_1 > 0$, $\bar{R}_2 > 0$ and $\bar{R}_1 + \bar{R}_2 = R_2$, the squared sample multiple correlation coefficent between the first and the remaining $p-1$ components of the random vector X. A corresponding maximal invariant in the parametric space under G is (ρ_1^2, ρ_2^2). The joint probability density function of $(\bar{R}_1, \bar{R}_2)$ is given by (see Giri (1979))

$$\begin{aligned} f_{\Delta}(\bar{r}_1, \bar{r}_2) &= K(1-\rho^2)^{-N/2}(1 - \bar{r}_1 - \bar{r}_2)^{\frac{1}{2}(N-p-1)}\prod_{i=1}^{2}(\bar{r}_i)^{\frac{1}{2}p_i - 1} \\ &\times \left[1 + \sum_{i=1}^{2}\bar{r}_i\left(\frac{1-\rho^2}{\bar{\gamma}_i} - 1\right)\right]^{-N/2} \\ &\times \sum_{\beta_1=0}^{\infty}\sum_{\beta_2=0}^{\infty}\prod_{i=1}^{2}\frac{\Gamma(\frac{1}{2}(N + p_i - \sigma_i) + \beta_i)\Gamma(\beta_i + \frac{1}{2})}{(2\beta_i)!\Gamma(\frac{1}{2}p_i + \beta_i)}(\bar{\theta}_i)^{\beta_i}, \end{aligned} \tag{4.37}$$

where

$$\bar{\gamma}_i = 1 - \sum_{j=1}^{i} \rho_j^2 \quad \text{with } \gamma_0 = 1\,,$$

$$\sigma_i = \sum_{j=1}^{i} p_j, \quad \bar{\alpha}_i^2 = \rho_i^2(1-\rho^2)/\bar{\gamma}_i\bar{\gamma}_{i-1}\,,$$

$$\bar{\theta}_i = 4\bar{r}_i\bar{\alpha}_i^2\left(1 + \sum_{i=1}^{2} \bar{r}_i\left[\frac{1-\rho^2}{\bar{\gamma}_i} - 1\right]\right)^{-1},$$

and K is the normalizing constant.

By straightforward computations the likelihood ratio test of H_{10} when $\Omega = \{(\mu, \Sigma) : \Sigma_{13} = 0\}$ rejects H_{10} whenever

$$\bar{r}_1 \geqslant c\,, \tag{4.38}$$

where the constant c depends on the size α of the test and under H_{10} $\bar{R}_1$ has a central beta distribution with parameter $(\frac{1}{2}p_1, \frac{1}{2}(N-p_1))$, and the likelihood ratio test of H_{20} when $\Omega = \{(\mu, \Sigma) : \Sigma_{12} = 0\}$ rejects H_{20} whenever

$$z = \frac{1-\bar{r}_1-\bar{r}_2}{1-\bar{r}_1} \leqslant c\,, \tag{4.39}$$

where the constant c depends on the size α of the test and under H_{20} the corresponding random variable Z is distributed independently of $\bar{R}_1$ as central beta with parameter $(\frac{1}{2}(N-p_1-p_2), \frac{1}{2}p_2)$.

Theorem 4.4.1. *For problem 7 the likelihood ratio test given in* (4.38) *is UMP invariant.*

Proof. Under H_{10} $\bar{\gamma}_i = 1$, $i = 0, 1, 2$. Hence $\bar{\alpha}_i^2 = 0$, $\bar{\theta}_i = 0$, $i = 1, 2$. Under $H_{1\lambda}$ $\rho_1^2 = \lambda$, $\rho_2^2 = 0$, $\bar{\gamma}_0 = 1$, $\bar{\gamma}_1 = 1 - \lambda\bar{\alpha}_1^2 = 1 - \lambda$, $\bar{\gamma}_2 = 1 - \lambda$, $\bar{\alpha}_2^2 = 0$, $\bar{\theta}_1 = 4\bar{r}_1\lambda$ and $\bar{\theta}_2 = 0$. Thus

$$\frac{f_{H_{1\lambda}}(\bar{r}_1, \bar{r}_2)}{f_{H_{10}}(\bar{r}_1, \bar{r}_2)} = K(1-\lambda)^{-N/2} \times \sum_{i=0}^{\infty} \frac{\Gamma(\frac{1}{2}N + i)(4\bar{r}_1\lambda)^i}{(2i)!}\,. \tag{4.40}$$

Using the Neyman-Pearson Lemma and (4.40) we get the theorem. □

Theorem 4.4.2. *For Problem 8 the likelihood ratio test is UMP invariant among all tests $\phi(\bar{R}_1, \bar{R}_2)$ based on $\bar{R}_1, \bar{R}_2$ satisfying*

$$E_{H_{20}}(\phi(\bar{R}_1, \bar{R}_2)|\bar{R}_1 = \bar{r}_1) = \alpha\,.$$

Proof. Under $H_{2\lambda}$ $\rho_1^2 = 0$, $\rho_2^2 = \lambda$, $\bar{\gamma}_0 = 1$, $\bar{\gamma}_1 = 1$, $\bar{\gamma}_2 = 1 - \lambda$, $\bar{\alpha}_1^2 = 0$, $\bar{\alpha}_2^2 = \lambda$, $\bar{\theta}_1 = 0$ and $\bar{\theta}_2 = 4\bar{r}_2\lambda(1 - \bar{r}_1\lambda)^{-1}$.

Hence

$$\begin{aligned} f_{H_{2\lambda}}(\bar{r}_2|\bar{r}_1)/f_{H_{20}}(\bar{r}_2|\bar{r}_1) &= f_{H_{2\lambda}}(\bar{r}_1, \bar{r}_2)/f_{H_{20}}(\bar{r}_1, \bar{r}_2) \\ &= K(1-\lambda)^{-N/2}(1 - \bar{r}_1\lambda)^{-N/2} \\ &\quad \times \sum_{i=0}^{\infty} \frac{\Gamma(\frac{1}{2}(N - p_1) + i)}{(2i)!} \left(\frac{4\bar{r}_2\lambda}{1 - \bar{r}_1\lambda} \right)^i . \end{aligned} \tag{4.41}$$

From (4.41) $f_{H_{2\lambda}}(\bar{r}_2|\bar{r}_1)$ has a monotone likelihood ratio in $\bar{r}_2 = (1-z)(1-\bar{r}_1)$. Now using Lehmann (1959) we get the theorem. □

Exercises

1. Prove Equations (4.3) and (4.5).
2. Prove (4.10).
3. Show that $\chi_m^2(\lambda)$ is distributed as χ_{m+2K}^2 where K is a Poisson random variable with parameter $\frac{1}{2}\lambda$.
4. Prove (4.31).
5. Let π_1, π_2 be two p-variate normal populations with means μ_1 and μ_2 and the same positive definite covariance matrix Σ. Let $X = (X_1, \ldots, X_p)'$ be distributed according to π_1 or π_2 and let $b = (b_1, \ldots, b_p)'$ be a real vector. Show that
$$\frac{[E_1(b'X) - E_2(b'X)]^2}{\mathrm{var}(b'X)}$$
is maximum for all b if $b = \Sigma^{-1}(\mu_1 - \mu_2)$ where E_i denote the expectation under π_i.
6. (Giri, 1994*a*). Let $X_\alpha = (X_{\alpha 1}, \ldots, X_{\alpha p})'$, $\alpha = 1, \ldots, N(> p)$ be independently identically distributed $p = 2p_1$-variate normal random vectors with mean $\mu = (\mu_1, \ldots, \mu_p)'$ and common covariance matrix Σ (positive definite). Let $\bar{X} = \frac{1}{N}\Sigma_{\alpha=1}^N X_\alpha$, $S = \Sigma_{\alpha=1}^N (X_\alpha - \bar{X})(X_\alpha - \bar{X})'$. Write

$$\Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} \\ \Sigma_{(21)} & \Sigma_{(22)} \end{pmatrix}, \quad S = \begin{pmatrix} S_{(11)} & S_{(12)} \\ S_{(21)} & S_{(22)} \end{pmatrix},$$
$$\bar{X} = (\bar{X}_1, \ldots \bar{X}_p) = (\bar{X}'_{(1)}, \bar{X}'_{(2)})',$$

where $\Sigma_{(ij)}$ and $S_{(ij)}$ are $p_1 \times p_1$ for all i, j and $\bar{X}_{(1)} = (\bar{X}_1, \ldots, \bar{X}_{p_1})'$

(a) Show that the likelihood ratio test of $H_0 : \mu_{(1)} = \mu_{(2)}$, with $\mu = (\mu'_{(1)}, \mu'_{(2)})'$, $\mu_{(1)} = (\mu_1, \ldots, \mu_{p_1})'$ rejects H_0 for large values of

$$T^2 = N(\bar{X}_{(1)} - \bar{X}_{(2)})'(S_{(11)} + S_{(22)} - S_{(12)} - S_{(21)})^{-1}(\bar{X}_{(1)} - \bar{X}_{(2)})$$

where T^2 is distributed as $\chi^2_{p_1}(\delta^2)/\chi^2_{N-p_1}$ with $\delta^2 = N\big(\mu_{(1)} - \mu_{(2)}\big)' \cdot \big(\Sigma_{(11)} + \Sigma_{(22)} - \Sigma_{(21)} - \Sigma_{(12)}\big)^{-1}\big(\mu_{(1)} - \mu_{(2)}\big)$ and $\chi^2_{p_1}(\cdot), \chi^2_{N-p_1}$ are independent.

(b) Show that the above test is UMPI.

7. (Giri, 1994*b*). In problem 5 let $\Gamma = \Sigma^{-1}\mu = (\Gamma_1, \ldots, \Gamma_{2p_1})' = (\Gamma'_{(1)}, \Gamma'_{(2)})'$ with $\Gamma_{(1)} = (\Gamma_1, \ldots, \Gamma_{p_1})'$.

(a) Show that the likelihood ratio test of $H_0 : \Gamma_{(1)} = \Gamma_{(2)}$ rejects H_0 for small values of Z

$$Z = \frac{1 + N(\bar{X}_{(1)} + \bar{X}_{(2)})'(S_{(11)} + S_{(22)} + S_{(21)} + S_{(12)})^{-1}(\bar{X}_{(1)} + \bar{X}_{(2)})}{1 + N\bar{X}'S^{-1}\bar{X}}$$

where Z is distributed as central beta with parameter $(\frac{1}{2}(N - p_1), \frac{1}{2}p_1)$ under H_0.

(b) Show that it is UMPI similar.

(c) Find the likelihood ratio test of $H_0 : \Gamma_{(1)} = \lambda\Gamma_{(2)}$ when λ is known.

8. (Giri, 1994*c*). In problem 6 write

$$\Sigma = \begin{pmatrix} \Sigma_{(11)} & \Sigma_{(12)} & \Sigma_{(13)} \\ \Sigma_{(21)} & \Sigma_{(22)} & \Sigma_{(23)} \\ \Sigma_{(31)} & \Sigma_{(33)} & \Sigma_{(32)} \end{pmatrix}, \quad S = \begin{pmatrix} S_{(11)} & S_{(12)} & S_{(13)} \\ S_{(21)} & S_{(22)} & S_{(23)} \\ S_{(31)} & S_{(32)} & S_{(33)} \end{pmatrix}$$

where $\Sigma_{(11)}$ and $S_{(11)}$ are 1×1, $\Sigma_{(22)}$ and $S_{(22)}$ are $p_1 \times p_1$ and $\Sigma_{(33)}$ and $S_{(33)}$ are $p_1 \times p_1$ with $2p_1 = p - 1$.

(a) Show that the likelihood ratio test of $H_0 : \Sigma_{(12)} = \Sigma_{(13)}$ rejects H_0 for large values of

$$\bar{R}_1^2 = (S_{(12)} - S_{(13)})(S_{(22)} + S_{(33)} - S_{(32)} - S_{(23)})^{-1}(S_{(21)} - S_{(31)})/S_{(11)},$$

the probability density function of $\bar{R}_1^2$ is given by

$$f(\bar{r}_1^2) = \frac{(1-\bar{\rho}_1^2)^{\frac{1}{2}(N-1)}(1-\bar{r}_1^2)^{\frac{1}{2}(N-p_1-3)}}{\Gamma(\frac{1}{2}(N-1))\Gamma(\frac{1}{2}(N-p_1))} \times \sum_{j=0}^{\infty} \frac{(\bar{\rho}_1^2)^j(\bar{r}_1^2)^{\frac{1}{2}p_1+j-1)}\Gamma^2(\frac{1}{2}(N-1)+j)}{j!\Gamma(\frac{1}{2}p_1+j)}$$

where

$$\bar{\rho}_1^2 = N(\Sigma_{(12)}-\Sigma_{(13)})(\Sigma_{(22)}+\Sigma_{(33)}-\Sigma_{(32)}-\Sigma_{(23)})^{-1} \times (\Sigma_{(21)}-\Sigma_{(31)})/\Sigma_{(11)}\,.$$

(b) Show that no optimum invariant test exists for this problem.

References

1. N. Giri, *Locally Minimax Tests of Multiple Correlations*, Canad. J. Statist. **7**, 53–60 (1979).
2. N. Giri, *Multivariate Statistical Inference*, Academic Press, N.Y., 1977.
3. N. Giri, *On the Likelihood Ratio Test of a Normal Multivariate Testing Problem*, Ann. Math. Statist. **35**, 181–189 (1964).
4. N. Giri, *On a Multivariate Testing Problem*, Calcutta Stat. Assoc. Bull. **11**, 55–60 (1962).
5. N. Giri, *On the Likelihood Ratio Test of A Multivariate Testing Problem II*, Ann. Math. Statist. **36**, 1061–1065 (1965).
6. N. Giri, *Admissibility of Some Minimax Tests of Muttivariate Mean*, Rapport de Research, Mathematics and Statistics, University of Montreal, 1994a.
7. N. Giri, *On a Test of Discriminant Coefficients*, Rapport de Research, Mathematics and Statistics, University of Montreal, 1994b.
8. N. Giri, *On Tests of Multiple Correlations with Partial Informations*, Rapport de Research, Mathematics and Statistics, Univerity of Montreal, 1994c.
9. N. Giri, *Multivariate Statistical Analysis*, Marcel Dekker, N. Y., 1996.
10. J. K. Ghosh, *Invariance in Testing and Estimation*, Tech. report no Math-Stat 2/672. Indian Statistical Institute, Calcutta, India, 1967.
11. E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, N.Y., 1959.
12. A. Wald, *Tests of Statistical Hypotheses Concerning Parameters when the Number of Observations is Large*, Trans. Amer. Math. Soc., **54**, 426–482 (1959).

Chapter 5

SOME MINIMAX TESTS IN MULTINORMALES

5.0. Introduction

The invariance principle, restricting its attention to invariant tests only, allows us to consider a subclass of the class of all available tests. Naturally a question arises, under what conditions, an optimum invariant test is also optimum among the class of all tests if such can at all be achieved. A powerful support for this comes from the celebrated unpublished work of Hunt and Stein, popularly known as the Hunt-Stein theorem, who towards the end of Second World War proved that under certain conditions on the transformation group G, there exists an invariant test of level α which is also minimax, i.e. minimizes the maximum error of second kind (1-power) among all tests. Though many proofs of this theorem have now appeared in the literature, the version of this theorem which appeared in Lehmann (1959) is probably close in spirit to that originally developed by Hunt and Stein. Pittman (1939) gave intuitive reasons for the use of best invariant procedure in hypothesis testing problems concerning location and scale parameters. Wald (1939) had the idea that for certain nonsequential location parameter estimation problems under certain restrictions on the group there exists an invariant estimator which is minimax. Peisakoff (1950) in his Ph.D. thesis pointed out that there seems to be a locuna in Wald's proof and he gave a general development of the theory of minimax decision procedures invariant under transformation group. Kiefer (1957) proved an analogue of the Hunt-Stein theorem for the continuous and discrete sequential decision problems and extended this theorem to other decision problems. Wesler (1959) generalized for modified minimax tests based on slices of the parametric space.

It is well-known that for statistical inference problems we can, without any loss of generality, characterize statistical tests as functions of sufficient statistic instead of sample observations. Such a characterization introduces considerable simplifications to the sample space without loosing any information concerning the problem at hand. Though such a characterization in terms of maximal invariant is too strong a result to expect, the Hunt-Stein theorem has made considerable contribution towards that direction. The Hunt-Stein theorem gives conditions on the transformation groups such that given any test ϕ for the problem of testing $H : \theta \in \Omega_H$ against the alternatives $K : \theta \in \Omega_K$, with $\Omega_H \cap \Omega_K$ a null set, there exists an invariant test ψ such that

$$\sup_{\theta \in \Omega_H} E_\theta \phi \geq \sup_{\theta \in \Omega_H} E_\theta \psi \,, \tag{5.1}$$

$$\inf_{\theta \in \Omega_K} E_\theta \phi \leq \inf_{\theta \in \Omega_K} E_\theta \psi \,. \tag{5.2}$$

In other words, ψ behaves at least as good as any ϕ in the worst possible cases. We shall present only the statements of this theorem. For a detailed discussion and a proof the reader is referred to Lehmann (1959) or Ghosh (1967).

Let $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ be a dominated family of distributions on the Euclidean space $(\mathfrak{X}, \mathcal{A})$, dominated by a σ-finite measure μ. Let G be the group of transformations, operating from the left on $\mathfrak{X}$, leave Ω invariant.

Theorem 5.0.1. (*Hunt-Stein Theorem*). *Let $\mathcal{B}$ be a σ-field of subsets of G such that for any $A \in \mathcal{A}$, the set of pairs (x, g) with $gx \in A$ is in $\mathcal{A} \times \mathcal{B}$ and for any $B \in \mathcal{B}$, $g \in G$, $Bg \in \mathcal{B}$. Suppose that there exists a sequence of distribution functions ν_n on $(G, \mathcal{B})$ which is asymptotically right invariant in the sense that for any $g \in G$, $B \in \mathcal{B}$*

$$\lim_{n \to \infty} |\nu_n(Bg) - \nu_n(B)| = 0 \,. \tag{5.3}$$

Then, given any test ϕ, there exists a test ψ which is almost invariant and satisfies conditions (5.1) and (5.2).

It is a remarkable feature of this theorem that its assumptions have nothing to do with the statistical aspects of the problem and they involve only the group G. However, for the problem of admissibility of statistical tests the situation is more complicated. If G is a finite or a locally compact group the best invariant test is admissible. For other groups the nature of $\mathcal{P}$ plays a dominant role.

The proof of Theorem 5.0.1 is straightforward if G is a finite group. Let m denote the number of elements of G. We define

$$\psi(x) = \frac{1}{m} \sum_{g \in G} \phi(gx) .$$

As observed in Chapter 2, invariant measures exist for many groups and they are essentially unique. But frequently they are not finite and as a result they cannot be taken as a probability measure.

We have shown in Chapter 2 that on the group $O(p)$ of orthogonal matrices of order p an invariant probability measure exists and this group satisfies the conditions of the Hunt-Stein theorem. The group $G_T(p)$ of nonsingular lower triangular matrices of order p also satisfies the conditions of this theorem (see Lehmann, 1959, p. 345).

5.1. Locally Minimax Tests

Let $(\mathfrak{X}, \mathcal{A})$ be a measurable space. For each point (δ, η) in the parametric space Ω, where $\delta \geq 0$ and η is of arbitrary dimension and its range may depend on δ, suppose that $p(\cdot\,; \delta, \eta)$ is a probability density function on $(\mathfrak{X}, \mathcal{A})$ with respect to some σ-finite measure μ. We are interested in testing at level α $(0 < \alpha < 1)$ the hypothesis $H_0 : \delta = 0$ against the alternative $H_1 : \delta = \lambda$, where λ is a positive specified canstant and in giving a sufficient condition for a test to be locally minimax in the sense of (5.7) below. This is a local theory in sense that $p(x; \lambda, \eta)$ is close to $p(x; o, \eta)$ when λ is small. Obviously, then, every test of level α would be locally minimax in the sense of trivial criterion obtained by not substracting α in the numerator and the denominator of (5.7).

It may be remarked that our method of proof of (5.7) consists merely of considering local power behaviour with sufficient accuracy to obtain an approximate version of the classical result that a Bayes procedure with constant risk is minimax. A result of this type can be proved under various possible types of conditions of which we choose a form which is more convenient in many applications and stating other possible generalizations and simplifications as remarks. Throughout this section expressions like $o(\lambda)$, $o(h(\lambda))$ are to be interpreted as $\lambda \to 0$.

For fixed α, $0 < \alpha < 1$ we shall consider critical regions of the form

$$R = \{x : \; V(x) \geq c_\alpha\} \tag{5.4}$$

where V is bounded and positive and has a continuous distribution function for each (δ, η), equicontinuous in (δ, η) for $\delta <$ some δ_0 and which satisfies

$$\begin{aligned} P_{0,\eta}(R) &= \alpha\,, \\ P_{\lambda,\eta}(R) &= \alpha + h(\lambda) + q(\lambda,\eta) \end{aligned} \tag{5.5}$$

where $q(\lambda,\eta) = o(h(\lambda))$ uniformly in η with $h(\lambda) > 0$ for $\lambda > 0$ and $h(\lambda) = o(1)$. We shall be concerned with *a priori* probability density functions $\xi_{0,\lambda}$ and $\xi_{1,\lambda}$ on sets $\delta = 0$, $\delta = \lambda$ respectively, for which

$$\frac{\int p(x;\lambda,\eta)\xi_{1,\lambda}(d\eta)}{\int p(x;0,\eta)\xi_{0,\lambda}(d\eta)} = 1 + h(\lambda)[g(\lambda) + r(\lambda)V(x))] + B(x,\lambda) \tag{5.6}$$

where $0 < c_1 < r(\lambda) < c_2 < \infty$ for λ sufficiently small and $g(\lambda) = o(1)$, $B(x,\lambda) = o(h(\lambda))$ uniformly in x.

Theorem 5.1.1. (*Locally Minimax Tests*) *If the region R satisfies* (5.5) *and for sufficiently small λ there exist $\xi_{0,\lambda}$ and $\xi_{1,\lambda}$ satisfying* (5.6) *then R is locally minimax of level α for testing $H_0 : \delta = 0$ against $H_1 : \delta = \lambda$ as $\lambda \to 0$, that is to say,*

$$\lim_{\lambda\to 0} \frac{\inf_\eta\ P_{\lambda,\eta}(R) - \alpha}{\sup_{\phi_\lambda \in Q_\alpha} \inf_\eta P_{\lambda,\eta}\{\phi_\lambda \text{ rejects } H_0\} - \alpha} = 1\,, \tag{5.7}$$

where Q_α is the class of all level α tests ϕ_λ.

Proof. Let

$$\tau_\lambda = [2 + h(\lambda)\{g(\lambda) + c_\alpha\, r(\lambda)\}]^{-1}\,.$$

Then

$$\frac{1-\tau_\lambda}{\tau_\lambda} = 1 + h(\lambda)\ [g(\lambda) + c_\alpha\, r(\lambda)]\,. \tag{5.8}$$

Using (5.7) and (5.8) the Bayes critical region relative to the *a priori* distribution

$$\xi_\lambda = (1-\tau_\lambda)\xi_{0,\lambda} + \tau_\lambda\ \xi_{1,\lambda}$$

and $(0-1)$ loss is given by

$$B_\lambda = \left\{x :\ U(x) + \frac{B(x,\lambda)}{r(\lambda)\ h(\lambda)} > c_\alpha\right\}. \tag{5.9}$$

Define for any subset A

$$P^*_{0,\lambda}(A) = \int_A P_{0,\eta}(A)\; \xi_{0,\lambda}(d\eta)\,,$$

$$P^*_{1,\lambda}(A) = \int_A P_{\lambda,\eta}(A)\; \xi_{1,\lambda}(d\eta)\,,$$

$$V_\lambda = R - B_\lambda\,,$$

$$W_\lambda = B_\lambda - R\,. \tag{5.10}$$

Using the fact that

$$\sup_x \left| \frac{B(x,\lambda)}{h(\lambda)} \right| = o(\lambda)$$

and the continuity assumption on the distribution function of U we get

$$P^*_{0,\lambda}(V_\lambda + W_\lambda) = o(\lambda)\,. \tag{5.11}$$

Also for $U_\lambda = V_\lambda$ or W_λ

$$P^*_{1,\lambda}(U_\lambda) = P^*_{0,\lambda}(U_\lambda)[1 + o(h(\lambda))]\,. \tag{5.12}$$

Let

$$r^*_\lambda(A) = (1-\tau_\lambda)\; P^*_{0,\lambda}(A) + \tau_\lambda(1 - P^*_{1,\lambda}(A))\,.$$

Using (5.9), (5.10) and (5.11) the integrated Bayes risk relative to ξ_λ is given by

$$\begin{aligned} r^*_\lambda(B_\lambda) &= r^*_\lambda(R) + (1-\tau_\lambda)[P^*_{0,\lambda}(W_\lambda) - P^*_{0,\lambda}(V_\lambda)] \\ &\quad + \tau_\lambda[P^*_{1,\lambda}(V_\lambda) - P^*_{1,\lambda}(W_\lambda)] \\ &= r^*_\lambda(R) + (1-2\tau_\lambda)[P^*_{0,\lambda}(W_\lambda) - P^*_{0,\lambda}(V_\lambda)] \\ &\quad + P^*_{0,\lambda}(V_\lambda + W_\lambda)\; o(h(\lambda)) \\ &= r^*_\lambda(R) + o(h(\lambda))\,. \end{aligned} \tag{5.13}$$

If (5.7) were false one could by (5.5) find a family of tests $\{\phi_\lambda\}$ of level α such that ϕ_λ has power function $\alpha + g(\lambda,\eta)$ on the set $\delta = \lambda$ with

$$\limsup_{\lambda\to 0}[\inf_\eta g(\lambda,\eta) - h(\lambda)]/h(\lambda) > 0\,.$$

The integrated risk r'_λ of ϕ_λ with respect to ξ_λ would then satisfy

$$\limsup_{\lambda\to 0}[r^*_\lambda(R) - r'_\lambda]/h(\lambda) > 0\,,$$

contradicting (5.13). □

Remarks.

(1) Let the set $\{\delta = 0\}$ be a single point and the set $\{\delta = \lambda\}$ be a convex finite dimensional Euclidean set where in each component η_i of η is $0(h(\lambda))$. If

$$\frac{p(x;\lambda,\eta)}{p(x;0,\eta)} = 1 + h(\lambda)\ U(x) + \sum_{i,j=1}^{k} s_i(x)\ a_{ij}(\lambda)\ \eta_j + B(x,\lambda,\eta) \quad (5.14)$$

where $s_i,\ a_{ij}$ are bounded and $\sup_{x,\eta}\ B(x,\lambda,\eta) = o(h(\lambda))$, and if there exists any $\xi_{1,\lambda}$ satisfying (5.6), then the degenerate measure $\xi^0_{1,\lambda}$ which assigns all its measure to the mean of $\xi_{1,\lambda}$ also satisfies (5.6).

(2) The assumption on B can be weakened to $P_{\lambda,\eta}\{|B(x,\lambda)| <\in h(\lambda)\} \to 0$ as $\lambda \to 0$ uniformly in η for each $\in > 0$. If the $\xi_{i,\lambda}{}'s$ are independent of λ the uniformity of the last condition is unnecessary. The boundedness of U and the equicontinuity of the distribution of U can be similarly weakened.

(3) The conclusion of Theorem 5.1.1 also holds if Q_α is modified to include every family $\{\phi_\lambda\}$ of tests of level $\alpha + o(h(\lambda))$. One can similarly consider the optimality of the family $\{U_\lambda\}$ rather than single U by replacing R by $R_\lambda = \{x : U_\lambda(x) \geq c_{\alpha,\lambda}\}$, where $P_{0,\eta}\{R_\lambda\} = \alpha - q_\lambda(\eta)$ with $q_\lambda(\eta) = o(h(\lambda))$.

(4) We can refine (5.7) by including one or more error terms. Specifically one may be interested to know if a level α critical region R which satisfies (5.7) with

$$\inf_\eta P_{\lambda,\eta}(R) = \alpha + c_1\lambda + o(\lambda), \quad \text{as } \lambda \to 0$$

also satisfies

$$\lim_{\lambda\to 0} \frac{\inf_\eta P_{\lambda,\eta}(R) - \alpha - c_1\lambda}{\sup_{\phi \in Q_\alpha} \inf_\eta P_{\lambda,\eta}\{\phi \text{ rejects } H_0\} - \alpha - c_1\lambda} = 1\,. \quad (5.15)$$

In the setting of (5.14) this involves two moments of the *a priori* distribution $\xi_{1,\lambda}$ rather than just one in Theorem 5.1.1. As further refinements are invoked more moments are brought in.

The theory of locally minimax test as developed above and the theory of asymptotically minimax test (far in distance from the null hypothesis) to be developed later in this chapter serve two purposes. First the obvious point of demonstrating such properties for their own sake. But well known valid doubts have been raised as to meaningfulness of such properties. Secondly, then, and in our opinion more important, these properties can give an indication of what to look for in the way of genuine minimax or admissibility property of certain tests, even though the later do not follow from the local or the asymptotic properties.

Example 5.1.1. (T^2-test). Consider Problem 1 of Chapter 4. Let $x_1, \ldots, x_N$ be independently and identically distributed $N_p(\mu, \Sigma)$ random vectors. Write $N\bar{X} = \sum_{i=1}^N X_i$, $S = \sum_1^N (X_i - \bar{X})(X_i - \bar{X})'$. Let $\delta = N\mu'\Sigma^{-1}\mu > 0$. For testing $H_0 : \delta = 0$ against the alternatives $H_1 : \delta > 0$ the Hotelling's T^2 test which rejects H_0 whenever $R = N\bar{X}'(S + N\bar{X}\bar{X}')^{-1}\bar{X} \geq c$, where c is chosen to yield level α, is UMPI (Theorem 4.1.1).

Note. For notational convenience we are writing R_1 as R.

Let $\delta = \lambda > 0$ (specified). We are concerned here to find the locally minimax test of H_0 against $H_1 : \delta = \lambda$ as $\lambda \to 0$ in the sense of (5.7). We assume that $N > p$, since it is easily shown that the denominator of (5.7) is zero in the degenerate case $N \leq p$. In our search for locally minimax test as $\lambda \to 0$ we may restrict attention to the space of sufficient statistic $(\bar{X}, S)$. The general linear group $G_l(p)$ of nousingular matrix of order p operating as

$$(\bar{x}, s; \mu, \Sigma) \longrightarrow (g\bar{x}, gsg'; g\mu,\ g\Sigma g')$$

leaves this problem invariant. However, as discussed earlier (see also James and Stein, 1960, p. 376), the Hunt-Stein theorem cannot be applied to the group $G_l(p)$, $p \geq 2$. However this theorem does apply to the subgroup $G_T(p)$ of nonsingular lower triangular matrices of order p. Thus, for each λ, there is a level α test which is almost invariant and hence for this problem which is invariant under $G_T(p)$ (see Lehmann, 1959, p. 225) and which minimizes, among all level α tests, the minimum power under H_1. From the local point of view, the denominator of (5.7) remains unchanged by the restriction to G_T invariant tests and for any level α test ϕ there is a G_T invariant level α test ϕ' for which the expression $\inf_\eta\ P_{\lambda,\eta}$ (ϕ' rejects H_0) is at least as large, so that a procedure which is locally minimax among G_T invariant level α tests, is locally minimax among all level α tests.

In the place of one-dimensional maximal invariant R under $G_l(p)$ we obtain a p-dimensional maximal invariant $(R_1, \ldots, R_p)$ as defined in Theorem 2.8.1 with $k = p$ and $d_i = 1$ for all i, satisfying

$$\sum_{j=1}^{i} R_j = N\bar{X}'_{[i]}(S_{[ii]} + N\bar{X}_{[i]}\bar{X}'_{[i]})^{-1}\bar{X}_{[i]}\,, \quad i = 1, \ldots, p \tag{5.16}$$

with $R_i \geq 0$, $\sum_{i=1}^{p} R_j = R \leq 1$ and the corresponding maximal invariant on the parametric space of (μ, Σ) is $(\delta_1, \ldots, \delta_p)$ (Theorem 2.8.1) where

$$\sum_{j=1}^{i} \delta_j = N\ \xi'_{[i]}\ \Sigma_{[ii]}^{-1}\xi_{[i]}, \quad i = 1, \ldots, p \tag{5.17}$$

with $\sum_{j=1}^{p} \delta_j = \delta$. The nuisance parameter in this reduced setup is $\eta = (\eta_1, \ldots, \eta_p)'$, with $\eta_i = \delta_i/\delta \geq 0$, $\sum_{j=1}^{p} \eta_j = 1$ and under H_0 $\eta = 0$. From Theorem 2.8.1 the Lebesgue density of $R_1, \ldots, R_p$ on the set

$$\left\{(r_1, \ldots, r_p) : r_i > 0,\ \sum_{1}^{p} r_j < 1\right\}$$

is given by

$$\begin{aligned} f(r_1, \ldots, r_p) = \frac{\pi^{-\frac{p}{2}}\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-p}{2}\right)} \prod_{1}^{p} r_i^{-\frac{1}{2}} \left(1 - \sum_{1}^{p} r_j\right)^{\frac{1}{2}(N-p-1)} \\ \times \exp\left\{-\frac{1}{2}\delta + \frac{1}{2}\sum_{j=1}^{p} r_j \sum_{i>j} \delta_i\right\} \\ \times \prod_{i=1}^{p} \phi\left(\frac{N-i+1}{2}, \frac{1}{2}; \frac{r_i\delta_i}{2}\right). \end{aligned} \tag{5.18}$$

We now verify the validity of Theorem 2.8.1 for $U = \sum_{j=1}^{p} R_j$, since those preceding (5.5) for the locally minimax tests are obvious. In (5.5) we take $h(\lambda) = b\lambda$ with b a positive constant. Of course $P_{\lambda,\eta}(R)$ does not depend on η. From (5.18) we get with $r = (r_1, \ldots r_p)'$,

$$\frac{f_{\lambda,\eta}(r)}{f_{0,\eta}(r)} = 1 + \frac{\lambda}{2}\left[-1 + \sum_{j=1}^{p} r_j \left(\sum_{i>j} \eta_i + (N-j+1)\eta_j\right)\right] + B(r, \lambda, \eta) \tag{5.19}$$

where $B(r, \lambda, \eta) = o(\lambda)$ uniformly in r, η as $\lambda \to 0$. The equation (5.6) is satisfied by letting $\xi_{0,\lambda}$ give measure one to the single point $\eta = 0$ while $\xi_{1,\lambda}$ gives measure one to the single point $\eta = \eta^* = (\eta_1^*, \ldots, \eta_p^*)'$ where

$$\eta_j^* = (N-j)^{-1}(N-j+1)^{-1}p^{-1}N(N-p)\,, \quad j = 1, \ldots, p$$

so that $\sum_{i>j} \eta_i^* + (N-j+1)\eta_j^* = \frac{N}{p}$ for all j. From Theorem 5.1.1 we get the following theorem.

Theorem 5.1.2. *For every $p, N, \alpha (N > p)$ Hotelling's T^2 test based on T^2 or equivalently on R is locally minimax for testing $H_0 : \delta = 0$ against the alternatives $H_1 : \delta = \lambda$ as $\lambda \to 0$.*

Example 5.1.3. Consider Problem 2 of Chapter 4. From Example 5.1.2 it follows that the maximal invariant in the sample space is $(R_1, \ldots, R_{p_1})$ with $\bar{R}_1 = \sum_1^{p_1} R_j$ and the corresponding maximal invariant in the parametric space is $(\delta_1, \ldots, \delta_{p_1})$ with $\bar{\delta}_1 = \sum_1^{p_1} \delta_j$. Under H_0, $\bar{\delta}_1 = 0$ and under H_1 $\bar{\delta}_1 = \lambda$. Now following Example 5.1.2 we prove Theorem 5.1.3.

Theorem 5.1.3. *For every p_1, N, α the UMPI test which rejects H_0 for large values of $\bar{R}_1$ is locally minimax for testing $H_0 : \bar{\delta}_1 = 0$ against the alternative $H_1 : \bar{\delta}_1 = \lambda$ as $\lambda \to 0$.*

Example 5.1.4. Consider Problem 3 of Chapter 4. We are concerned here to find the locally minimax test of $H_0 : \bar{\delta}_1 = 0$, $\bar{\delta}_2 = 0$ against the alternatives $H_1 : \bar{\delta}_1 = 0$, $\bar{\delta}_2 = \lambda > 0$. As usual we assume that $N > p$ the determinator of (5.7) is not zero. To find the locally minimax test we restrict attention to the space of sufficient statistic $(\bar{X}_1\ S)$. The group (G_2, T_2) operating as $(\bar{X}, S) \to (g\bar{X} + t_2,\ gSg')$, $g \in G_2$, $t_2 \in T_2$, which leaves the problem invariant, does not satisfy the condition of the Hunt-Stein theorem. However this theorem holds for the subgroup $(G_T(p_1+p_2),\ T_2)$. A maximal invariant in the sample space under this subgroup is $R_1, \ldots, R_{p_1+p_2}$ such that

$$\sum_{j=1}^{i} R_j = N\bar{X}_{[i]}' \ (S_{[ii]} + N\bar{X}_{[i]}\ \bar{X}_{[i]}')^{-1}\bar{X}_{[i]}\,,\ \ i = 1, \ldots, p_1 + p_2 \tag{5.20}$$

with $R_i \geq 0$, $\bar{R}_1 = \sum_{j=1}^{p_1} R_j$, $\bar{R}_1 + \bar{R}_2 = \sum_{j=1}^{p_1+p_2} R_j$, and the corresponding maximal invariant in the parametric space is $\delta_1, \ldots, \delta_{p_1+p_2}$ such that

$$\sum_{j=1}^{i} \delta_j = N\xi_{[i]}'\Sigma_{[ii]}^{-1}\xi_{[i]}, \quad i = 1, \ldots, p_1 + p_2 \tag{5.21}$$

with $\delta_i \geq 0$, $\bar{\delta} = \sum_{j=1}^{p_1} \delta_j$, $\bar{\delta}_1 + \bar{\delta}_2 = \sum_{j=1}^{p_1+p_2} \delta_j$. Under H_0 $\delta_i = 0$ for all i and under H_1 $\delta_i = 0, i = 1, \ldots, p_1$, $\bar{\delta}_2 = \delta = \lambda > 0$. The nuisance parameter in this reduced set up is $\eta = (\eta_1, \ldots, \eta_{p_1+p_2})'$ with $\eta_i = \frac{\delta_i}{\delta}$. Under H_0 $\eta = 0$ and under H_1, $\eta_i = 0$ $i = 1, \ldots, p_1$, $\eta_i > 0$, $i = p_1 + 1, \ldots, p_1 + p_2$ with $\sum_{i=p_1+1}^{p_1+p_2} \eta_i = 1$. Equation (5.19) in this case reduces to

$$\frac{f_{\lambda,\eta}(r)}{f_{0,\eta}(r)} = 1 + \frac{\lambda}{2}\left[-1 + \bar{r}_1 + \sum_{j=p_1+1}^{p_1+p_2} r_j\left(\sum_{i>j}\eta_i + (N-j+1)\eta_j\right)\right] + B(r,\eta,\lambda) \tag{5.22}$$

as $\lambda \to 0$ with $B(r,\eta,\lambda) = o(\lambda)$ uniformly in r, η.

The set $\bar{\delta}_2 = 0$, $\bar{\delta}_1 = 0$ is a single point $\eta = 0$, so $\xi_{0,\lambda}$ assigns measure one to the single point $\eta = 0$. Since the set $\{\bar{\delta}_1 = 0, \ \bar{\delta}_2 = \lambda \text{ (fixed)}\}$ is a convex p_2-dimensional Euclidean set wherein each component η_i is $o(h(\lambda))$, any probability measure $\xi_{1,\lambda}$ can be replaced by the degenerate measure $\xi_{1\lambda}^*$ which assigns measure one to the mean $(0, \ldots, 0, \eta_{p_1+1}^0, \ldots, \eta_{p_1+p_2}^0$ of $\xi_{1,\lambda})$. Hence from (5.22)

$$\begin{aligned}
&\frac{\int f_{\lambda,\eta}(r)\ \xi_{1,\lambda}(d\eta)}{\int f_{0,\eta}(r)\ \xi_{0,\lambda}(d\eta)} \\
&\quad = \int\left(1 + \frac{\lambda}{2}\left[-1 + \bar{r}_1 + \sum_{j=p_1+1}^{p_1+p_2} r_j\left(\sum_{i>j}\eta_i + (N-j+1)\eta_j\right)\right]\right. \\
&\qquad \left. + B(r,\lambda,\eta)\right)\xi_{1,\lambda}(d\eta) \\
&\quad = 1 + \frac{\lambda}{2}\left[-1 + \bar{r}_1 + \sum_{j=p_1+1}^{p_1+p_2} r_j\left(\sum_{i>j}\eta_i^0 + (N-j+1)\eta_j^0\right)\right] + B(r,\lambda)
\end{aligned} \tag{5.23}$$

where $B(r,\lambda) = o(h(\lambda))$ uniformly in r.

Let R_k be the rejection region defined by

$$R_k = \{X : U(X) = \bar{R}_1 + k\bar{R}_2 \geq C_\alpha\} \tag{5.24}$$

where k is chosen such that (5.23) is reduced to yield (5.6) and the constant C_α depends on the level α of the test for the chosen k. Choose

$$\eta_j^0 = \left[\frac{(N-j-1)\cdots(N-p_1-p_2)}{(N-j+1)\cdots(N-p_1-p_2+2)}\right]\left[\frac{(N-p_1)}{p_2(N-p_1-p_2+1)}\right],$$

$$j = p_1 + 1, \ldots, p_1 + p_2 - 1,$$
$$\eta^0_{p_1+p_2} = \frac{(N - p_1)}{p_2(N - p_1 - p_2 + 1)} \tag{5.25}$$

so that

$$\sum_{i>j} \eta^0_i + (N - j + 1)\eta^0_j = \frac{N - p_1}{p_2}, \quad j = p_1 + 1, \ldots, p_1 + p_2 .$$

The test ϕ^* with rejection region

$$R^* = \left\{ X : U(X) = \bar{R}_1 + \frac{N - p_1}{p_2} \bar{R}_2 \geq C_\alpha \right\} \tag{5.26}$$

with $P_{0,\lambda}(R^*) = \alpha$ satisfies (5.6) as $\lambda \to 0$. Furthermore any region R_k of structure (5.24) must have $k = \frac{N-p_1}{p_2}$ to satisfy (5.6) for some $\xi_{1,\lambda}$. From Theorem 4.1.4, for any invariant region R^*, $P_{\lambda,\eta}(R^*)$ depends only on λ and hence from (5.5) $q(\lambda, \eta) = 0$ and thus ϕ^* is LBI for this problem as $\lambda \to 0$. Hotelling's T^2 test based on $\bar{R}_1 + \bar{R}_2$ does not coincide with ϕ^* and hence it is locally worse. It is easy to verify that the power function of Hotelling's test which depends only on $\bar{\delta}_2$, $\bar{\delta}_1$ being zero, has positive derivative everywhere, in particular, at $\bar{\delta}_2 = 0$. Thus from (5.5), with $R = R^*$, $h(\lambda) > 0$. Hence we get the following theorem.

Theorem 5.1.4. For Problem 3 the LBI test ϕ^* is locally minimax as $\lambda \to 0$.

Remarks. Hotelling's T^2 test and the likelihood ratio test are also invariant under (G_2, T_2) and therefore thier power functions are functions of $\bar{\delta}_2$ only. From the above theorem it follows that neither of these two tests maximises the derivative of the power function at $\bar{\delta}_2 = 0$. So Hotelling's T^2 test and the likelihood ratio test are not locally minimax for this problem.

Example 5.1.5. Consider Problem 6 of Chapter 4. The affine group (G, T), transforming

$$(\bar{X}, S; \mu, \Sigma) \to (g\bar{X} + t,\ gSg'; g\mu + t,\ g\Sigma g')$$

where $g \in G$, $t \in T$ leaves the problem of testing $H_0; \rho^2 = 0$ again $H_1 : \rho^2 = \lambda > 0$ (specified) invariant. However the subgroup $(G_T(p), T)$, where $G_T(p)$ is the multiplicative group of nonsingular lower triangular matrices of order p whose first column contains only zeros expect for the first element, satisfies Hunt-Stein conditions. The action of the translation group T is to reduce the mean μ to zero and $S = \sum_{\alpha=1}^{N} X^{\alpha}(X^{\alpha})'$ is sufficient where N has been reduced by one from what it was originally. We treat the latter formulation considering $X^1, \ldots, X^N$ to have zero mean and positive definite covariance matrix Σ and consider only the subgroup $G_T(p)$ for invariance. A maximal invariant in the sample space under $G_T(p)$ is $R = (R_2, \ldots, R_p)'$, where

$$\sum_{j=1}^{i} R_j = \frac{S_{(12)[i]}\ S_{(22)[i]}^{-1}\ S_{(21)[i]}}{S_{(11)}}, \quad i = 2, \ldots, p \tag{5.27}$$

where $C_{[i]}$ denotes the upper left-hand corner submatrix of C of order i and $b_{[i]}$ denotes the i-vector consisting of the first i components of a vector b and

$$S = \begin{pmatrix} S_{(11)} & S_{(12)} \\ S_{(21)} & S_{(22)} \end{pmatrix}$$

where $S_{(22)}$ is a square matrix of order $(p-1)$. As usual we assume that $N \geq p$ which implies that S is positive definite with probability one. Hence $R_i \geq 0$, $R^2 = U = \sum_{i=2}^{p} R_j \leq 1$. ($R^2$ is the squared sample multiple correlation coefficient). The corresponding maximal invariant in the parametric space is $\Delta = (\delta_2, \ldots, \delta_p)'$, where

$$\sum_{i=2}^{i} \delta_j = \frac{\Sigma_{(12)[i]}\Sigma_{(22)[i]}^{-1}\ \Sigma_{(21)[i]}}{\Sigma_{(11)}}, \quad i = 2, \ldots, p \tag{5.28}$$

with $\delta_i \geq 0$, $\rho^2 = \sum_{j=2}^{p} \delta_j$. Let

$$\eta = (\eta_2, \ldots, \eta_p)', \quad \eta_i = \delta_i/\rho^2 .$$

The joint distribution of $R_2, \ldots, R_p$ is (see Giri and Kiefer, 1964)

$$\begin{aligned} f_{\lambda,\eta}(r) = {} & \frac{(1-\lambda)^{\frac{1}{2}N}\left(1 - \sum_{j=2}^{p} r_j\right)^{\frac{1}{2}(N-p-1)}}{\pi^{\frac{1}{2}(p-1)}\Gamma(\frac{1}{2}(N-p+1))\left(1 + \sum_{j=2}^{p} r_j[(1-\lambda)/\gamma_i - 1]\right)^{\frac{N}{2}}} \\ & \times \prod_{j=2}^{p}\left[r_i^{\frac{1}{2}}\gamma_i^{\frac{1}{2}}(\pi_i+1)^{\frac{1}{2}(N-i+2)}\Gamma\Big(\frac{1}{2}(N-i+2)\Big)\right]^{-1} \\ & \times \sum_{\beta_2=0}^{\infty} \cdots \sum_{\beta_p=0}^{\infty} \Gamma\left(\sum_{j=2}^{p}\beta_j + \frac{N}{2}\right) \\ & \times \prod_{j=2}^{p}\left[\frac{\Gamma(\frac{1}{2}(N-i+2)+\beta_j)}{(2\beta_j)!}\right]\left[\frac{4r_j(1-\lambda)/\gamma_j(1+\pi_j^{-1})}{1 + \sum_{j=2}^{p} r_j[(1-\lambda)/\gamma_j - 1]}\right]^{\beta}, \end{aligned} \tag{5.29}$$

where $\gamma_j = 1 - \lambda \sum_{i=2}^{j} \eta_i$, $\pi_j = \delta_j / \gamma_j$. The expression $1/(1+\pi_j^{-1})$ means 0 if $\delta_j = 0$. From (5.29) we obtain

$$\frac{f_{\lambda,\eta}(r)}{f_{0,0}(r)} = 1 + \frac{N\lambda}{2}\left[-1 + \sum_{j=2}^{b} r_j \left(\sum_{i>j} \eta_i + (N-j+2)\eta_j \right)\right] + B(r,\lambda,\eta)\,, \quad (5.30)$$

where $B(r,\lambda,\eta) = o(\lambda)$ uniformly in r, η.

From Theorem 4.3.1 it follows that the assumption of Theorem 5.1.1 are satisfied with $U = \sum_{j=2}^{p} R_j = R^2$ and $h(\lambda) = b\lambda$, $b > 0$. As in example 5.1.3 letting $\xi_{1,\lambda}$ give measure one to $\eta^0 = (\eta_2^0, \ldots, \eta_p^0)$ where $\eta_j^0 = (N-j+1)^{-1}(N-j+2)^{-1}(p-1)^{-1}N(N-p+1)$, $j = 2, \ldots, p-1$ we get (5.6). Hence we prove the following theorem.

Theorem 5.1.5. *For every p, N and α the R^2 test which rejects H_0 for large values of the squared sample multiple correlation coefficient R^2 is locally minimax for testing H_0 against $H_1 : \rho^2 = \lambda \to 0$.*

Remarks. It is not a coincidence that (5.19) and (5.29) have the same form. (5.18) involves the ratio of noncentral to central chisquare distribution while (5.29) involves similar ratio with random noncentrality parameter. The first order terms in the expressions which involve only mathematical expectations of these quantities correspond each other.

Example 5.1.5. Consider problems 7 and 8 of Sec. 4.3. The group G as defined there does not satisfy the condition of the Hunt-Stein theorem. However this theorem does apply to the solvable subgroup G_T of $p \times p$ lower triangular matrices g of the form

$$g = \begin{pmatrix} g_{11} & 0 & 0 & \ldots & 0 \\ 0 & g_{22} & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & g_{2p} & \cdot & & g_{pp} \end{pmatrix}. \quad (5.31)$$

From Example 5.1.4 a maximal invariant in the sample space under G_T is $R = (R_2, \ldots, R_p)'$ and the corresponding maximal invariant in the parametric space is $\Delta = (\delta_2, \ldots, \delta_p)'$. From (5.29) R has a single distribution under H_{10} and H_{20} and it has a distribution which depends continuously on the p_1-dimension parameter $\Gamma_{1\lambda}$ under $H_{1\lambda}$ and on the p-dimension parameter $\Gamma_{2\lambda}$ under $H_{2\lambda}$, where

$$\Gamma_{1\lambda} = \{\Delta : \delta_i \geq 0,\ i = 2, \ldots, p_1 + 1, \quad \delta_i = 0,\ i = p_1 + 2, \ldots, p, \quad \sum_{i=2}^{p_1+1} \delta_i = \rho_1^2 = \lambda\},$$

$$\Gamma_{2\lambda} = \{\Delta : \delta_i = 0,\ i = 2, \ldots, p_1 + 1, \quad \delta_i \geq 0\,,\ i = p_1 + 2, \ldots, p, \quad \sum_{i=p_1+2}^{p} \delta_i = \rho_2^2 = \lambda\}. \tag{5.32}$$

Let

$$\eta = (\eta_2, \ldots, \eta_p)'$$

with

$$\eta_i = \begin{cases} \delta_i/\rho^2\,, & \text{if } \rho^2 > 0 \\ 0\,, & \text{if } \rho^2 = 0\,. \end{cases}$$

Because of the compactness of the reduced parameter space $\{0\}$ and $\Gamma_{1\lambda}$ and the continuity of $f_{\lambda,\eta}(r)$ in η we conclude from Wald (1950) that every minimax test for the reduced problem in terms of the maximal invariant R is Bayes. Thus any test based on R with constant power on $\Gamma_{1\lambda}$ is minimax for problem 7 if and only if it is Bayes. From (5.29) as $\lambda \to 0$ we get (for testing H_{10} against $H_{1\lambda}$)

$$\frac{f_{\lambda,\eta}(r)}{f_{0,\eta}(r)} = 1 + \frac{N\lambda}{2}\left[-1 + \sum_{j=2}^{p_1+1} r_j \left(\sum_{i>j} \eta_i + (N-j+1)\eta_j\right)\right] + B(r, \lambda, \eta) \tag{5.33}$$

where $B(r, \lambda, \eta) = o(\lambda)$ uniformly in r and η. It is obvious that the assumption of Theorem 5.1.1 is satisfied with

$$U = \sum_{j=2}^{p_1+1} R_j = \bar{R}_1\,,\ h(\lambda) = b\lambda$$

where $b > 0$. The set $\rho^2 = 0$ is a single point $\eta = 0$, so $\xi_{0\lambda}$ assigns measure one to the point $\eta = 0$. The set $\Gamma_{1\lambda}$ is a convex p_1-dimensional set wherein each component $\eta_i = o(h(\lambda))$. So for a Bayes solution any probability measure $\xi_{1,\lambda}$ can be replaced by the degenerate measure $\xi_{1\lambda}^*$ which assigns the probability one to mean $\eta^0 = (\eta_2^0, \ldots, \eta_{p_1+1}^0)$ of $\xi_{1\lambda}$. Choosing $\xi_{1\lambda}^*$ which assigns probability one to the point whose jth coordinate is $\eta_j^0 = (N-j+1)^{-1}(N-j+2)^{-1}p_2^{-1}(N-$

$p+1)(N-p_1)$, $j=2,\ldots,p_1+1$ we get (5.6) from (5.33). The condition (5.5) follows from Theorem 4.4.1. Hence we prove the following theorem.

Theorem 5.1.5. *For Problem 7 the likelihood ratio test of H_{10} defined in (4.38), is locally minimax against the alternatives $H_{1\lambda}$ as $\lambda \to 0$.*

In problem 8 as $\lambda \to 0$ (for testing H_{20} against $H_{2\lambda}$)

$$\frac{f_{\lambda,\eta}(r)}{f_{0,\eta}(r)} = 1+\frac{N\lambda}{2}\left[-1+\bar{r}_1+\sum_{j=p_1+2}^{p} r_j\left(\sum_{i>j}\eta_i+(N-j+2)\eta_j\right)\right]+B(r,\lambda,\eta)\,, \tag{5.34}$$

where $B(r,\lambda,\eta)=o(h(\lambda))$ uniformly in r,η. Since the set $\Gamma_{2\lambda}$ is a p_2-dimensional Euclidean set wherein each component $\eta_i = o(h(\lambda))$, using the same argument as in problem 7 we can write

$$\begin{aligned}
&\frac{\displaystyle\int f_{\lambda,\eta}(r)\xi_{1,\lambda}(d\eta)}{\displaystyle\int f_{0,\eta}(r)\xi_{0,\lambda}(d\eta)} \\
&\quad = \int\left\{1+\frac{N\lambda}{2}\left[-1+\bar{r}_1+\sum_{j=p_1+1}^{p} r_j\left(\sum_{i>j}\eta_i+(N-j+2)\eta_j\right)\right]\right. \\
&\qquad \left. +B(r,\lambda,\eta)\right\}\xi_{1,\lambda}(d\eta) \\
&\quad = 1+\frac{N\lambda}{2}\left[-1+\bar{r}_1+\sum_{j=p_1+1}^{p} r_j\left(\sum_{i>j}\eta_i^0+(N-j+1)\eta_j^0\right)\right]+B(r,\lambda)
\end{aligned} \tag{5.35}$$

where $B(r,\lambda) = o(h(\lambda))$ uniformly in r and $\xi_{1\lambda}$ assigns measures one to $(\eta^0_{p_1+2},\ldots,\eta^0_p)'$.

Let R_k be the rejection region, given by,

$$R_k = \{x : U(x) = \bar{r}_1 + k\bar{r}_2 \geq c_\alpha\} \tag{5.36}$$

where k is chosen such that (5.36) is reduced to yield (5.5) and c_α depends on the level of significance α of the test for the chosen k. Now choosing

$$\eta_j^0 = \frac{(N-p+2)(N-p+1)}{(N-j+2)(N-j+1)}\,\eta_p^0\,, \quad j=p_1+1,\ldots,p\,,$$

$$\eta_p^0 = \frac{N-p_1}{(N-p+2)p_2}$$

we conclude that the test ϕ^* with the rejection region

$$R^* = \left\{x : U(x) = \bar{r}_i + \frac{N - p_1}{p_2}\bar{r}_2 \geq c_\alpha\right\}$$

with $P_{0,\lambda}(R^*) = \alpha$ satisfies (5.6) as $\lambda \to 0$. Moreover any region R_k of the form (5.36) must have $k = \frac{N-p_i}{p_2}$ in order to satisfy (5.36) for some $\xi_{1,\lambda}$. From (4.37) it is easy to conclude that the test ϕ^* is LBI for problem 8 as $\lambda \to 0$. The R^2-test which rejects H_{20} whenever $r^2 = \bar{r}_1 + \bar{r}_2 \geq c_\alpha$ does not coincide with ϕ^* and is locally worse. From Sec. 4.3 it is evident that the power function of R^2 test depends only on ρ^2 and has positive derivatives everywhere in particular at $\rho^2 = 0$. From (5.5) with $R = R^*$, $h(\lambda) > 0$. So we get the following theorem.

Theorem 5.1.6. *For Problem 8 the LBI test ϕ^* which rejects H_{20} whenever $\bar{r}_1 + \frac{N-p_1}{p_2}\bar{r}_2 \geq c_\alpha$ where c_α depends on the size α of the test, is locally minimax against the alternatives $H_{2\lambda}$ as $\lambda \to 0$.*

Remarks. In order for an invariant test of H_{20} against $H_{2\lambda}$ to be minimax, it has to be minimax among all invariant tests also. However, since for an invariant test the power function is constant on each contour $\rho^2 = \rho_2^2 = \lambda$, "minimax" simply means "most powerful". The rejection region of the most powerful test is obtained from (5.34), from which it is easy to conclude that $[f_{\lambda,\eta}(r)/f_{0,\eta}(r)]$ depends non-trivially on λ so that no test can be most powerful for every value of α.

5.2. Asymptotically Minimax Tests

We treat here the setting of (5.1) when $\lambda \to \infty$. Expressions like $o(1)$, $o(H(\lambda))$ are to be interpreted as $\lambda \to \infty$. We shall be concerned here in maximizing a probability of error which tends to zero. The reader, familiar with the large sample theory, may recall that in this setting it is difficult to compare directly approximations to such small probabilities for different families of tests and one instead compares their logarithms. While our considerations are asymptotic in a sense not involving sample sizes, we encounter the same difficulty which accounts for the form (5.39) below.

Assume that the region

$$R = \{x : U(x) \geq c_\alpha\} \tag{5.37}$$

satisfies in place of (5.5)

$$\begin{aligned} &P_{0,\eta}(R) = \alpha\,, \\ &P_{\lambda,\eta}(R) = 1 - \exp\{-H(\lambda)\ (1 + o(1))\} \end{aligned} \tag{5.38}$$

where $H(\lambda) \to \infty$ as $\lambda \to \infty$ and $o(1)$ term is uniform in η. Suppose that

$$\frac{\int p(x,\lambda,\eta)\xi_{1,\lambda}(d\eta)}{\int p(x,0,\eta)\xi_{0,\lambda}(d\eta)} = \exp\{H(\lambda)[G(\lambda) + R(\lambda)U(x)] + B(x,\lambda)\} \tag{5.39}$$

where $\sup_x |B(x,\lambda)| = o(H(\lambda))$, and $0 < c_1 < R(\lambda) < c_2 < \infty$. One other regularity assumption is that c_α is a point of increase from the left of the distribution function of U when $\delta = 0$ uniformly in η, i.e.

$$\inf_\eta P_{0,\eta}(U \geq c_\alpha - \epsilon) > \alpha \tag{5.40}$$

for every $\epsilon > 0$.

Theorem 5.2.1. *If U satisfies* (5.38)–(5.40) *and for sufficiently large λ there exist $\xi_{0,\lambda}$ and $\xi_{1,\lambda}$ satisfying* (5.39), *then the rejection region R is asymptotically logarithmically minimax of level α for testing $H_0 : \delta = 0$ against the alternatives $H_1 : \delta = \lambda$ as $\lambda \to \infty$, i.e.,*

$$\lim_{\lambda\to\infty} \frac{\inf_\eta[-\log(1 - P_{\lambda,\eta}\{R\})]}{\sup_{\phi_\lambda \in Q_\alpha} \inf_\eta[-\log(1 - P_{\lambda,\eta}(\phi_\lambda \text{ rejects } H_0))]} = 1\,. \tag{5.41}$$

Proof. Assume that (5.41) does not hold. Then there exists an $\epsilon > 0$ and an unbounded sequence Γ of values λ with corresponding tests ϕ_λ in Q_α whose critical region satisfies

$$P_{\lambda,\eta}\{R\} > 1 - \exp\{-H(\lambda)\ (1 + 5\epsilon)\} \tag{5.42}$$

for all η. There are two possible cases, (5.43) and (5.46). If $\lambda \in \Gamma$ and

$$-1 - G(\lambda) \leq R(\lambda)\ \ c_\alpha + 2\epsilon\,, \tag{5.43}$$

consider the *a priori* distribution ξ_λ (see Theorem 5.1.1) given by $\xi_{i,\lambda}$ and τ_λ satisfying

$$\tau_\lambda/(1 - \tau_\lambda) = \exp\{H(\lambda)\ (1 + 4\epsilon)\}\,. \tag{5.44}$$

Using (5.42) and (5.44) the integrated risk of any Bayes procedure B_λ must satisfy

$$\begin{aligned} r^*_\lambda(B_\lambda) \leq r^*_\lambda(\phi_\lambda) &\leq (1 - \tau_\lambda)\alpha + \tau_\lambda \exp\{-H(\lambda)\ (1 + 5\epsilon)\} \\ &= (1 - \tau_\lambda)[\alpha + \exp\{-\epsilon\ H(\lambda)\}]\,. \end{aligned} \tag{5.45}$$

From (5.39) a Bayes critical region is

$$B_\lambda = \{x : U(x) + B(x,\lambda)/R(\lambda)H(\lambda) \geq [-(1+4\epsilon) - G(\lambda)]/R(\lambda)\}\,.$$

Hence, if λ is so large that

$$\sup_x \left| \frac{B(x,\lambda)}{H(\lambda)R(\lambda)} \right| < \epsilon c_2\,,$$

we conclude from (5.43) that

$$B_\lambda \supset \{x : U(x) > c_\alpha - \epsilon/c_2\} = B'_\lambda(\text{say})\,.$$

The assumption (5.40) implies that $P_{0,\eta}\{B'_\lambda\} > \alpha + \epsilon'$ with $\epsilon' > 0$, contradicting (5.45) for large λ.

On the other hand if $\lambda \in \Gamma$ and

$$-1 - G(\lambda) > R_\lambda C_\alpha + 2\epsilon\,, \tag{5.46}$$

consider the *a priori* distribution ξ_λ given by $\xi_{i,\lambda}$, and τ_λ satisfying

$$\tau_\lambda/(1-\tau_\lambda) = \exp\{H(\lambda)\ (1+\epsilon)\}\,. \tag{5.47}$$

Using (5.39) a Bayes critical region is

$$B_\lambda = \{x : U(x) + B(x,\lambda)/R(\lambda)H(\lambda) \geq [-(1+\epsilon) - G(\lambda)]/R(\lambda)\}\,.$$

Thus, if $\sup_x |\frac{B(x,x)}{R(\lambda)\ B(\lambda)}| < \epsilon/2c_2$, we conclude from (5.46) that $B_\lambda \subset R$, so that, by (5.37) and (5.47),

$$\begin{aligned} r^*(B_\lambda) &> \tau_\lambda \exp\{-H(\lambda)[1+o(1)]\} \\ &= (1-\tau_\lambda)\exp\{H(\lambda)(\epsilon - o(1))\}\,. \end{aligned} \tag{5.48}$$

Since

$$\begin{aligned} r^*(B_\lambda) \leq r^*(\phi_\lambda) &\leq (1-\tau_\lambda)\alpha + \tau_\lambda \exp\{-H(\lambda)(1+5\epsilon)\} \\ &= (1-\tau_\lambda)[\alpha + \exp\{-4\epsilon\ H(\lambda)\}]\,, \end{aligned}$$

we contradict (5.48) for sufficiently large λ. □

Example 5.2.1. (T^2-test) In Problem 1 of Chapter 4 let $U(X) = \sum_1^p R_j$. Since $\phi(a,b;x) = \exp\{x(1+o(1))\}$ as $x \to \infty$ we get, using (5.18),

$$\frac{f_{\lambda,\eta}(r)}{f_{0,\eta}(r)} = \exp\left\{\frac{\lambda}{2}\left[-1 + \sum_{j=1}^p r_j \sum_{i \geq j} \eta_i\right](1 + B(r,\eta,\lambda))\right\} \tag{5.49}$$

with $\sup_{r,\eta}|B(r,\eta,\lambda)| = o(1)$ as $\lambda \to \infty$. From (4.5) putting $\eta_p = 1$, $\eta_i = 0$, $i < p$, in (5.49) the density of U being independent of η we see that

$$P_{\lambda,\eta}\{U < c_\alpha\} = \exp\{(c_\alpha - 1)[1 + o(1)]\} \tag{5.50}$$

as $\lambda \to \infty$. Hence (5.38) is satisfied with $H(\lambda) = \frac{1}{2}(1 - c_\alpha)$. Now letting $\xi_{1,\lambda}$ assign measure one to the point $\eta_1 = \cdots = \eta_{p-1} = 0$, $\eta_p = 1$ and $\xi_{0,\lambda}$ assign measure one to (0,0) we get (5.39). Finally (5.40) is trivial. From Theorem 5.2.1 we get Theorem 5.2.2.

Theorem 5.2.2. *For every p, N, α, Hotelling's T^2 test is asymptotically minimax for testing $H_0 : \delta = 0$ against $H_1 : \delta = \lambda$ as $\lambda \to \infty$.*

From Sec. (5.1) we conclude that no critical region of the form $\sum a_i\, R_i > c$ other than Hotelling's would have been locally minimax, many regions of this form are asymptotically minimax.

Theorem 5.2.3. *$c < 1$ and $1 = a_1 \le a_2 \le \cdots \le a_p$, then the critical region $\{\sum a_i\, R_i > c\}$ is asymptotically minimax among tests of same size as $\lambda \to \infty$.*

Proof. The maximum of $\sum_j r_j \sum_{i>j} \eta_i$ subject to $\sum_i a_i r_i \le c$ is achieved at $r_1 = c$, $r_2 = \cdots = r_p = 0$. Hence the integration $f_{\lambda,\eta}(r)$ over a small region near that point yields (5.50) with c_α replaced by c. Since a_j's are nondecreasing in j it follows from (5.49) that $\xi_{1,\lambda}$ can be chosen to yield (5.39) with $U = \sum_1^p a_i R_i$. Again (5.40) is trivial. □

Example 5.2.2. Consider again Problem 2 of Chapter 4. From Example 5.2.1 with $p = p_1$, $R = (R_1, \ldots, R_{p_1})'$, $\eta = (\eta_1, \ldots, \eta_{p_1})'$ we get

$$\frac{f_{\lambda,\eta}(r)}{f_{0,\eta}(r)} = \exp\left\{\frac{\lambda}{2}\left[-1 + \sum_{j=1}^{p} r_j \sum_{i \ge j} \eta_i\right](1 + B(r,\lambda,\eta))\right\} \tag{5.51}$$

with $\sup_{r,\eta}|B(r,\lambda,\eta)| = o(1)$ as $\lambda \to \infty$. Using the same argument of the above theorem we now prove the following.

Theorem 5.2.3. *For every p_1, N, α the UMPI test is asymptotically minimax for testing $H_0 : \bar{\delta}_1 = 0$ against $H_1 : \bar{\delta} = \lambda$ as $\lambda \to \infty$.*

Example 5.2.3. In Problem 3 of Chapter 4 we want to find the asymptotically minimax test of $H_0 : \bar{\delta}_1 = \bar{\delta}_2 = 0$ against the alternatives $H_1 : \bar{\delta}_1 = 0$,

$\bar{\delta}_2 = \lambda$ as $\lambda \to \infty$. In the notations of Examples 5.1.4 and 5.2.1 we get, $\bar{r}_1 = \sum_1^{p_1} r_j$, $\bar{r}_1 + \bar{r}_2 = \sum_1^{p_1+p_2} r_j$, $r = (r_1, \ldots, r_{p_1+p_2})'$, $\eta = (\eta_1, \ldots, \eta_{p_1+p_2})'$, $\eta_i = 0$, $i = 1, \ldots, p_1 + p_2$ under H_0 and $\eta_i = 0$, $i = 1, \ldots, p_1$ under H_1. Hence

$$\frac{f_{\lambda,\eta}(r)}{f_{0,\eta}(r)} = \exp\left\{\left[-1 + \bar{r}_1 + \sum_{j=1}^{p_1+p_2} r_j \sum_{i \geq j} \eta_i\right](1 + B(r, \lambda, \eta))\right\} \tag{5.52}$$

where $\sup_{r,\eta} |B(r, \lambda, \eta)| = o(1)$ as $\lambda \to \infty$. From (4.9) when $\bar{\delta}_2 = \lambda \to \infty$, $\bar{\delta}_1 = 0$ we get

$$\frac{f(\bar{r}_1, \bar{r}_2|\lambda)}{f(\bar{r}_1, \bar{r}_2|0)} = \exp\left\{\frac{\lambda}{2}[-1 + \bar{r}_1 + \bar{r}_2](1 + B(\bar{r}_1, \bar{r}_2, \lambda))\right\} \tag{5.53}$$

where $\sup_{\bar{r}_1, \bar{r}_2} |B(\bar{r}_1, \bar{r}_2, \lambda)| = o(1)$ as $\lambda \to 0$. Letting $\xi_{0\lambda}$ in (5.39) assign measure one to the single point $\eta = 0$ we get from (5.52)

$$\begin{aligned} &\frac{\int f_{\lambda,\eta}(r)\, \xi_{1,\lambda}(d\eta)}{\int f_{0,\eta}(r)\, \xi_{0,\lambda}(d\eta)} \\ &\quad = \int \exp\left\{\frac{\lambda}{2}\left[-1 + \bar{r}_1 + \sum_{j=p_1+1}^{p_1+p_2} r_j \sum_{i \geq j} \eta_i\right](1 + B(r, \eta, \lambda))\right\} \xi_{1,\lambda}(d\eta) \end{aligned} \tag{5.54}$$

Let R_k be the rejection region of size α, based on $\bar{R}_1, \bar{R}_2$ given by

$$R_k = \{x :\ U(x) = \bar{R}_1 + k\,\bar{R}_2 \geq c_\alpha\} \tag{5.55}$$

where k is chosen such that (5.54) is reduced to yield (5.39) and R_k for the chosen k satisfies (5.38) and (5.40). Now letting $\eta_1 = \cdots = \eta_{p_1+p_2-1} = 0$, $\eta_{p_1+p_2} = 1$ we see that (5.54) is reduced to yield (5.39) with $U(x) = \bar{R}_1 + \bar{R}_2$. From (5.53)

$$P(U(x) \leq c''_\alpha) = \exp\left\{\frac{\lambda}{2}(c''_\alpha - 1)(1 + o(1))\right\}. \tag{5.56}$$

Hence Hotelling's test which rejects H_0 for large values of $\bar{R}_1 + \bar{R}_2$ satisfies (5.38) with $H(\lambda) = \frac{\lambda}{2}(1 - c''_\alpha)$. The fact that Hotelling's test satisfies (3.39) is trivial. Since the coefficient of r_{p_1+1} in the expression inside the brackets in the exponent of (5.54) is one, any rejection region R_k must have $k = 1$ to satisfy (5.39) for some $\xi_{1,\lambda}$. From Theorem 5.2.1 a region R_k with $k \neq 1$ and which satisfies (5.38) and (5.40) cannot be minimax as $\lambda \to \infty$. Using Theorem 4.1.5 we now prove the following:

Theorem 5.2.4. *Hotelling's test which is asymptotically best invariant is asymptotically minimax for testing* $H_0 : \bar{\delta}_1 = \bar{\delta}_2 = 0$ *against* $H_1 : \bar{\delta}_1 = 0$, $\bar{\delta}_2 = \lambda$ *as* $\lambda \to \infty$.

Remarks. From Theorem 5.2.3 it is obvious that there are other asymptotically minimax tests, not of the form R_k, for this problem. It is easy to see that

$$P_{\lambda,\eta}\left(\bar{R}_1 + (1-c)^{-1}\bar{R}_2 \geq 1\right) = 1 - \exp\left\{-(\log \lambda^2 - \log\left((1-c)\,(1+o(1))\right)\right\},$$

$$P_{\lambda,\eta}\left(\bar{R}_1 + \frac{N-p_1}{p_2}\bar{R}_2 \geq c_\alpha\right) = 1 - \exp\left\{-\frac{\lambda}{2}(1-c_\alpha)(1+o(1))\right\}.$$

Thus the fact that the likelihood ratio test which rejects H_0 whenever $\frac{1-\bar{R}_1-\bar{R}_2}{1-\bar{R}_1} \leq c$ and the locally minimax test satisfy (5.40) and (5.38) is obvious in both cases. But from Theorem 5.2.4 these two tests are not assymptotically minimax for this problem.

5.3. Minimax Tests

In this section we discuss the genuine minimax character of Hotelling's T^2 and the test based on the square of the sample multiple correlation coefficient R^2. These problems have remained elusive even in the simplest case of $p = 2$ or 3 dimensions. In the case of Hotelling's T^2 test Semika (1940) proved the UMP character of T^2 test among all level α tests with power functions depending only on $\sigma = N\mu'\Sigma^{-1}\mu$. In 1956 Stein proved the admissibility character of T^2 test by a method which could not be used to prove the admissibility character of R^2-test. Giri, Kiefer and Stein (1964a) attacked the problem of minimax property of Hotelling's T^2-test among all level α tests and proved its minimax propery for the very special case of $p = 2$, $N = 3$ by solving a Fredholm integral equation of first kind which is transformed into an "overdetermined" linear differential equation of first order. Linnik, Pliss and Šalaeveski (1969) extended this result to $N = 4$ and $p = 2$ using a slightly more complicated argument to construct an overdetermined boundary problem with linear differential operator. Later Šalaveski (1969) extended this result to the general case of $p = 2$.

5.3.1. *Hotelling's T^2 test*

In the setting of Examples 5.1.1 and 5.2.1 we give first the details of the proof of the minimax property of Hotelling's T^2 test as developed by Giri, Kiefer and Stein (1964a), then give the method of proof by Lunnik, Pliss and Sālaeveski (1966). In this setting a maximal invariant under $G_T(p)$ is $R = (R_1, \ldots, R_p)'$ with $\sum_1^p R_i = N\bar{X}'(S + N\bar{X}\bar{X})^{-1}\bar{X}$ and the corresponding maximal invariant in the parametric space is $\Delta = (\delta_1, \ldots, \delta_p)'$ with $\sum_1^p \delta_i = N\mu'\Sigma^{-1}\mu = \delta$. From (5.18) R has a single distribution under H_0 and its distribution under $H_1 : \ \delta = \lambda$ (fixed) depends continuously on a $p-1$ dimensional parameter

$$\Gamma = \left\{ \Delta = (\delta_1, \ldots, \delta_p)' : \ \delta_i \geq 0 \,, \ \sum_1^p \delta_i = \lambda \right\}.$$

Thus there is no UMPI test under $G_T(p)$ as it was under $G_l(p)$. Let us write $f_\Delta(r)$ as the pdf of R as given in (5.18). Because of the compactness of the reduced parametric spaces $\{0\}$ and Γ and the continuity of $f_\Delta(r)$ in Δ we conclude from Wald (1950) that every minimax test for the reduced problem in terms of R is Bayes. In particular Hotelling's T^2 test which rejects H_0 whenever $U = \sum_1^p R_i > c$, which is also G_T-invariant and has a constant power function on each contour $N\mu'\Sigma^{-1}\mu = \lambda$, maximizes the minimum power over H_1 if and only if there is a probability measure ξ on Γ such that for some constant c_1

$$\int_\Gamma \frac{f_\Delta(r)}{f_0(r)} \, \xi(d\Delta) \qquad \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \ c_1 \tag{5.57}$$

according as

$$\sum_1^p r_i \qquad \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \ c$$

except possibly for a set of measure zero where c depends on the specified level of significance α and c_1 may depend on c and the specified λ. An examination of the integrand of (5.57) allows us to replace it by its equivalent

$$\int_\Gamma \frac{f_\Delta(r)}{f_0(r)} \, \xi(d\Delta) = c_1 \qquad \text{if} \quad \sum_1^p r_i = c \,. \tag{5.58}$$

Obviously (5.57) implies (5.58). On the other hand, if there are a ξ and a constant c_1 for which (5.58) is satisfied and if $r^* = (r_1^*, \ldots, r_p^*)'$ is such that $\sum_1^p r_i^* = c' > c$, then writing $f = \frac{f_\Delta}{f_0}$ and $r^{**} = \frac{c}{c'} r^*$ we conclude that

$$f(r^*) = f\left(\frac{c'}{c}r^{**}\right) > f(r^{**}) = c_1$$

because of the form of f and the fact that $c'/c > 1$ and $\sum_1^p r_i^{**} = \frac{c}{c'}\sum_1^p r_i^* = c$. This and similar argument for the case $c' < c$ show that (5.58) implies (5.57). Of course we do not assert that the left-hand side of (5.58) still depends only on $\sum_1^p r_i$ if $\sum_1^p r_i \neq c$.

The computation is somewhat simplified by the fact that for fixed c and λ we can at this point compute the unique value of c_1 for which (5.58) can possibly be satisfied.

Let $\hat{R} = (R_i, \ldots, R_{p-1})'$ and let $f_\triangle(\hat{r}/u)$ be the conditional pdf of $\hat{R}$ given $\sum_{i=1}^p R_i = u$ with respect to the Lebesgue measure, which is continuous in $\hat{r}$ and u with $r_i > 0$ and $\sum_1^{p-1} r_i < u < 1$. Denote by $f_\delta^{**}(u)$ the pdf of $U = \sum_1^p R_i$, which is continuous for $0 < u < 1$ and vanishes elsewhere and which depends on $\triangle$ only through δ. Then (5.58) can be written as

$$\int f_\triangle(\hat{r}|c)\xi(d\triangle) = \left[c_1 \frac{f_0^{**}(c)}{f_\delta^{**}(c)}\right] f_0(\hat{r}|c) \tag{5.59}$$

if $r_i > 0$, $\sum_1^{p-1} r_i < c$. The left hand integral in (5.59), being a probability mixture of probability densities is itself a probability density in $\hat{r}$ as is $f_0(\hat{r}|c)$. Hence the expression inside square brakets of (5.59) is equal to one. From (4.5)

$$f_\delta^{**}(c) = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N-p}{2}\right)} \exp\left\{-\frac{1}{2}\delta\right\}$$
$$\cdot\, c^{\frac{1}{2}(p-2)}(1-c)^{\frac{1}{2}(N-p-2)}\phi\left(\frac{N}{2}, \frac{p}{2}; \frac{c\delta}{2}\right). \tag{5.60}$$

Using (5.60) we rewrite (5.58) as

$$\int_\Gamma \exp\left\{\frac{1}{2}\sum_1^p r_j \sum_{i>j} \delta_i\right\} \prod_{i=1}^p \phi\left(\frac{N-i+1}{2}, \frac{1}{2}; \frac{r_i\delta_i}{2}\right)\xi(d\triangle) = \phi\left(\frac{N}{2}, \frac{p}{2};\ \frac{c\lambda}{2}\right) \tag{5.61}$$

for all r with $r_i > 0$, $\sum_1^p r_i = c$. Write Γ_1 for the unit $(p-1)$-simplex

$$\Gamma_1 = \left\{(\beta_1, \ldots, \beta_p) : \beta_i \geq 0\,,\ \sum_1^p \beta_i = 1\right\}$$

where $\beta_i = \delta_i/\lambda$. Let $t_i = \frac{\gamma r_i}{c}$, $\gamma = c\lambda$ and ξ^* be the probability measure on Γ_1 associated with ξ on Γ. Since $\xi^*(A) = \xi(\delta A)$, (5.61) reduces to

$$\int_{\Gamma_1} \exp\left\{\sum_{j=1}^{p} t_j \sum_{i>j} (\frac{1}{2}\beta_i)\right\} \prod_{i=1}^{p} \phi\left(\frac{1}{2}(N-i+1),\ \frac{1}{2};\ \frac{1}{2}t_i\beta_i\right) d\xi^*(\beta_1,\ldots,\beta_p)$$
$$= \phi\left(\frac{N}{2}, \frac{p}{2}; \frac{1}{2}\gamma\right) \tag{5.62}$$

for all $(t_1,\ldots,t_p)$ with $\sum_1^p t_i = \gamma$ and $t_i > 0$ and hence, by analyticity for all $(t_1,\ldots,t_p)$ with $\sum_1^p t_1 = \gamma$.

From (5.62) it is evident that such ξ^*, if it exists, depends on c and λ only through their product γ. Note that for $p = 1$, Γ_1 is a single point but the dependence on γ in other cases is genuine.

Case $\mathbf{p = 2, N = 3}$. Solution of Giri, Kiefer and Stein

Since $\phi(\frac{3}{2}, \frac{1}{2}; \frac{1}{2}x) = (1+x)\exp(\frac{1}{2}x)$, from (5.62) we obtain

$$\int_0^1 [1 + (\gamma - t_2)(1-\beta_2)]\,\phi\left(1, \frac{1}{2}; \frac{1}{2}\beta_2 t_2\right)\, d\xi^*(\beta_2) = e^{\frac{1}{2}(t_2-\gamma)}\phi\left(\frac{3}{2}, 1; \frac{1}{2}\gamma\right) \tag{5.63}$$

with $t_1 = \gamma - t_2$, $\beta_1 = 1 - \beta_2$. We could now try to solve (5.63) for ξ^* by using the theory of Meijer transform with kernal $\phi(1, \frac{1}{2}; \frac{1}{2}x)$. Instead we expand both sides of (5.63) as power series in t_2. Let

$$\mu_i = \int_0^1 (\beta_2)^i d\xi^*(\beta_2)\,, \quad i = 0, 1, \ldots$$

be the ith moment of β_2. From (5.63) we obtain

(a) $1 + \gamma - \gamma\mu_1 = B$,

(b) $-(2r-1)\,\mu_{r-1} + (2r+\gamma)\mu_r - \gamma\mu_{r+1} = B\left[\frac{\Gamma(r+\frac{1}{2})}{\Gamma(r)\Gamma\left(\frac{1}{2}\right)}\right]$ (5.64)

where $B = e^{-\frac{1}{2}\gamma}\phi(\frac{3}{2}, 1; \frac{1}{2}\gamma)$. We could now try to show that the sequence $\{\mu_i\}$ given by (5.64) satisfies the classical necessary and sufficient condition for it to be the moment sequence of a probability measure on [0,1] or, equivalently, that the Laplace transform

$$\sum_{j=0}^{\infty} \mu_j(-t)^j/j!$$

is completely monotone on $[0, \infty)$, but we have been unable to proceed successfully in this way. Instead, we shall obtain a function $m_\gamma(x)$ which we prove below to be the Lebesgue density $d\xi^*(x)/dx$ of an absolutely continuous probability measure ξ^* satisfying (5.64) and hence (5.63).

The generating function

$$\psi(t) = \sum_{j=0}^{\infty} \mu_j \, t^j$$

of the sequence $\{\mu_i\}$ satisfies a differential equation which is obtained by multiplying (5.64) (b) by t^{r-1} and summing from 1 to ∞:

$$\begin{aligned} &2t^2(1-t)\psi'(t) - (t^2 - \gamma t + \gamma)\psi(t) \\ &\quad = B\, t[(1-t)^{-1/2} - 1] + \gamma[t(1-\mu) - 1] \\ &\quad = B\, t(1-t)^{-1/2} - t - \gamma\,. \end{aligned} \tag{5.65}$$

This is solved by treatment of the corresponding homogeneous equation and by variation of parameter to yield

$$\psi(t) = \frac{e^{-\gamma/2t}}{(1-t)^{1/2}} \int_0^t e^{\gamma/2T}\left[\frac{-1}{2T(1-T)^{1/2}} - \frac{\gamma}{2T^2(1-T)^{1/2}} + \frac{B}{2T(1-T)}\right] dT\,, \tag{5.66}$$

the integration being understood to start from the origin along the negative real axis of the complex plane. The constant of integration has been chosen to make ψ continuous at 0 with $\psi(0) = 1$, and (5.66) defines a single-valued function on the complex plane minus a cut along the real axis from 1 to ∞. The analyticity of ψ on this region can easily be demonstrated by considering the integral of ψ on a closed curve about 0 avoiding 0 and the cut, making the inversion $w = \frac{1}{t}$, shrinking the path down to the cut $0 \le w \le 1$ and using (5.67) below. Now, if there existed an absolutely continuous ξ^* whose suitably regular derivative m_γ satisfied

$$\int_0^1 m_\gamma(x)/(1-tx)dx = \psi(t)\,, \tag{5.67}$$

we could obtain m_γ by using the inversion formula

$$m_\gamma(x) = (2\pi i x)^{-1} \lim_{\epsilon \to 0}[\psi(x^{-1} + i\epsilon) - \psi(x^{-1} - i\epsilon)]\,. \tag{5.68}$$

However, there is nothing in the theory of the Stieltjes transform which tells us that an $m_\gamma(x)$ satisfying (5.68) does satisfy (5.67) and, hence (5.63), so we use (5.68) as a formal device to obtain an m_γ which we shall then prove satisfied (5.63).

From (5.66) and (5.68) we obtain, for $0 < x < 1$,

$$m_\gamma(x) = \frac{e^{-\gamma x/2}}{2\pi\, x^{1/2}(1-x)^{1/2}} \left\{ \int_0^\infty \left[\frac{B}{1+u} - \frac{u^{1/2}}{(1+u)^{3/2}}\right] + B\int_0^x \frac{e^{\gamma u/2}}{1-u} du \right\}. \tag{5.69}$$

In order to prove that $d\xi^*(x) = m_\gamma(x)\,dx$ satisfies (5.63) with ξ^* a probability measure we must prove that

$$\begin{aligned}
&\text{(a)} \;\; m_\gamma(x) \geq 0 \text{ for almost all } x,\ 0 \leq x \leq 1, \\
&\text{(b)} \;\; \textstyle\int_0^1 m_\gamma(x)\, dx = 1 \\
&\text{(c)} \;\; \mu_1 = \textstyle\int_0^1 x m_\gamma(x)\, dx \text{ satisfies (5.64) (a)} \\
&\text{(d)} \;\; \mu_r = \textstyle\int_0^1 x^r m_\gamma(x)\, dx \text{ satisfies (5.64) (b) for all } r \geq 1\,.
\end{aligned} \tag{5.70}$$

The first condition follows from (5.69) and the fact that $B > 1$ and $u^{3/2}(1+u) < (1+u)^{3/2}$ for $u > 0$. The condition (d) follows from the fact that $m_\gamma(x)$ as defined in (5.69) satisfies the differential equation

$$m'_\gamma(x) + m_\gamma(x)\left[\frac{\gamma}{2} + (1-2x)/2x(1-x)\right] = B/2\pi x^{1/2}(1-x)^{3/2}\,, \tag{5.71}$$

so that an integration by parts yields, for $r \geq 1$,

$$\begin{aligned}
(r+1)\mu_r - r\,\mu_{r-1} &= \int_0^1 [(r+1)x^r \; - r\, x^{r-1}] m_\gamma(x)\, dx \\
&= \int_0^1 (x^r - x^{r+1}) m'_\gamma(x)\, dx \\
&= \mu_r(1-\gamma/2) + \gamma\, \mu_{r+1}/2 - \mu_{r-1}/2 \\
&\quad + B\Gamma\left(r + \frac{1}{2}\right)/2\pi^{\frac{1}{2}} r!
\end{aligned}$$

which is (5.64) (b).

To prove (5.70) (b) and (e) we need the following identities involving confluent hypergeometric function $\phi(a,b;x)$. The materials presented here can be found, for example, in Erdélyi (1953), Chapter 6 or Stater (1960). The confluent hypergeometric function has the following integral representation

$$\phi(a,b;x) = \sum_{j=0}^{\infty}[\,\Gamma(a+j)\Gamma(b)/\Gamma(a)\Gamma(b+j)\;j!\,]x^j$$
$$= \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 e^{xt}t^{a-1}(1-t)^{b-a-1}dt \tag{5.72}$$

when $b > a > 0$. The associated solution ψ to the hypergeometric equation has the representation

$$\psi(a,b;x) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)}\int_0^1 e^{-xt}t^{a-1}(1+t)^{b-a-1}dt \tag{5.73}$$

if $a > 0$. We shall use the fact the general definition of ψ, as used in what follows when $a = 0$, satisfies

$$\psi(0,b;x) = 1\,. \tag{5.74}$$

The functions ψ and ϕ satisfy the following differential properties, identities and integrals.

$$\frac{d}{dx}\phi(a,b;x) = \left(\frac{a}{b}-1\right)\phi(a,b+1;x) + \phi(a,b;x)\,, \tag{5.75}$$

$$\frac{d}{dx}\psi(a,b;x) = \psi(a,b;x) - \psi(a,b+1;x)\,, \tag{5.76}$$

$$\frac{d}{dx}\psi(a,b;x) = x^{-1}a[(a-b+1)\psi(a+1,b;x) - \psi(a,b;x)]\,, \tag{5.77}$$

$$(a-b+1)\phi(a,b;x) - a\phi(a+1,b;x) + (b-1)\phi(a,b-1;x) = 0\,, \tag{5.78}$$

$$b\phi(a,b;x) - b\phi(a-1,b;x) - x\phi(a,b+1;x) = 0\,, \tag{5.79}$$

$$\psi(a,b;x) - a\psi(a+1,b;x) - \psi(a,b-1;x) = 0\,, \tag{5.80}$$

$$(b-a)\psi(a,b;x) - x\psi(a,b+1;x) + \psi(a-1,b;x) = 0\,, \tag{5.81}$$

$$\int_0^{\infty} e^{-x}(x+y)^{-1}\phi\left(\frac{1}{2},1;x\right)dx = \psi\left(\frac{1}{2},1;y\right) \quad \text{for } y > 0\,, \tag{5.82}$$

$$\int_0^\infty e^{-x}(x+y)^{-2}\phi\left(\frac{1}{2},1;x\right)dx$$
$$= -\left[\Gamma\left(\frac{1}{2}\right)/2y\right]\left[\psi\left(\frac{3}{2},1;y\right)-\psi\left(\frac{1}{2},1;y\right)\right]. \tag{5.83}$$

Using (5.73) , in terms of hypergeometric functions, for $0 < x < 1$, the m_γ, defined in (5.69), can be written as

$$\begin{aligned} m_\gamma(x) &= \frac{e^{-\gamma x/2}}{2\pi[x(1-x)]^{1/2}}\left\{\int_0^\infty e^{-\gamma u/2}\left[e^{-\gamma/2}\phi\left(\frac{3}{2},1;\gamma/2\right)(1+u)^{-1}\right.\right. \\ &\quad \left.\left. - u^{\frac{1}{2}}(1+u)^{-\frac{3}{2}}\right] du + e^{-\gamma/2}\phi\left(\frac{3}{2},1;\gamma/2\right)\int_0^x e^{\gamma u/2}(1-u)^{-1}du\right\} \\ &= \frac{e^{-\gamma x/2}}{2\pi[x(1-x)]^{1/2}}\left\{e^{-\gamma/2}\phi\left(\frac{3}{2},1;\gamma/2\right)\psi(1,1;\gamma/2)-\Gamma\left(\frac{3}{2}\right)\psi\left(\frac{3}{2},1;\gamma/2\right)\right. \\ &\quad \left. + \phi\left(\frac{3}{2},1;\gamma/2\right)\left[\int_{1-x}^\infty v^{-1}e^{-\gamma v/2}dv - \int_1^\infty v^{-1}e^{-\gamma v/2}dv\right]\right\} \\ &= \frac{1}{2\pi[x(1-x)]^{\frac{1}{2}}}\left\{e^{-\gamma/2}\phi\left(\frac{3}{2},1;\gamma/2\right)\psi(1,1;\gamma(1-x)/2)\right. \\ &\quad \left. - e^{-\gamma x/2}\Gamma\left(\frac{3}{2}\right)\psi\left(\frac{3}{2},1;\gamma/2\right)\right\}. \end{aligned} \tag{5.84}$$

We now prove (5.70) (b) to establish that $m_\gamma(x)$ is an honest probability density function. From (5.73), (5.72) and (5.82) we get

$$\begin{aligned} &\int_0^1 \frac{1}{2\pi[x(1-x)]^{\frac{1}{2}}}\psi(1,1;\gamma(1-x)/2)\,dx \\ &\quad = \int_0^1 \frac{1}{2\pi[x(1-x)]^{\frac{1}{2}}}\psi(1,1;\gamma x/2)\,dx \\ &\quad = \int_0^1 \frac{dx}{2\pi[x(1-x)]^{\frac{1}{2}}}\int_0^\infty (1+t)^{-1}e^{-\gamma xt/2}\,dt \\ &\quad = \frac{1}{2}\int_0^\infty (1+t)^{-1}\phi\left(\frac{1}{2},1;\gamma t/2\right)e^{-\gamma t/2}dt \\ &\quad = \frac{1}{2}\Gamma\left(\frac{1}{2}\right)\psi\left(\frac{1}{2},1;\gamma/2\right) \end{aligned} \tag{5.85}$$

From (5.85), (5.84) and (5.72) we get

$$H(z) = \frac{4\,e^z}{\Gamma\left(\frac{1}{2}\right)} \int_0^1 m_{2z}(x)\,dx$$
$$= 2\phi\left(\frac{3}{2},1;z\right)\psi\left(\frac{1}{2},1;z\right) - \phi\left(\frac{1}{2},1;z\right)\psi\left(\frac{3}{2},1;z\right)$$

where $2z = \gamma$.

We now show that

$$H'(z) - H(z) = 0\,, \tag{5.86}$$

from which it follows that

$$H(z) = c\,e^z$$

for some constant c.

By direct evaluation in terms of elementary integrals when $\gamma = 0$ we get (using (5.85))

$$\int_0^1 m_0(x)\,dx = 1\,;$$

hence, (5.70) (b) follows from (5.86). To prove (5.86), we use (5.75) and (5.76), which yield

$$\begin{aligned} H'(z) - H(z) &= \phi\left(\frac{3}{2},2;z\right)\psi\left(\frac{1}{2},1;z\right) + 2\phi\left(\frac{3}{2},1;z\right)\psi\left(\frac{1}{2},1;z\right) \\ &\quad - 2\phi\left(\frac{3}{2},1;z\right)\psi\left(\frac{1}{2},2;z\right) + \frac{1}{2}\phi\left(\frac{1}{2},2;z\right)\psi\left(\frac{3}{2},1;z\right) \\ &\quad - \phi\left(\frac{1}{2},1;z\right)\psi\left(\frac{3}{2},1;z\right) + \phi\left(\frac{1}{2},1,z\right)\psi\left(\frac{3}{2},2,z\right). \end{aligned} \tag{5.87}$$

To this expression we now add the following four left-hand side expressions, each of which equals zero

$$\psi\left(\frac{3}{2},1;z\right)\left[-\frac{1}{2}\ \phi\left(\frac{1}{2},2;z\right) - \frac{1}{2}\phi\left(\frac{3}{2},2;z\right) + \phi\left(\frac{1}{2},1;z\right)\right]\,;$$
$$\psi\left(\frac{3}{2},2,z\right)\left[\phi\left(\frac{3}{2},1;z\right) - \phi\left(\frac{1}{2},1;z\right) - z\phi\left(\frac{3}{2};2,z\right)\right]\,;$$
$$2\phi\left(\frac{3}{2},1,z\right)\left[\psi\left(\frac{1}{2},2;z\right) - \frac{1}{2}\psi\left(\frac{3}{2},2;z\right) - \psi\left(\frac{1}{2},1;z\right)\right]\,;$$
$$-\phi\left(\frac{3}{2},2;z\right)\left[-\frac{1}{2}\psi\left(\frac{3}{2},1;z\right) - z\psi\left(\frac{3}{2},2;z\right) + \psi\left(\frac{1}{2},1;z\right)\right]\,;$$

to obtain $H' - H = 0$, as desired.

We now verify (5.70) (c). From (5.72) and (5.79), with $a = \frac{3}{2}$, $b = 1$ we obtain

$$\int_0^1 \frac{(1+\gamma y)\ e^{\gamma y/2}}{2\pi[y(1-y)]^{1/2}}\ dy = \phi\left(\frac{1}{2}, 1; \gamma/2\right) + \gamma\phi\left(\frac{3}{2}, 2; \gamma/2\right)/4$$
$$= \frac{1}{2}\phi\left(\frac{3}{2}, 1; \gamma/2\right). \tag{5.88}$$

Using (5.70) (b), which we have just proved, and (5.85) and (5.88) we rewrite (5.70) (c) as

$$\begin{aligned}
1 &= \left[\phi\left(\frac{3}{2}, 1; \gamma/2\right)\right]^{-1} \int_0^1 [1+\gamma(1-x)]\ m_\gamma(x)\ dx \\
&= \left[\phi\left(\frac{3}{2}, 1; \gamma/2\right)\right]^{-1} \int_0^1 \frac{(1+\gamma y)}{2\pi[y(1-y)]^{1/2}} \left\{\phi\left(\frac{3}{2}, 1; \gamma/2\right)\psi\left(1, 1, \frac{\gamma y}{2}\right)\right. \\
&\quad \left. - e^{\gamma y/2}\Gamma\left(\frac{3}{2}\right)\psi\left(\frac{3}{2}, 1; \gamma/2\right)\right\} dy \\
&= \int_0^1 \frac{\gamma y\ \psi(1, 1; \gamma y/2)}{2\pi[y(1-y)]^{1/2}}\ dy + \int_0^1 \frac{\psi(1, 1;\ \gamma y/2)}{2\pi[y(1-y)]^{1/2}}\ dy \\
&\quad - \frac{\Gamma\left(\frac{3}{2}\right)\ \psi\left(\frac{3}{2}, 1; \gamma/2\right)}{\phi\left(\frac{3}{2}, 1; \gamma/2\right)} \int_0^1 \frac{(1+\gamma y)\ e^{\gamma y/2}}{2\pi[y(1-y)]^{1/2}}\ dy \\
&= \int_0^1 \frac{\gamma y \psi(1, 1; \gamma y/2)}{2\pi[y(1-y)]^{1/2}}\ dy \\
&\quad + \frac{1}{2}\Gamma\left(\frac{1}{2}\right)\psi\left(\frac{1}{2}, 1; \gamma/2\right) - \frac{1}{2}\Gamma\left(\frac{3}{2}\right)\psi\left(\frac{3}{2}, 1; \gamma/2\right).
\end{aligned} \tag{5.89}$$

Using (5.81), with $a = 1$, $b = 0$ and (5.72), (5.73), (5.83) we get

$$\begin{aligned}
\int_0^1 \frac{\gamma y \psi(1, 1; \gamma y/2)}{2\pi[y(1-y)]^{1/2}}\ dy &= \int_0^1 \frac{[\psi(0, 0; \gamma y/2) - \psi(1, 0; \gamma y/2)]}{\pi[y(1-y)]^{1/2}}\ dy \\
&= 1 - \int_0^1 \frac{dy}{\pi[y(1-y)]^{1/2}} \int_0^\infty (1+t)^{-2} e^{-\gamma y t/2}\ dt \\
&= 1 - \int_0^\infty (1+t)^{-2}\phi\left(\frac{1}{2}, 1; \gamma t/2\right) e^{-\gamma t/2}\ dt \\
&= 1 + \Gamma\left(\frac{1}{2}\right)\left[\psi\left(\frac{3}{2}, 1; \gamma/2\right)/2 - \psi\left(\frac{1}{2}, 1; \gamma/2\right)\right]/2.
\end{aligned} \tag{5.90}$$

Thus (5.89) and (5.90) imply (5.31) and, hence (5.70) (c).

Case p = 2, $N = 4$. Solution of Linnik, Pliss and Ŝalaevski

From (5.62) the problem consists of showing (with $\beta_1 = \beta$) on the interval $0 < \beta < 1$ there exists a probability density function $\xi(\beta)$ for which

$$\int_0^1 e^{t_1(1-\beta)/2}\phi\left(2, \frac{1}{2}; t_1\beta/2\right)\phi\left(\frac{3}{2}, \frac{1}{2}; (\gamma - t_i)(1-\beta)/2\right)\xi(\beta)d\beta$$
$$= \phi(2, 1; \gamma/2)\,. \qquad (5.91)$$

With a view to solving this problem we now consider the equation

$$\int_0^1 e^{t_1(1-\beta)/2}\phi\left(2, \frac{1}{2}; t_1\beta/2\right)\phi\left(\frac{3}{2}, \frac{1}{2}; (\gamma - t_1)(1-\beta)/2\right)\xi(\beta)d\beta$$
$$= \int_0^1 \phi\left(\frac{3}{2}, \frac{1}{2}; \gamma(1-\beta)/2\right)\xi(\beta)d\beta\,. \qquad (5.92)$$

Using the fact that

$$\phi\left(\frac{3}{2}, \frac{1}{2}; \frac{x}{2}\right) = (1 + x)\, e^{x/2}$$

we rewrite (5.92) as

$$\int e^{-\gamma\beta/2}\phi\left(2, \frac{1}{2}; \frac{t_1\beta}{2}\right)[1 + \gamma - \gamma\beta - t_1(1-\beta)]\xi(\beta)d\beta$$
$$= \int_0^1 e^{-\gamma\beta/2}\,[1 + \gamma - \gamma\beta]\,\xi(\beta_1)d\beta\,. \qquad (5.93)$$

From (5.93) we get

$$\int_0^1 e^{-\gamma\beta/2}[(r - 2r^2)\beta^{r-1} + (1 + \gamma + \gamma r + 2r^2)\beta^r$$
$$- (\gamma + \gamma r)\beta^{r+1}]\xi(\beta)d\beta = 0\,, \qquad \text{for } r = 1, 2, \ldots\,. \qquad (5.94)$$

Let

$$e^{\gamma\beta/2}f(\beta) = \sum_{r=1}^{\infty} a_{r-1}\beta^{r-1}$$

be an arbitrary function represented by a series with radius of convergence greater than one. Multiplying the rth equation in (5.94) by a_{r-1} and taking the sum over r from 1 to ∞ we get

$$\int_0^1 \Bigg[(2\beta^3 - 2\beta^2)f'' + (-5\beta + 6\beta^2 - \gamma\beta^2 + \gamma\beta^3)f'$$
$$+ \left(-1 + 3\beta - \frac{\gamma\beta}{2} + \gamma\beta^2 \right) f \Bigg] \xi(\beta) d\beta$$
$$= \int_0^1 L(f)\xi(\beta)\, d\beta$$
$$= 0 \tag{5.95}$$

where $L(f)$ denotes the differential form inside the square brakets. Applying the green formula to $L(f)$ and choosing a small $\epsilon > 0$ we get

$$\int_\epsilon^{1-\epsilon} L(f)\xi(\beta)d\beta = \int_\epsilon^{1-\epsilon} L^*(\xi)f(\beta)d\beta + L[f,\xi]\,|_\epsilon^{1-\epsilon}\,, \tag{5.96}$$

where the adjoint form $L^*(\xi)$ is given by

$$L^*(\xi) = 2\beta^2(\beta - 1)\xi'' + \beta(-3 + 6\beta + \gamma\beta - \gamma\beta^2)\xi'$$
$$+ \left(3\beta + \frac{3\gamma\beta}{2} - 2\gamma\beta^2 \right)\xi \tag{5.97}$$

and the bilinear form $L[f,\xi]$ is given by

$$L[f,\xi] = 2\beta^2(1-\beta)(f\xi' - f'\xi) - [\beta + \gamma\beta^2(1-\beta)]\; f\xi\,. \tag{5.98}$$

Using the thoery of Frobenius (see Ince, 1926) we conclude that

$$\xi_{01} = \sum_{k=0}^{\infty} a_k\beta^k\,, \qquad \xi_{02} = \beta^{-\frac{1}{2}} \sum_{k=0}^{\infty} c_k\; \beta^k \tag{5.99}$$

with $a_0 \neq 0$, $c_0 \neq 0$; are two fundamental solution of $L^*(\xi) = 0$ on the interval (0,1). Similarly

$$\xi_{11} = \sum_{k=0}^{\infty} g_k(1-\beta)^k\,, \qquad \xi_{12} = (1-\beta)^{-\frac{1}{2}} \sum_{k=0}^{\infty} h_k(1-\beta)^k \tag{5.100}$$

with $g_0 \neq 0$, $h_0 \neq 0$ are also two fundamental solutions of the same equation on the interval (0,1). Consequently any arbitrary solution of the equation $L^*(\xi) = 0$ is integrable on [0,1] and we assert that there also exists a solution of the equation $L^*(\xi) = 0$ for which

$$\lim_{\epsilon - 0} L[f,\xi]\,|_\epsilon^{1-\epsilon} = 0\,. \tag{5.101}$$

It can also be shown that any linear combination of the solutions in (5.99) (similarly for solutions in (5.100)) satisfies (5.101) provided that $\xi(\cdot) = \xi_{12}(\cdot)$. Thus $\xi = \xi_{12}$ is a solution of (5.95) and hence of (5.91).

In the following paragraph we show that ξ_{12} does not vanish in the interval (0,1). Write

$$\beta = 1 - x\,, \quad \theta(x) = h_0^{-1}\sqrt{x}\,\xi_{12}(1-x)\,. \tag{5.102}$$

The equation $L^*(\xi) = 0$ becomes

$$2x(1-x)\theta'' + (-1 + 4x - \gamma x + \gamma x^2)\theta' + \frac{3}{2}(1+\gamma x)\theta = 0\,. \tag{5.103}$$

Substituting $u(x) = \theta'(x)/\theta(x)$ we obtain the Riccati equation

$$u' = u^2 - \frac{1 - 4x + \gamma x - \gamma x^2}{2x(1-x)} - \frac{3 + 3\gamma x}{4x(1-x)}\,. \tag{5.104}$$

Let us recall that $\theta(x) = \sum_{k=0}^{\infty}(\frac{h_k}{h_0})x^k$, $|x| < 1$. Assume that $\theta(x)$ vanishes at some point in the interval (0,1). Since $\theta(0) = 1$, among the roots of the equation $\theta(x) = 0$ there will be a least root. Let us denote it by x_0. Then the function $\theta(x)$ is analytic on the circle $|x| < x_0$. From (5.103) we get $\theta'(0) = \frac{3}{2}$ which implies that $u(0) = -\frac{3}{2}$. Hence $u(x)$ cannot vanish in the interval $(0, x_0)$, since at any point where it vanishes we would necessarily have $u'(x) \geq 0$ whereas (5.104) shows that at this point $u'(x) < 0$. The fact that the function $u(x)$ has a negative value implies that it decreases unboundedly as we approach the point x_0 from the left ($u' \neq 0$ since $u \neq 0$) which in turn contradicts (5.104). As a result $\xi = \xi_{12}$ does not vanish in the interval (0,1).

Let us now return to (5.91), which, as we have seen, is satisfied by $\xi = \xi_{12}$. By the method of Laplace transforms it is easy to obtain the relation

$$\begin{aligned}\int_0^1 x^{b-1}(1-x)^{c-b-1}e^{q\,x}\phi(a,b;px)\;\phi(a-b,c-b;q(1-x))\;dx \\ = \frac{\Gamma(b)\Gamma(c-b)}{\Gamma(c)}\phi(a,c;p+q)\,, \quad 0 < b < c\,.\end{aligned} \tag{5.105}$$

Letting $t_1 = \gamma\tau$, and multiplying both sides of (5.91) by $\tau^{-\frac{1}{2}}(1-\tau)^{-\frac{1}{2}}$ and integrating with respect to τ from 0 to 1 we obtain with $a = 2$, $b = \frac{1}{2}$, $c = 1$, $p = \gamma\beta/2$, $q = \gamma(1-\beta)/2$,

$$\int_0^1 \phi\Big(2,1;\frac{\gamma}{2}\Big)\xi(\beta)d\beta = \int_0^1 \phi\Big(\frac{3}{2},\frac{1}{2};\frac{\gamma(1-\beta)}{2}\Big)\xi(\beta)d\beta\,. \tag{5.106}$$

Now to obtain (5.91) it is sufficient to make use of the homogeneity of the above relation and to normalize the function ξ_{12}.

5.3.2. R^2-test

Consider the setting of Example 5.1.5 for the problem of testing $H_0 : \rho^2 = 0$ against the alternatives $H_1 : \rho^2 = \lambda > 0$. We have shown in Chapter 4 that among all tests based on sufficient statistic $(\bar{X}, S)$ of H_0 against the alternatives $\rho^2 > 0$ the R^2-test is UMP invariant under the affine group (G, T) of transformations (g, t), $g \in G$, $t \in T$ operating as $(\bar{X}, S) \to (g\bar{X} + t,\ gSg')$ where g is a nousingular matrix of order p of the form

$$g = \begin{pmatrix} g_{(11)} & 0 \\ 0 & g_{(22)} \end{pmatrix}$$

with $g_{(11)}$ of order 1 and t is a p-vector.

For $p > 2$ this result does not imply the minimax property of R^2-test as the group (G, T) does not satisfy the condition of the Hunt-Stein theorem. As discussed in Example 5.1.5 we assume the mean to be zero and consider only the subgroup $G_T(p)$ for invariance. This subgroup satisfies the conditions of the Hunt-Stein theorem. A maximal invariant under $G_T(p)$ is $R = (R_2, \ldots, R_p)'$ with a single distribution under H_0 but with a distribution which depends continuously on a $(p-2)$-dimensional parameter $\triangle = (\delta_2, \ldots, \delta_p)'$ with $\sum_{j=2}^{p} \delta_j = \rho^2 = \lambda$ under H_1. The Lebesgue density function $f_{\lambda,\eta}(r)$ of R under H_1 is given in (5.29). Because of the compactness of the reduced parameter spaces $\{0\}$ and

$$\Gamma = \left\{ (\delta_2, \ldots, \delta_p)', \quad \delta_i \geq 0, \quad \sum_{2}^{p} \delta_j = \lambda \right\} \tag{5.107}$$

and the continuity of $f_{\lambda,\eta}(r)$ in Γ it follows that every minimax test for the reduced problem in terms of R is Bayes. In particular, the R^2-test which has a constant power on the contour $\rho^2 = \lambda$ and which is also $G_T(p)$ invariant, maximizes the minimum power over H_1 if and only if there is a probability measure on Γ such that for some constant c_1

$$\int \frac{f_{\lambda,\eta}(r)}{f_{0,0}(r)} \xi(d\triangle) \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \ c_1 \tag{5.108}$$

according as

$$\sum_{2}^{p} r_i \left\{ \begin{matrix} > \\ = \\ < \end{matrix} \right\} \ c$$

except possibly for a set of measure zero. An examination of the integrand in (5.108) allows us to replace it by the equivalent

$$\int \frac{f_{\lambda,\eta}(r)}{f_{0,0}(r)}\xi(d\triangle) = c_1 \qquad \text{if } \sum_2^p r_i = c\,. \tag{5.109}$$

Obviously (5.108) implies (5.109). On the other hand if there is a ξ and a constant c_1 for which (5.109) is satisfied and if $\bar{r} = (\bar{r}_2, \ldots, \bar{r}_p)'$ such that $\sum_2^p \bar{r}_i = c' > c$, then writing $f = \frac{f_{\lambda,\eta}}{f_{0,0}}$ and $r^* = c\bar{r}/c'$ we conclude that

$$f(\bar{r}) = f(c'r^*/c) > f(r^*) = c_1$$

because of the form of f and the fact that $c'/c > 1$ and $\sum_2^p r_i^* = c$. Note that in (5.29) $\gamma_i^{-1}(1-\lambda) - 1 = -\sum_{j>1}\delta_j/\gamma_i$ and that $\gamma_i > 0$. This and similar argument for the case $c' < c$ show that (5.108) implies (5.109). The remaining computations in this section is somewhat simplified by the fact that for fixed c and λ we can at this point compute the unique value of c_1 for which (5.109) can possibly be satisfied.

Let $\hat{R} = (R_2, \ldots, R_{p-1})'$ and write $f_{\lambda,\eta}(\hat{r}|u)$ for the version of conditional Lebesgue density of $\hat{R}$ given that $\sum_2^p R_i = U = u$ which is continuous in $\hat{r}$ and for $r_i > 0$ and $\sum_1^{p-1} r_i < u < 1$ and zero elsewhere. Also write $f_\delta^*(u)$ for the Lebesgue density of $R^2 = \sum_2^p R_i = U$ which is continuous for $0 < u < 1$ and vanishes elsewhere and depends on $\triangle$ only through $\delta = \sum_2^p \delta_j$. Then (5.109) can be written as

$$\int f_{\lambda,\eta}(\hat{r}|c)\xi(d\triangle) = \left[c_1 \frac{f_0^*(c)}{f_\lambda^*(c)}\right]\ f_{0,0}(\hat{r}|c) \tag{5.110}$$

for $r_i > 0$ and $\sum_2^{p-1} r_i < c$. The integral in (5.110), being a probability mixture of probability densities, is itself a probability density in $\hat{r}$, as is $f_{0,0}^*(\hat{r}|c)$. Hence the expression in square brackets equals one. From (4.35) with $0 < c < 1$

$$\begin{aligned} f_\lambda^*(c) = {} & \frac{(1-\lambda)^{\frac{1}{2}(N-1)}\ \Gamma(\frac{1}{2}(N-1))}{\Gamma((N-p)/2)\ \Gamma((p-1)/2)}\ c^{\frac{1}{2}(p-3)}(1-c)^{\frac{1}{2}(N-p-2)} \\ & \cdot F\left(\frac{1}{2}(N-1),\ \frac{1}{2}(N-1); \frac{1}{2}(p-1); c\lambda\right) \end{aligned} \tag{5.111}$$

where $F(a,b;c;x)$ is the ordinary ${}_2F_1$ hypergeometric series given by

$$\begin{aligned} F(a,b;c;x) &= \sum_{r=0}^{\infty} \frac{\Gamma(a+r)\Gamma(b+r)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+r)} x^r \\ &= \sum_r \frac{(a)_r\ (b)_r}{(c)_r} x^r \end{aligned} \tag{5.112}$$

where we write $(a)_r = \Gamma(a+r)/\Gamma(a)$. From (5.110) the value of c_1 which satisfies (5.109) is given by

$$\begin{aligned} c_1 = &\frac{(1-\lambda)^{\frac{1}{2}(N-1)}\ \Gamma(\frac{1}{2}(N-1))}{\Gamma(\frac{1}{2}(N-p))\ \Gamma(\frac{1}{2}(p-1))}\ c^{\frac{1}{2}(p-3)}(1-c)^{\frac{1}{2}(N-p-2)} \\ &\cdot F\left(\frac{1}{2}(N-1), \frac{1}{2}(N-1); \frac{1}{2}(p-1); c\lambda\right). \end{aligned} \tag{5.113}$$

From (5.113) and (5.29) the condition (5.109) becomes

$$\begin{aligned} &\int_\Gamma \left[1 + \sum_2^p r_i(1-\lambda)/\gamma_i - 1)\right]^{\frac{1}{2}(N-1)} \sum_{\beta_2=0}^{\infty} \cdots \sum_{\beta_p=0}^{\infty} \frac{\Gamma(\sum_2^p \beta_j + \frac{1}{2}(N-1))}{\Gamma(\frac{1}{2}(N-1))} \\ &\quad \times \prod_{j=2}^{p} \frac{\Gamma(\frac{1}{2}(N-j+1)+\beta_j)}{\Gamma(\frac{1}{2}(N-j+1))\ (2\beta_j)!} \left[\frac{4r_j(1-\lambda)/\gamma_j(1+\pi_j^{-1})}{1+\sum_2^p r_j((1-\lambda)/\gamma_j - 1)}\right]^{\beta_j} \xi(d\triangle) \\ &= F\left(\frac{1}{2}(N-1),\ \frac{1}{2}(N-1);\ \frac{1}{2}(p-1); c\lambda\right) \end{aligned} \tag{5.114}$$

for all r with $r_i > 0$ and $\sum_2^p r_i = c$.

Case p = 3, N = 4 (N = 3 if mean known). Solution of Giri and Kiefer

In this case (5.113) can be written as

$$\begin{aligned} &\int_\Gamma \sum_{n=0}^{\infty} \left\{ \frac{(1+2n)}{(1-r_2\lambda)^{\frac{3}{2}}} \left(\frac{r_3\ \delta_3}{(1-\delta_3)(1-r_2\lambda)}\right)^n \right. \\ &\qquad \left. + \frac{r_2\delta_2(1-\lambda)(2n+1)(2n+3)}{(1-\delta_2)(1-r_2\lambda)^{\frac{5}{2}}} \left(\frac{r_3\ \delta_3}{(1-\delta_2)(1-r_2\lambda)}\right)^n \right\} \xi(d\triangle) \\ &= F\left(\frac{3}{2}, \frac{3}{2}; 1; c\lambda\right). \end{aligned} \tag{5.115}$$

One could presumably solve for ξ by using the theory Meijer transforms with kernal $F(\frac{3}{2}, \frac{3}{2}; 1; x)$. Instead they proceeded as for Hotelling's T^2 problem, by expanding (5.115) in a power series. Let Γ_1 be the unit one-demensional simplex

$$\Gamma_1 = \left\{(\beta_1, \beta_2) : \beta_i \geq 0, \sum_1^2 \beta_i = 1\right\}$$

and make the change of variables

$$\begin{aligned}
t_1 &= \frac{r_2}{1-r_2}, & t_1 + t_2 &= \frac{r_2 + r_3}{1 - r_2 - r_3}, \\
\eta_1 &= \frac{\delta_2}{1-\delta_2}, & \eta_1 + \eta_2 &= \frac{\lambda}{1-\lambda} = \lambda'(\text{say}), \\
\beta_i &= \frac{\eta_i}{\lambda}, & i &= 1, 2, \\
c^* &= \frac{c}{1-c}, & y &= t_2\lambda'/(1 + t_1 + \lambda')(1 + c^*).
\end{aligned} \tag{5.116}$$

Let ξ^* be a measure for β_2 on Γ_1 associated with ξ in the obvious way. Denote by μ_i

$$\mu_i = \int_0^1 (\beta_2)^i \xi^*(d\beta_2), \quad i = 1, 2, \ldots$$

the ith moment of ξ^*. Writing $z = c\lambda = c^*\lambda'/(1 + c^*)(1 + \lambda')$ we obtain from (5.115)

$$\begin{aligned}
&(1-z)\sum_{n=0}^{\infty} y^n(2n+1)\mu_n + (z - y)\sum_{n=0}^{\infty} y^n(2n+1)(2n+3)(\mu_n - \mu_{n+1}) \\
&= (1-z)^{\frac{5}{2}} F\left(\frac{3}{2}, \frac{3}{2}; 1; z\right)(1-y)^{-3/2}.
\end{aligned} \tag{5.117}$$

Let $B_z = (1-z)^{\frac{5}{2}} F(\frac{3}{2}, \frac{3}{2}; 1; z)$. Equating coefficients of like powers of y on both sides of (5.117) we obtain the following set of equations as equivalent to (5.115)

$$\begin{aligned}
&\text{(a)}\ 1 + 2z - 3z\mu_1 = B_z \\
&\text{(b)}\ -(2n-1)\mu_{n-1} + (2n + z(2n+2))\mu_n - z(2n+3)\mu_{n+1} \\
&\quad = B_z \tfrac{\Gamma(n+\frac{1}{2})}{\Gamma(\frac{1}{2})n!}, \ n \geq 1.
\end{aligned} \tag{5.118}$$

From this it is clear that ξ^*, if it exists, depends on c and λ only through their product. One could now try to show that the sequence $\{\mu_i\}$ defined by (5.118) with $\mu_0 = 1$ (for ξ^* to be a probability measure) satisfies the classical necessary and sufficient conditions for it to be a moment sequence of a genuine probability measure on [0,1], or, equivalently that the Laplace transform $\sum_0^\infty \mu_j(-t)^j/j!$ is completely monotone on $[0,\infty)$. Unfortunately we have been unable to proceed successfully in this way. We now obtain a function $m_z(x)$, which we then prove to be the Lebesgue density $d\xi^*(x)/dx$ of an absolutely continuous probability measure ξ^* satisfying (5.118) and hence (5.115). That proof does not rely on the somewhat heuristic development which follows, but we nevertheless sketch that development to give an idea of where the $m_z(x)$ of (5.123) comes from.

The generating function

$$\phi(t) = \sum_{j=0}^{\infty} \mu_j t^j$$

of the sequence $\{\mu_j\}$ satisfies a differential equation, which is obtained from (5.118) by multiplying it by t^{n-1} and summing with respect to n from 1 to ∞,

$$\begin{aligned} &2(1-t)(t-z)\phi'(t) - t^{-1}(t^2 - 2zt + z)\phi(t) \\ &= B_z(1-t)^{-\frac{1}{2}} - 1 - zt^{-1}\,. \end{aligned} \tag{5.119}$$

Solving (5.119) by treatment of the corresponding homogeneous equation and by variation of parameter, we get

$$\begin{aligned} \phi(t) = &\left[\frac{t-z}{t(1-t)}\right]^{\frac{1}{2}} \int_0^t \left[\frac{B_z \tau^{\frac{1}{2}}}{2(1-\tau)\,(\tau-z)^{\frac{3}{2}}}\right. \\ &\left. - \frac{\tau^{\frac{1}{2}}}{(\tau-z)^{\frac{3}{2}}\,(1-z)^{\frac{1}{2}}} + \frac{1}{2[\tau(1-\tau)\,(\tau-z)]^{\frac{1}{2}}}\right] d\tau\,. \end{aligned} \tag{5.120}$$

The constant of integration has been chosen to make $\phi(t)$ continuous at $t=0$ with $\phi(0) = 1$ and (5.120) defines a single valued analytic function on the complex plane cut from 0 to z and from 1 to ∞. If there did exist an absolutely continuous ξ^* whose suitably regular derivative m_z satisfied

$$\int_0^1 \frac{m_z(x)}{(1-tx)} dx = \phi(t)\,, \tag{5.121}$$

we could obtain m_z by using the simple inversion formula

$$m_z(x) = \frac{1}{2\pi i x} \lim_{\varepsilon \downarrow 0} \left[\phi(x^{-1} + i\varepsilon) - \phi(x^{-1} - i\varepsilon)\right]. \tag{5.122}$$

Since there is nothing in the theory of Stieltjes transforms which tells us that an m_z satisfying (5.122) does satisfy (5.121), we use (5.122) a formal device to obtain m_z which we shall prove satisfies (5.115). From (5.120) and (5.122) we obtain, for $0 < x < 1$,

$$\begin{aligned} m_z(x) = \frac{(1-zx)^{\frac{1}{2}}}{2\pi\ x^{\frac{1}{2}}(1-x)^{\frac{1}{2}}} \biggl\{ & B_z \int_0^x \frac{du}{(1-u)(1-zu)^{\frac{3}{2}}} \\ & + \int_0^\infty \biggl[\frac{B_z u^{\frac{1}{2}}}{(1+u)(z+u)^{\frac{3}{2}}} + \frac{1}{[u(1+u)(z+u)]^{\frac{1}{2}}} \\ & - 2 \frac{u^{\frac{1}{2}}}{(1+u)^{\frac{1}{2}}(z+u)^{\frac{3}{2}}} \biggr] \, du \biggr\} \\ = \frac{(1-zx)^{\frac{1}{2}}}{2\pi(x(1-x))^{\frac{1}{2}}} & \{B_z Q_z(x) + c_z\} \text{ (say)}. \end{aligned} \tag{5.123}$$

We can evaluate c_z by making the change of variables $v = (1+u)^{-1}$ and using (5.126) below.

We obtain

$$\begin{aligned} c_z = \frac{2}{3}\ B_z F\left(\frac{3}{2}, 1; \frac{5}{2}; 1-z\right) + \pi\ F\left(\frac{1}{2}, \frac{1}{2}; 1; 1-z\right) \\ - \pi\ F\left(\frac{3}{2}, \frac{1}{2}; 2; 1-z\right) \end{aligned}$$

and

$$\begin{aligned} Q_z = {} & 2(1-z)^{-1}[1-(1-zx)^{-\frac{1}{2}}] \\ & + (1-z)^{-\frac{3}{2}} \ \log\ \left[\frac{(1-zx)^{\frac{1}{2}} + (1-z)^{\frac{1}{2}}}{(1-zx)^{\frac{1}{2}} - (1-z)^{\frac{1}{2}}} \cdot \frac{1-(1-z)^{\frac{1}{2}}}{1+(1-z)^{\frac{1}{2}}} \right]. \end{aligned}$$

Now to show that $\xi^*(dx) = m_z(x)\ dx$ satisfies (5.115) with ξ^* a probability measure we must show that

$$\begin{aligned} &\text{(a)}\ \ m_z(x) \geq 0 \text{ for allmost all } x,\ 0 < x < 1; \\ &\text{(b)}\ \ \textstyle\int_0^1 m_z(x)\ dx = 1\,; \\ &\text{(c)}\ \ \mu_1 = \textstyle\int_0^1 x m_z(x)\ dx \text{ satisfies (5.118a)}; \\ &\text{(d)}\ \ \mu_n = \textstyle\int_0^1 x^n\ m_z(x)\ dx \text{ satisfies (5.118b) for } n \geq 1\,. \end{aligned} \tag{5.124}$$

The first condition will follow from (5.123) and the positivity of B_z and c_z for $0 < z < 1$. The former is obvious. To prove the positivity of c_z, we first note that

$$F\left(\frac{3}{2}, \frac{3}{2}; 1; z\right) \geq (1-z)^{-2};$$

this is seen by comparing the two power series, the coefficient of z^j being $[(\frac{3}{2})_j/j!]^2$ and $(j+1)$, and the ratio of the former to the latter being $\prod_{i=1}^{j}(i+\frac{1}{2})^2/i(i+1) \geq 1$. Thus we have $B_z \geq 1-z$. Substituting this lower bound into the expression for c_z and writing $u = 1-z$ the resulting lower bound for c_z has a power series in u (convergent for $|u| < 1$) whose constant term is zero and whose coefficent of u^j for $j \geq 1$ is $[(j+\frac{1}{2})^{-1} - \Gamma^2(j+\frac{1}{2})/\Gamma(j)\Gamma(j+1)(j+1)]$. Using the logarithmic convexity of the Γ-function, i.e. $\Gamma^2(j+\frac{1}{2}) < \Gamma(j)\Gamma(j+1)$, the coefficient of u^j for $j \geq 1$ is $> (j+\frac{1}{2})^{-1} - (j+1)^{-1} > 0$ Hence $c_z > 0$ for $0 < z < 1$.

To prove (5.124)(d) we note that $m_z(x)$ defined by (5.118) satisfies the differential equation

$$m_z'(x) + \frac{1}{2}m_z(x)\left[\frac{1-2x+zx^2}{x(1-x)(1-zx)}\right] = B_z/2\pi x^{\frac{1}{2}}(1-x)^{\frac{3}{2}}(1-zx)\,, \quad (5.125)$$

so that an integration by parts yields, $n \geq 1$

$$\begin{aligned}
&-z(n+2)\mu_{n+1} + (1+z)(n+1)\mu_n - n\mu_{n-1} \\
&= \int_0^1 \{-z(n+2)x^{n+1} + (1+z)(n+1)x^n - nx^{n-1}\}\, m_z(x)\, dx \\
&= \int_0^1 x^n(1-(1+z)x+zx^2)\, m_z'(x)\, dx \\
&= \frac{1}{2}\left\{-\mu_{n-1} + 2\mu_n - z\mu_{n+1} + B_z\frac{\Gamma(x+\frac{1}{2})}{n!\,\Gamma(\frac{1}{2})}\right\},
\end{aligned}$$

which is (5.118)(b).

The proofs of (5.124) (b) and (c) depends on the following identities involving hypergeometric functions $F(a,b;c;x)$ which has the following integral representation when $Re(c) > Re(b) > 0$:

$$F(a,b;c;x) = \frac{\Gamma(c)}{\Gamma(b)\,\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tx)^{-a}dt\,. \qquad (5.126)$$

We will also use the representation

$$2x\, F\left(\frac{1}{2}, 1; \frac{3}{2}; x\right) = \log\left(\frac{1+x}{1-x}\right) \qquad (5.127)$$

and the identities

$$F(a,b;c;x) = F(b,a;c;x)\,, \tag{5.128}$$

$$\begin{aligned}(c-a-1)\ F(a,b;c;x) + aF(a+1,b;c;x)&\\ -(c-1)\ F(a,b;c-1;x) = 0\,,&\end{aligned} \tag{5.129}$$

$$\begin{aligned}&\lim_{c\to -n}[\Gamma(c)]^{-1}F(a,b;c;x)\\ &\quad = \frac{(a)_{n+1}\ (b)_{n+1}}{(n+1)!}x^{n+1}\ F(a+n+1,b+n+1;n+2;x)\\ &\qquad\qquad\qquad \text{for } n = 0,1,2,\dots\,,\end{aligned} \tag{5.130}$$

$$\begin{aligned}&\frac{\Gamma(1+\lambda+\mu)\Gamma(1+\nu+\mu)}{\Gamma(\lambda+\mu+\nu+\frac{3}{2})\Gamma(\frac{1}{2}+\mu)}\\ &\quad = F\left(\frac{1}{2}+\lambda,-\frac{1}{2}-\nu;1+\lambda+\mu;x\right)F\left(\frac{1}{2}-\lambda,\frac{1}{2}+\nu;1+\nu+\mu;1-x\right)\\ &\qquad + F\left(\frac{1}{2}+\lambda,\frac{1}{2}-\nu;1+\lambda+\mu;x\right)\ F\left(-\frac{1}{2}-\lambda,\frac{1}{2}+\nu;1+\nu+\mu;1-x\right)\\ &\qquad - F\left(\frac{1}{2}+\lambda,\frac{1}{2}-\nu;1+\lambda+\mu;x\right)\ F\left(\frac{1}{2}-\lambda,\frac{1}{2}+\nu;1+\nu+\mu;1-x\right);\end{aligned}$$

$$F(a,b;c;x) = (1-x)^{c-a-b}F(c-a,c-b;c;x)\,. \tag{5.131}$$

We now prove (5.124) (b) and (c). From (5.123), using (5.126) and (5.127) we obtain

$$\begin{aligned}\int_0^1 m_z(x)\ dx &= (1-z)^{\frac{3}{2}}F\left(\frac{3}{2},\frac{3}{2};1;z\right)\ F\left(-\frac{1}{2},\frac{1}{2};1;z\right)\\ &\quad\times\left[1-F\left(\frac{1}{2},1;\frac{3}{2};1-z\right)+\frac{1}{3}(1-z)\ F\left(\frac{3}{2},1;\frac{5}{2};1-z\right)\right]\\ &\quad -(1-z)^{\frac{3}{2}}\ F\left(\frac{3}{2},\frac{3}{2};1;z\right)+(\pi/2)\ \left[F\left(\frac{1}{2},\frac{1}{2};1;1-z\right)\right.\\ &\quad \left. -F\left(\frac{3}{2},\frac{1}{2};2;1-z\right)\times F\left(-\frac{1}{2},\frac{1}{2};1;z\right)\right.\\ &\quad + B_z\int_0^1 \frac{(1-zx)^{\frac{1}{2}}}{2\pi(1-z)^{\frac{3}{2}}\ x^{\frac{1}{2}}\ (1-x)^{\frac{1}{2}}}\log\left(\frac{1+(i-z)^{\frac{1}{2}}(i-zx)^{-\frac{1}{2}}}{1-(1-z)^{\frac{1}{2}}(1-zx)^{-\frac{1}{2}}}\right)\ dx\,.\end{aligned} \tag{5.132}$$

The first expression in square brackets in (5.132) varishes, as is easily seen from the power series of F in (5.111). Using (5.127), the integral in (5.132) can be written as

$$\frac{1}{\pi(1-z)}\sum_{n=0}^{\infty}\frac{(1-z)^n}{2n+1}\int_0^1\frac{dx}{x^{\frac{1}{2}}\,(1-x)^{\frac{1}{2}}(1-zx)^n}$$

$$=(1-z)^{-1}+(1-z)^{-1}\sum_{n=1}^{\infty}\frac{(1-z)^n}{2n+1}\,F\left(n,\frac{1}{2};1;z\right)$$

$$=(1-z)^{-1}+\sum_{n=0}^{\infty}\frac{(1-z)^n}{2n+3}F\left(n+1,\frac{1}{2};1;z\right)$$

$$=(1-z)^{-1}+\frac{1}{2}\sum_{m=0}^{\infty}\frac{z^m(\frac{1}{2})m}{(m!)^2}\sum_{n=0}^{\infty}\frac{(n+m)!\,(1-z)^n}{n!\,(n+\frac{3}{2})}$$

$$=(1-z)^{-1}+\frac{1}{2}\sum_{m=0}^{\infty}\frac{z^m\,(\frac{1}{2})_m}{(m!)^2}\,(1-z)^{-\frac{3}{2}}\int_0^{1-z}\frac{m!\,t^{\frac{1}{2}}}{(1-t)^{m+1}}\,dt$$

$$=(1-z)^{-1}+\frac{1}{2}(1-z)^{-\frac{3}{2}}\int_0^{1-z}\frac{t^{\frac{1}{2}}\,dt}{(1-t)^{\frac{1}{2}}\,(1-z-t)^{\frac{1}{2}}}$$

$$=(1-z)^{-1}+(\pi/4)\,(1-z)^{-\frac{1}{2}}F\left(\frac{1}{2},\frac{3}{2};2;1-z\right). \tag{5.133}$$

From (5.132) and (5.133) we get

$$\begin{aligned}\int_0^1 m_z(x)\,dx &= (\pi/2)\,F\left(-\frac{1}{2},\frac{1}{2};1;z\right)\left[F\left(\frac{1}{2},\frac{1}{2};1;1-z\right)\right.\\ &\qquad\left.-F\left(\frac{3}{2},\frac{1}{2};2;1-z\right)\right]\\ &\qquad+(\pi/4)(1-z)^2\,F\left(\frac{3}{2},\frac{3}{2};1;z\right)\,F\left(\frac{3}{2},\frac{1}{2};2;1-z\right).\end{aligned} \tag{5.134}$$

Hence, from (5.129) and (5.133) we obtain from (5.134)

$$\begin{aligned}\int_0^1 m_z(x)\,dx &= (\pi/4)\,\left\{F\left(-\frac{1}{2},\frac{1}{2};1;z\right)\,F\left(\frac{1}{2},\frac{1}{2};2;1-z\right)\right.\\ &\qquad+F\left(\frac{3}{2},\frac{1}{2};2;1-z\right)\,\left[F\left(-\frac{1}{2},-\frac{1}{2};1;z\right)\right.\\ &\qquad\left.\left.-F\left(-\frac{1}{2},-\frac{1}{2};1;z\right)\right]\right\}.\end{aligned} \tag{5.135}$$

Using (5.129) with $a = b = -\frac{1}{2}$, $c = 1 + \varepsilon$ and (5.130) with $n = 0$ and $c = \varepsilon \to 0$, (5.132) reduces to

$$(\pi/4)F\left(-\frac{1}{2},\frac{1}{2};1;z\right)F\left(\frac{1}{2},\frac{1}{2};2;1-z\right)+(z/2)\,F\left(\frac{3}{2},\frac{1}{2};2;1-z\right)F\left(\frac{1}{2},\frac{1}{2};2;z\right) \tag{5.136}$$

Now, by (5.131), the expression (5.133) equals one if we have

$$\begin{aligned} F\left(\frac{3}{2},\frac{1}{2};2;1-z\right)&\left[F\left(-\frac{1}{2},\frac{1}{2};1;z\right)\right.\\ &\left.-F\left(-\frac{1}{2},-\frac{1}{2};1;z\right)+(z/2)\,F\left(\frac{1}{2},\frac{1}{2};2;z\right)\right]=0\,. \end{aligned} \tag{5.134}$$

The expression inside the square brackets is easily seen to be zero by computing the coefficent of z^n. Thus we prove (5.124)(b).

We now verify (5.124) (c). The integrand in (5.132) is unaltered by multiplication by x, and in place of (5.133) we obtain $(1-z)^{-1}/2-z^{-1}/3+[\pi/4z(1-z)^{\frac{1}{2}}]\,F(-\frac{1}{2},\frac{3}{2};2;1-z)$. The analogue of (5.134) is

$$\begin{aligned} \mu_1 &= \int_0^1 x\ m_z(x)\ dx\\ &= (\pi/4)\ F\left(-\frac{1}{2},\frac{3}{2};2;z\right)\left[F\left(\frac{1}{2},\frac{1}{2};1;1-z\right)-F\left(\frac{3}{2},\frac{1}{2},2;1-z\right)\right]\\ &+(\pi/4)(1-z^2)/z\left[F\left(\frac{3}{2},\frac{3}{2};1;z\right)F\left(-\frac{1}{2};\frac{3}{2};2;1-z\right)\right.\\ &-[(1-z)^{\frac{5}{2}}/3z]\ F\left(\frac{3}{2},\frac{3}{2};1;z\right). \end{aligned} \tag{5.135}$$

To verify (5.124) (c) we then have to prove the following identity (by (5.118) (a))

$$\begin{aligned} (1+2z)/3z = (\pi/4)\ F\left(-\frac{1}{2},\frac{3}{2};2;z\right)&\left[F\left(\frac{1}{2},\frac{1}{2};1;1-z\right)\right.\\ &\left.-F\left(\frac{3}{2},\frac{1}{2};2;1-z\right)\right]+(\pi/4)[(1-z)^2/z]\\ &\times F\left(\frac{3}{2},\frac{3}{2};1;z\right)\ F\left(-\frac{1}{2},\frac{3}{2};2;1-z\right). \end{aligned} \tag{5.136}$$

From (5.131) with $c = 1$, $a = b = \frac{3}{2}$, then (5.129) with $a = \frac{1}{2}$, $b = -\frac{1}{2}$, $c = 1$, and then (5.130) with $\lambda = \mu = 0$, $\nu = 1$ we can rewrite (5.136) as

$$\begin{aligned}
&(4/3\pi)(1+2z) \\
&= z\, F\Big(-\frac{1}{2},\frac{3}{2};2;z\Big)\left[F\Big(\frac{1}{2},\frac{1}{2};1;1-z\Big) - F\Big(\frac{3}{2},\frac{1}{2};2;1-z\Big)\right] \\
&\quad + (3z/2)F\Big(\frac{1}{2},-\frac{1}{2};2;z\Big)\, F\Big(-\frac{1}{2},\frac{3}{2};2;1-z\Big) + (4/3\pi)(1-z) \\
&\quad + (1-z)\, F\Big(\frac{1}{2},\frac{3}{2};2;1-z\Big)\left[F\Big(\frac{1}{2},-\frac{1}{2};1;z\Big) - F\Big(\frac{1}{2},-\frac{3}{2};1;z\Big)\right].
\end{aligned} \tag{5.137}$$

Using (5.129) with $a = -\frac{3}{2}$, $b = \frac{1}{2}$, $c = 1+\epsilon$, and then (5.130) with $n = 0$, $c = \epsilon \to 0$, the expression inside the square brackets in the last term of (5.137) can be simplified to $\frac{1}{2}\ z\ F(-\frac{1}{2},\frac{3}{2};2;z)$. Thus we are faced with the problem of establishing the following identity

$$\begin{aligned}
4/\pi &= F\Big(-\frac{1}{2},\frac{3}{2};2;z\Big)\, F\Big(\frac{1}{2},\frac{1}{2};1;1-z\Big) \\
&\quad + \frac{3}{2}F\Big(\frac{1}{2},-\frac{1}{2};2;z\Big)\, F\Big(-\frac{1}{2},\frac{3}{2};2;1-z\Big) \\
&\quad - [(1+z)/2]\, F\Big(\frac{3}{2},-\frac{1}{2};2;z\Big)\, F\Big(\frac{1}{2},\frac{3}{2};2;1-z\Big),
\end{aligned} \tag{5.138}$$

which finally by (5.126) reduces to

$$\begin{aligned}
0 = F\Big(-\frac{1}{2},\frac{3}{2};2;z\Big)&\left[F\Big(\frac{1}{2},\frac{1}{2};1;1-z\Big) + \frac{3}{2}\, F\Big(\frac{3}{2},-\frac{1}{2};2;1-z\Big)\right. \\
&\left. -\frac{3}{2}\, F\Big(\frac{1}{2},-\frac{1}{2};2;1-z\Big) + ((1-z)/2-1)\, F\Big(\frac{3}{2},\frac{1}{2};2;1-z\Big)\right].
\end{aligned} \tag{5.139}$$

The expression inside the square brackets in (5.139) has a power series in $1-z$, the value of which is seen to be zero by computing the coefficents of various power of $1-z$. Hence we prove (5.124) (c).

5.3.3. ε-*Minimax test (Linnik, 1966)*

Consider the setting of Sec. 5.3.1. for Hotelling's T^2 lest for testing H_0 : $\delta = 0$ against alternatives $H_1 : \delta^2 > 0$ at level $\alpha\varepsilon(0,1)$. The Hotelling T^2 test ϕ_N is ε-minimax if

$$\sup_{\phi} \inf_{\theta \in H_1} E_\theta(\phi) - \inf_{\theta \in H_1} E_\theta(\phi_N) \leq \varepsilon \tag{5.139}$$

for all $N > N_0(\varepsilon)$ where $\theta = (\mu, \Sigma)$ and ϕ runs through all tests of level $\leq \alpha$.

For $N \to \infty$, $\lambda = \frac{2\psi}{N}, 0 < \psi \leq 2(\log N)^{\frac{1}{3}}$, (5.26) has the approximate solution

$$d\xi^*(\beta_i, \ldots, \beta_p) = \frac{\Gamma(\frac{p}{2})}{[\Gamma(\frac{1}{2})]^p} (\beta_1 \cdots \beta_p)^{-\frac{1}{2}} \tag{5.140}$$

which is a pdf on the simplex Γ_1.

If we substitute (5.140) in the left side of (5.62) we obtain a discrepancy with its right side which is of the order of $0_\varepsilon(N^{-1+\varepsilon})$, for $\varepsilon > 0$. Thus, if $\alpha = \alpha_N$ satisfies the condition

$$O(\exp(-\log N)^{1/\varepsilon}) \leq \alpha \leq 1 - O(\exp(-\log N)^{1/\varepsilon}) \tag{5.141}$$

and we have

$$\exp(-(\log N)^{1/\varepsilon}) \leq \delta \leq (\log(N))^{1/\varepsilon} \tag{5.142}$$

then

$$\sup_{\phi} \inf_{\theta \varepsilon H_1} E_\theta(\phi) - \inf_{\theta \varepsilon H_1} E_\theta(\phi_N) = 0_\varepsilon(1/N^{1-\varepsilon})$$

for $\varepsilon > 0$ and hence the Hotelling T^2 test is ε-minimax for testing H_0 against $H_1 : \delta > 0$.

Exercises

1. Prove in details (5.65).
2. Prove (5.72) and (5.73).
3. Prove (5.119) and (5.120).
4. Prove (5.127) and (5.130).
5. Prove that in Problem 8 of Chapter 4 no invariant test under (G_T, T_2) is minimax for testing Ho against H_1 for every choice of λ.

References

1. M. Behara, and N. Giri, *Locally and asymptotically minimax test of some multivariate decision problems*, Archiv der Mathmatik, 4, 436–441, (1971).
2. A. Erdelyi, *Higher Transcedental Functions,* McGraw-Hill, N.Y., 1953.
3. J. K. Ghosh, *Invariance in Testing and Estimation*, Publication no SM 67/2, Indian Statistical Institute, Calculta India, (1967).
4. N. Giri, *Multivariate Statistical Inference*, Academic Press, N.Y., (1977).
5. N. Giri, *Locally and asymptotically minimax tests of a multivariate problem.*, Ann. Math. Statist. **39**, 171–178 (1968).
6. N. Giri, and J. Kiefer, *Local and asymptotic minimax properties of multivariate test*, Ann. Math. Statist. **39**, 21–35 (1964).
7. N. Giri, and J. Kiefer, *Minimax character of R^2-test in the simplest case*, Ann. Math. Statist, **34**, 1475–1490 (1964).
8. N. Giri, J. Kiefer, and C. Stein, *Minimax properties of T^2-test in the simplest case*, Ann. Math. Statist. **34**, 1524–1535 (1963).
9. E. L. Ince, *Ordinary Differential Equations*, Longmans, Green and Co., London (1926).
10. W. James, and C. Stein, *Estimation with quadratic loss.*, Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 361–379 (1960).
11. E. L. Lehmann, *Testing Statistical Hypotheses*, Wiley, N.Y. (1959).
12. Ju. V. Linnik, V. A. Pliss and Šalaevski, *On Hotelling's Test*, Doklady, 168, 719–722 (1966).
13. J. Kiefer, *Invariance, minimax sequential estimation, and continuous time processes*, Ann. Math. Statist. **28**, 573–601, (1957).
14. Ju. V. Linnik, *Appoximately minimax detection of a vector signal on a Gaussian background*, Soviet Math. Doklady, 7, 966–968, (1966).
15. M. P. Peisakoff, *Transformation Parameters*, Ph. D. thesis, Princeton University (1950).
16. E. J. G. Pittman, *Tests of hypotheses concerning location and scale parameters*, Biometrika **31**, 200–215 (1939).
17. O. V. Šalaevski, *Minimax character of Hotelling's T^2 test*, Soviet Math. Doklady **3**, 733–735 (1968).
18. J. B. Semika, *An optimum poroperty of two statistical test*, Biometrika **33**, 70–80 (1941).
19. L. J. Slater, *Confluent Hypergeometric Functions*, Cambridge University Press (1960).
20. C. Stein, *The admissibility of Hotelling's T^2 test*, Ann. Math. Statist, **27**, 616–623 (1956).
21. A. Wald, *Contributions to the theory of statistical estimation and testing hypotheses*, Ann. Math. Statist **10**, 299–320, (1939).
22. A. Wald, *Statistical Decision Functions*, Wiley, N.Y. (1950).
23. O. Wesler, *Invariance theory and a modified principle*, Ann. Math. Statist. **30**, 1–20 (1959).

Chapter 6

LOCALLY MINIMAX TESTS IN SYMMETRICAL DISTRIBUTIONS

6.0. Introduction

In multivariate analysis the role of multivariate normal distribution is of utmost importance for the obvious reason that many results relating to the univariate normal distribution have been successfully extended to the multivariate normal distribution. However, in actual practice, the assumption of multinormality does not always hold and the verification of multinormality in a given set of data is, often, very cumbersome, if not impossible. Very often, the optimum statistical procedures derived under the assumption of multivariate normal remain optimum when the underlying distribution is a member of a family of elliptically symmetric distributions.

6.1. Elliptically Symmetric Distributions

Definition 6.1.1. (Elliptically Symmetric Distributions (Univariate)). A random vector $X = (X_1, \cdots, X_p)'$ with values in E^p is said to have a distribution belonging to the family of univariate elliptically symmetric distributions with location parameter $\mu = (\mu_1, \ldots, \mu_p)' \in E^p$ and scale matrix Σ (positive definite) if its probability density function can be expressed as a function of the quadratic form $(x-\mu)'\Sigma^{-1}(x-\mu)$ and is given by

$$f_X(x) = |\Sigma|^{-1/2} q\Big((x-\mu)'\Sigma^{-1}(x-\mu)\Big), \quad x \in E^p, \tag{6.1}$$

where q is a function on $[0,\infty)$ satisfying

$$\int q(y'y)dy = 1$$

for $y \in E^p$.

We shall denote a family of elliptically symmetric distributions by $E_p(\mu, \Sigma)$. It may be verified that

$$E(X) = \mu, \qquad \text{cov}(X) = \alpha\Sigma$$

where $\alpha = p^{-1}E\Big((X-\mu)'\Sigma^{-1}(X-\mu)\Big)$. In other words all distributions in the class $E_p(\mu, \Sigma)$ have the same mean and the same correlation matrix. The family $E_p(\mu, \Sigma)$ contains a class of probability densities whose contours of equal density have the same elliptical shape as the multivariate normal but it contains also long-tailed and short-tailed distributions relative to multivariate normal. This family of distributions satisfies most properties of the multivariate normal. We refer to Giri (1996) for these results.

Definition 6.1.2. (Multivariate Elliptically Symmetric Distribution). A $n \times p$ random matrix

$$X = (X_{ij}) = (X_1, \dots, X_n)'$$

where $X_i = (X_{i1}, \dots, X_{ip})'$ is said to have a multivariate elliptically symmetric distribution with the same location parameter $\mu = (\mu_1, \dots, \mu_p)'$ and the same scale matrix Σ (positive definite) if its probability density function is given by

$$f_X(x) = |\Sigma|^{-n/2} q\left(\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right), \tag{6.2}$$

with $x_i \in E^p, i = 1, \dots, n$ and q is a function on $[0, \infty)$ of the sum of quadratic forms $(x_i-\mu)'\Sigma^{-1}(x_i-\mu)$, $i = 1, \dots, n$.

We have modified the definition 6.1.2. for statistical applications. Elliptically symmetric distributions are becoming increasingly popular because of their frequent use in filtering and stochastic control, random signal input, stock market data analysis and robustness studies of multivariate normal statistical procedures. The family $E_p(\mu, \Sigma)$ includes, among others, the muiltivariate normal, the multivariate Cauchy, the Pearsonian type II and type IV, the Laplance distributions. The family of spherically symmetric distributions which includes the multivariate-t distribution, the contaminated normal and the compound normal distributions is a subfamily of $E_p(\mu, \Sigma)$.

Write $f_X(x)$ in (6.2) as

$$f_X(x) = \beta(\theta)q(\Psi(x|\theta)) \tag{6.3}$$

where $\theta = (\mu, \Sigma) \in \Omega, x \in \chi$ and Ψ is a measurable function from χ onto $\mathcal{Y}$. Let us assume that $\mathcal{Y}$ is a nonempty open subset of R^m and is independent of θ and q is fixed integrable function from $\mathcal{Y}$ to $[0, \infty)$ independently of θ.

Let G be a group of transformations which leaves the problem of testing $H_0 : \theta \in \Omega_{H_0}$ against $H_1 : \theta \in \Omega_{H_1}$ invariant.

Assume that χ is a Cartan G-space and the function ψ satisfies

$$\Psi(gx|\theta) = \bar{g}\Psi(x|\theta) \tag{6.4}$$

for all $x \in \chi$, $g \in G$, $\theta \in \Omega$ and $\bar{g} \in \bar{G}$ where $\bar{G}$ is the induced group of transformations on Ψ corresponding to $g \in G$ on χ and $\bar{G}$ acts transitively on the range space $\mathcal{Y}$ of Ψ.

Using (2.21a) the ratio R of the probability density function of the maximal invariant $T(X)$ under G on χ; with $\theta_1 \in \Omega_{H_1}$, $\theta_0 \in \Omega_{H_0}$; is given by

$$R = \frac{dP^T_{\theta_1}}{dP^T_{\theta_0}} = \frac{\displaystyle\int_G f(gx|\theta_1)\delta(g)\mu(dg)}{\displaystyle\int_G f(gx|\theta_0)\delta(g)\mu(dg)} \tag{6.5}$$

where $\delta(g)$ is the Jacobian of the inverse transformation $x \to gx$.

Theorem 6.1.1. *If $\bar{G}$ acts transitively on the range space $\mathcal{Y}$ of Ψ, then R is independent of q for all $\theta_1 \in \Omega_{H_1}$, $\theta_0 \in \Omega_{H_0}$.*

Proof. Since $\bar{G}$ acts transitively on $\mathcal{Y}$, for any $y_0 \in \mathcal{Y}$ there exists an element $h(x, \theta_1 : y_0)$ of G such that

$$h(x, \theta_1 : y_0)\Psi(x|\theta_1) = y_0\,. \tag{6.6}$$

Using the invariance property of μ and replacing g by $gh(x, \theta_1 : y_0)$ we get

$$\begin{aligned}\beta(\theta_1)\int_G \delta(g)q(\Psi(gx|\theta_1))d\mu(g) &= \beta(\theta_1)\int_G \delta(g)q(\bar{g}\Psi(x|\theta_1))d\mu(g)\\ &= \beta(\theta_1)\delta(h(x,\theta_1 : y_0)^{-1})\Delta_r(h(x,\theta_1 : y_0))\\ &\quad\times\int_G q(\bar{g}y_0)\delta(g)d\mu(g)\end{aligned}$$

where $\Delta_r\ (\cdot)$ is the right modular function of μ. Hence

$$R = \frac{\beta(\theta_1)\delta(h(x,\theta_1 : y_0)^{-1})\Delta_r(h(x,\theta_1 : y_0))}{\beta(\theta_0)\delta(h(x,\theta_0 : y_0)^{-1})\Delta_r(h(x,\theta_0 : y_0))}$$

is independent of q. □

6.2. Locally Minimax Tests in $E_p(\mu, \Sigma)$

In recent years considerable attentions are being focussed on the study of robustness of commonly used test procedures concerning the parameters of multivariate normal populations in the family of symmetric distributions. For an up-to-date reference we refer to Kariya and Sinha (1988). The criterion mostly used in such studies is the locally best invariant (LBI) property of the test procedure. Giri (1988) gave the following formulation of the locally minimax test in $E_p(\mu, \Sigma)$. Let F be the space of values of the maximal invariant $Z = T(X)$ and let the distribution of $T(X)$ depend on the parameter (δ, η), with $\delta > 0$ and $\eta \in R^p$. For each (σ, η) in the parametric space of the distribution of Z suppose that $f(z : \delta, \eta)$ is the probability density function of Z with respect to some σ-finite measure. Suppose the problem of testing $H_0 : \theta \in \Omega_{H_0}$ against $H_1 : \theta \in \Omega_{H_1}$ reduces to that of testing $H_0 : \delta = 0$ against $H_1 := \lambda > 0$, in the presence of the nuisance parameter η, in terms of Z. We are concerned with the local theory in the sense that $f(z : \lambda, \eta)$ is close to $f(z : 0, \eta)$ when λ is small for all q in (6.2). Throughout this chapter notations like $o(1)$, $o(h(\lambda))$ are to be interpreted a $\lambda \to 0$ for an q in (6.2).

For each α, $0 < \alpha < 1$ we now consider a critical region of the form

$$R^* = \{U(x) = U(T(x)) \geq C_\alpha\} \tag{6.7}$$

where U is bounded, positive and has a continous distribution for (δ, η), equicontinuous in (δ, η) with $\delta < \delta_0$ for any q (6.2) and R^* satisfies

$$P_{0,\eta}(R^*) = \alpha, P_{\lambda,\eta}(R^*) = \alpha + h(\lambda) + r(\lambda, \eta) \tag{6.8}$$

for any q in (6.2) where $r(\lambda, \eta) = o(h(\lambda))$ uniformly in η with $h(\lambda) > 0$ and $h(\lambda) = o(1)$. Without any loss of generality we shall assume $h(\lambda) = b\lambda$ with $b > 0$.

Remarks.

(1) The assumption $P_{0,\eta}(R^*) = \alpha$ for any q implies that the distribution of U under H_0 is independent of the choice of q. The test with critical region R^* satisfying the assumption is called null robust.

(2) The assumption $P_{\lambda,\eta}(R^*) = \alpha + h(\lambda) + r(\lambda, \eta)$ for any q implies that the distribution of U under H_1 is independent of the choice of q as $\lambda \to 0$. The test with critical region R^* satisfying the assumption is called locally non-null as $\lambda \to 0$.

Let ξ_0, ξ_λ be the probability measures on the sets $\{\delta = 0\}$, $\{\delta = \lambda\}$, respectively, satisfying

$$\frac{\int f(z : \lambda, \eta)\xi_\lambda(d\eta)}{\int f(z : 0, \eta)\xi_0(d\eta)} = 1 + h(\lambda)[g(\lambda) + r(\lambda)u] + B(z, \lambda) \tag{6.9}$$

for any q in (6.2) where $0 < c_1 < r(\lambda) < c_2 < \infty$ for λ sufficiently small, $g(\lambda) = o(1)$ and $B(z, \lambda) = o(\lambda)$ uniformly in z.
Note. If the set $\delta = 0$ is a single point, ξ_0 assigns measure 1 to that point. In this case we obtain

$$\frac{\int f(z : \lambda, \eta)\xi_\lambda(d\eta)}{\int f(z : 0, \eta)\xi_0(d\eta)} = \int \frac{f(z : \lambda, \eta)}{f(z : o, \eta)}\xi_\lambda(d\eta)\,. \tag{6.10}$$

Since $\bar{G}$ acts transitively on the range space $\mathcal{Y}$ of Ψ, the right-hand side of (6.10) is independent of q.

Theorem 6.2.1. *If R^* satisfies* (6.8) *and for sufficiently small λ, there exist ξ_λ and ξ_0 satisfying* (6.10), *then R^* is locally minimax for testing H_0 : $\delta = 0$ against the alternative $H_1 : \delta = \lambda$ for any q in* (6.2) *as $\lambda \to 0$, i.e.*

$$\lim_{\lambda\to 0} \frac{\inf_\eta P_{\lambda,\eta}(R^*) - \alpha}{\sup_{\phi_\lambda \in Q_\alpha} \inf_\eta P_{\lambda,\eta}\{\phi_\lambda \ rejects\ H_0\} - \alpha} = 1 \tag{6.11}$$

for any q in (6.2), *where Q_α is the class of test ϕ_λ of level α.*

Proof. Let

$$\tau_\lambda = (2 + h(\lambda)[g(\lambda) + C_\alpha r(\lambda)])^{-1}.$$

A Bayes critical region for (0, 1) losses with respect to the *a priori* $\xi^*_\lambda = \tau_\lambda\xi_\lambda + (1 - \tau_\lambda)\xi_0$ is given by

$$\begin{aligned} B_\lambda(z) &= \left\{ z : \int \frac{f(z : \lambda, \eta)}{f(z : 0, \eta}\xi_\lambda(d\eta) \ge \frac{1 - \tau_\lambda}{\tau_\lambda} \right\} \\ &= \left\{ z : u(z) + \frac{B(z, \lambda)}{r(\lambda)h(\lambda)} \ge C_\alpha \right\}. \end{aligned} \tag{6.12}$$

By (6.10) $B_\lambda(z)$ holds for any q in (6.2). Write $B_\lambda(z) = B_\lambda$, $V_\lambda = R^* - B_\lambda$ and $W_\lambda = B_\lambda - R^*$. Since $\sup_z |B_\lambda(z)/h(\lambda)| = o(1)$ and the distribution of U is continuous, letting

$$P^*_{0,\lambda}(A) = \int P_{o,\eta}(A)\xi_0(d\eta)$$
$$P^*_{1,\lambda}(A) = \int P_{\lambda,\eta}(A)\xi_\lambda(d\eta)$$

we get

$$P^*_{0,\lambda}(W_\lambda + V_\lambda) = o(1).$$

Since, with $A = V_\lambda$ or W_λ,

$$P^*_{1,\lambda}(A) = P^*_{0,\lambda}(A) \quad (1 + o(h(\lambda))$$

writing

$$r^*_\lambda(A) = (1-\tau_\lambda)P^*_{0,\lambda}(A) + \tau_\lambda(1 - P^*_{1,\lambda}(A))$$

we get the integrated Bayes risk with respect to ξ^*_λ as

$$\begin{aligned} r^*_\lambda(B_\lambda) &= (1-\tau_\lambda)P^*_{0,\lambda}(B_\lambda) + \tau_\lambda(1 - P^*_{1,\lambda}(B_\lambda)), \\ r_\lambda(R^*) &= (1-\tau_\lambda)P^*_{0,\lambda}(R) + \tau_\lambda(1 - P^*_{1,\lambda}(R^*)) . \end{aligned}$$

Hence for any q in (6.2)

$$\begin{aligned} r^*_\lambda(B_\lambda) &= r^*_\lambda(R^*) + (1-\tau_\lambda)(P^*_{0,\lambda}(W_\lambda) - P^*_{0,\lambda}(V_\lambda)) \\ &\quad + \tau_\lambda(P^*_{1,\lambda}(V_\lambda) - P^*_{1,\lambda}(W_\lambda)) \\ &= r^*_\lambda(R^*) + (1-2\tau_\lambda)(P^*_{0,\lambda}(W_\lambda) - P^*_{0,\lambda}(V_\lambda)) \\ &\quad + P^*_{0,\lambda}(V_\lambda + W_\lambda)o(h(\lambda)) \\ &= r^*_\lambda(R^*) + o(h(\lambda)). \end{aligned} \tag{6.13}$$

If (6.11) is false for all q in (6.2), then by (6.8) we can find a family of tests $\{\phi_\lambda\}$ of level α such that ϕ_λ has power function $\alpha + r(\lambda, \eta)$ on the set $\{\delta = \lambda\}$ for all q in (6.2) satisfying

$$\lim_{\lambda \to 0} \sup_{\eta} (\inf[r(\lambda,\eta) - h(\lambda)]/h(\lambda) > 0.$$

Hence, the integrated Bayes risk r'_λ of ϕ_λ with respect to ξ^*_λ, then satisfies

$$\lim_{\lambda \to 0} \sup (r^*_\lambda(R^*) - r'_\lambda)/h(\lambda) > 0$$

for all q in (6.2) contradicting (6.13) □

6.3. Examples

Examples 6.3.1. Let $X = (X_{ij}) = (X_1', \ldots, X_n')$, $X_i' = (X_{i1}', \ldots, X_{ip})$, $i = 1, \ldots, n$ be a $n \times p$ $(n \geq p)$ matrix with probability density function

$$f_X(x) = |\Sigma|^{-n/2} q\Big(\sum_{i=1}^{n} (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \Big) \tag{6.14}$$

where q is from $[0, \infty)$ to $[0, \infty)$, $\mu = (\mu_1, \ldots, \mu_p)' \in E^p$ and Σ is a $p \times p$ positive definite matrix. We shall assume that q is thrice contunuously differentiable. Write for any $b = (b_1, \ldots, b_p)' = (b_{(1)}', b_{(2)}')'$, with $b_{(1)} = (b_1, \ldots, b_{p_1})', b_{(2)} = (b_{p_1+1}, \ldots, b_p)', b_{[i]} = (b_1, \ldots, b_i)'$ and for any $p \times p$ matrix

$$A = (a_{ij}) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{ij} are $p_i \times p_j$ submatrices of A with $p_1 + p_2 = p$. We shall denote by

$$A_{[i]} = \begin{pmatrix} a_{11}, \ldots, a_{1i} \\ \cdot, \ldots, \cdot \\ \cdot, \ldots, \cdot \\ a_{i1}, \ldots, a_{ii} \end{pmatrix} .$$

We are interested here in the following three problems.

Problem 1. To test $H_{10} : \mu = 0$ against $H_{11} : \mu \neq 0$ when Σ is unknown.

Problem 2. To test $H_{20} : \mu_{(1)} = 0$ against $H_{21} : \mu_{(1)} \neq 0$ when $\mu_{(2)}$, Σ are unknown.

Problem 3. To test $H_{30} : \mu = 0$ against $H_{31} : \mu_{(1)} = 0$, $\mu_{(2)} \neq 0$ when Σ is unknown. The normal analogues of these problems have been considered in Chapter 4.

Problem 1.

Let $G_l(p)$ be the multiplicative group of $p \times p$ nonsingular matrices g. Problem 1 remains invariant under $G_l(p)$ transforming

$$(\bar{X}, S; \mu, \Sigma) \to (g\bar{X}, gSg'; g\mu, g\Sigma g')$$

where $\bar{X} = \frac{1}{n}\Sigma_{i=1}^{n} X_i$, $S = \Sigma_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})'$. A maximal invariant in the space of $(\bar{X}, S)$ is $T^2 = n\bar{X}' S^{-1} \bar{X}'$ or equivalently $R = n\bar{X}'(S + n\bar{X}\bar{X}')^{-1}\bar{X} =$

$T^2/(1+T^2)$. A corresponding maximal invariant in the parametric space under $G_l(p)$ is $\delta = n\mu'\Sigma^{-1}\mu$. If q is convex and nonincreasing, then the Hotelling T^2 test which rejects H_{10} whenever $R \geq C$ is uniformly most powerful invariant with respect to $G_l(p)$ for testing H_{10} against H_{11} (Kariya (1981)). As stated in Chapter 5, the group $G_l(p)$ does not satisfy the conditions of the Hunt–Stein theorm. However this theorem does apply for the subgroup $G_T(p)$ of $p \times p$ nonsingular lower triangular matrices with $p \geq 2$. From Example 5.1.1. the maximal invariant under $G_T(p)$ is $(R_1, \ldots, R_p)'$, whose distribution depends on $\Delta = (\delta_1, \ldots, \delta_p)'$. The ratio $\bar{R}$ of probability densities of $(R_1, \ldots, R_p)'$ under $H_{11} : \delta = \lambda$ and $H_{10} : \delta = 0$ is given by (with $g = (g_{ij}) \in G_T = G_T(p)$)

$$\bar{R} = \frac{\int_{G_T} p_1(gx) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg}{\int_{G_T} p_0(gx) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg} \tag{6.15}$$

where

$$\begin{aligned} dg &= \prod dg_{ij}, \\ p_1(x) &= q(\text{tr } x'x - 2n\bar{x}\mu + \lambda), \\ p_0(x) &= q(\text{tr } x'x). \end{aligned}$$

Since $x'x > 0$, there exists a $g \in G_T$ such that

$$gx'xg' = I, \;\; \sqrt{n} g\bar{x} = y = (\sqrt{R}_1, \ldots, \sqrt{R}_p)' .$$

Hence

$$\bar{R} = \frac{\int_{G_T} q(\text{tr}(gg' - 2\Delta' gy + \lambda)) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg}{\int_{G_T} q(\text{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg} . \tag{6.16}$$

Under the assumption that q is thrice differentiable, we expand q as

$$\begin{aligned} &q(\text{tr}(gg' - 2\Delta' gy + \lambda)) \\ &\quad = q(\text{tr}(gg')) + (-2\text{tr}(\Delta' gy) + \lambda) q^{(1)}(\text{tr } gg') \\ &\qquad + \frac{1}{2}(-2\text{tr}(\Delta' gy) + \lambda)^2 (\text{tr } (gg')) \\ &\qquad + \frac{1}{6}(-2\text{tr}(\Delta' gy) + \lambda)^3 q^3(z) \end{aligned} \tag{6.17}$$

where $z = \mathrm{tr}(gg') + (1-\alpha)(-2\mathrm{tr}(\Delta' gy) + \lambda)$, $0 \le \alpha \le 1$ and $q^{(i)}(x) = \frac{d^i q}{dx^i}$. Since the measures

$$q^{(k)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg, \quad k = 1, 2, 3$$

are invariant under the change of sign of g to $-g$, we conclude that

$$\int_{G_T} \mathrm{tr}(\Delta' gy) q^{(k)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg = 0,$$
$$\int_{G_T} g_{ij} g_{lm} q^{(k)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ij}^2)^{(n-i)/2} dg = 0 \tag{6.18}$$

for $i \neq l$, $j \neq m$. Now letting

$$D = \int_{G_T} q(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg,$$

from (6.15)–(6.18) we get

$$\begin{aligned} \bar{R} = 1 &+ \frac{\lambda}{D} \int_{G_T} q^{(1)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg \\ &+ \frac{1}{2D} \int_{G_T} (\mathrm{tr}(\Delta' gy))^2 q^{(2)}(\mathrm{tr}(gg')) \prod_{i=1}^{2} (g_{ii}^2)^{(n-i)/2} dg \\ &+ \frac{\lambda^2}{2D} \int_{G_T} q^{(2)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg \\ &+ \frac{1}{6D} \int_{G_T} (-2\mathrm{tr}(\Delta' gy) + \lambda)^3 q^{(3)}(z) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg. \end{aligned} \tag{6.19}$$

The first integral in (6.19) is a finite constant β_1. To evaluate the second integral in (6.19) we first note that

$$\mathrm{tr}(\Delta' gy) = \sum_{j=1}^{p} r_j^{1/2} \left[\sum_{i>j} \delta_i^{\frac{1}{2}} g_{ij} + \delta_j^{1/2} g_{jj} \right]. \tag{6.20}$$

From (6.17) and (6.19) the second integral in (6.18) reduces to

$$\int_{G_T} \left(\sum_j r_j \sum_{i>j} \delta_i g_{ij}^2 \right) q^{(2)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg. \tag{6.21}$$

To evaluate the above integral we need to evaluate the following two integrals.

$$\begin{aligned} I_1 &= \int_{G_T} g_{ii}^2 q^{(2)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg, \\ I_2 &= \int_{G_T} g_{ij}^2 q^{(2)}(\mathrm{tr}(gg')) \prod_{i=1}^{p} (g_{ii}^2)^{(n-i)/2} dg. \end{aligned} \tag{6.22}$$

Define

$$\begin{aligned} L &= \mathrm{tr}(gg') \\ e_i &= g_{ii}^2/L, \quad i = 1, \ldots, p; \\ e_{p+i} &= g_{i+1,i}^2/L, \quad i = 1, \ldots, p-1; \\ e_{p+p-1+i} &= g_{i+2,i}^2/L, \quad i = 2, \ldots, p-2; \\ &\vdots \\ e_{p(p+1)/2} &= g_{p1}^2/L \end{aligned} \tag{6.23}$$

and

$$\begin{aligned} K &= \int_{G_T} q^{(2)}(\mathrm{tr}(gg'))dg \\ N &= \int_{G_T} L^{1+\frac{1}{2}\Sigma_1^p(n-i)} q^{(2)}(L)dg. \end{aligned}$$

Since

$$K^{-1} q^{(2)}(\mathrm{tr}(gg'))$$

is a spherical density of $g_{ij}'s$, L and $e = (e_1, \ldots, e_{p(p+1)/2})'$ are independent and e has a Dirichlet distribution $D(1/2, \ldots, 1/2)$. From Kariya and Eaton (1977) the probability density function of e is given by

$$\begin{aligned} p(e) &= \Gamma\left(\frac{p(p+1)}{4}\right) / [\Gamma(1/2)]^{p(p+1)/2} \\ &\times \prod_{i=1}^{p(p+1)/2-1} (e_i)^{1/2-1} \left(1 - \sum_{1}^{p(p+1)/2-1} e_i\right)^{1/2-1}. \end{aligned} \tag{6.24}$$

Using (6.22) and (6.23) we get

$$I_1 = (n-i+1)\frac{NM}{2C}, \quad I_2 = \frac{NM}{2C} \tag{6.25}$$

where

$$M = E\left(\prod_{i=1}^{p}(e_i)^{(n-i)/2}\right),$$

$$C = p(p+1)/4 + 1/2\sum_{i=1}^{p}(n-i)\,. \tag{6.26}$$

Let $\eta = (\eta_1, \ldots, \eta_p)'$, $\eta_i = \delta_i/\lambda$. From (6.18) we get

$$\bar{R} = 1 + \lambda D\left(\beta_1 + \frac{MN}{2C}\left(\sum_{j=1}^{p} r_j\left(\sum_{i>j}\eta_i + (n-j+1)\eta_j\right)\right)\right) + B(y,\eta,\lambda) \tag{6.27}$$

where $B(y,\eta,\lambda) = o(\lambda)$ uniformly in $\mathcal{Y}$ and η. Thus, from (6.27), with

$$U(x) = \sum_{j=1}^{p} r_j = r\,,$$

the equation (6.9) is satisfied by letting ξ_0 give measure one to the single point $\eta = 0$ while ξ_λ gives measure one to the single point $\eta = \eta^* = (\eta_1^*, \ldots, \eta_p^*)$ whose jth corrdinate η_j^* satisfies

$$\eta_j^* = (n-j)^{-1}(n-j+1)^{-1}p^{-1}n(n-p),\;\; j = 1, \ldots, p \tag{6.28}$$

so that

$$\sum_{i>j}\eta_i^* + (n-j+1)n_j^* = n/p. \tag{6.29}$$

Since $G_l(p)$ and $G_T(p)$ satisfy the condition of Theorem 6.1.1., $\bar{R}$ does not depend on q. The first equation in (6.8) follows from the following lemma.

Lemma 6.3.1. *Let $Y = (Y_1, \ldots, Y_n)'$, $Y_i = (Y_{i1}, \ldots, Y_{ip})'$, $i = 1, \ldots, n$ be an $n \times p$ random matrix with spherically symmetric probability density function*

$$f_Y(y) = q(\mathrm{tr}\; yy') = q\left(\sum_{i=1}^{n}\mathrm{tr}\; y_i y_i'\right), \tag{6.30}$$

and let

$$Y = \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix}$$

where $Y_{(1)}, Y_{(2)}$ are of dimensions $k \times p$, $(n-k) \times p$ respectively and $n - k \geq p \geq k$. The distribution of $Y_{(1)}\left(Y'_{(2)}Y_{(2)}\right)^{-1}Y_{(1)}$ does not depend on q.

Proof. Since $n \geq p$, $Y'Y$ is positive definite with probability one. Hence there exists a $p \times p$ symmetric positive definite matrix A such that $Y'Y = AA$. Transform Y to U, given by $Y = UA$. As the Jacobian of the transformation $Y \to U$ is $|A|^n$,

$$f_{U,A}(u, a) = q(\text{tr } aa')|a|^n .$$

Thus U has the uniform distribution and U is independent of A. Partition U as

$$U = \begin{pmatrix} U_{(1)} \\ U_{(2)} \end{pmatrix}$$

where $U_{(1)}$ is $k \times p$ and $U_{(2)}$ is $(n-k) \times p$. From $Y = UA$ we get $Y_{(i)} = U_{(i)}A$, $i = 1, 2$. Thus

$$\begin{aligned} Y_{(1)}(Y'_{(2)}Y_{(2)})^{-1}Y'_{(1)} &= U_{(1)}A(A(U'_{(2)}U_{(2)})A)^{-1}AU'_{(1)} \\ &= U_{(1)}(U'_{(2)}U_{(2)})^{-1}U'_{(1)} . \end{aligned}$$

Since $Y_{(1)}(Y'_{(2)}Y_{(2)})^{-1}Y'_{(1)}$ is a function of U alone, it has a completely specified distribution. □

Corollary 6.3.1. If $k = 1$, $Y_{(1)}(Y'_{(2)}Y'_{(2)})^{-1}Y'_{(1)}$ is distributed as Hotelling's T^2 (central i.e. $\delta = 0$).

Proof. Since $Y_{(1)}(Y'_{(2)}Y_{(2)})^{-1}Y'_{(1)}$ has a completely specified distribution for all q in (6.30), we can, without any loss of generality, assume that $Y_1, \ldots, Y_n$ are independently distributed $N_p(0, I)$. This implies that $Y'_{(2)}Y_{(2)} = \Sigma_{i=2}^{n} Y_i'Y_i'$ has Wishart distribution $W_p(n-1, I)$ independently of Y_1. Hence $Y'_{(1)}(Y'_{(2)}Y_{(2)})^{-1}Y'_{(1)}$ has Hotelling's T^2 distribution. □

From this it follows that when $\delta = 0$, $T^2 = n\bar{Y}'S^{-1}\bar{Y}$, with $\bar{Y} = 1/n \sum_{i=1}^{n} Y_i$, $S = \sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ has Hotelling's T^2 distribution for all q in (6.30).

To establish the second equation in (6.8) we proceed as follows. Using (2.21) we get for any $g \in G_l(p)$, (which we will write as G_l)

$$\frac{f_{T^2}(t^2|\delta)}{f_{T^2}(t^2|0)} = \frac{\int_{G_l} q(\text{tr}(gg' - 2\alpha' gy + \delta))|gg'|^{(n-p)/2} dg}{\int_{G_L} q(\text{tr } gg')|gg'|^{(n-p)/2} dg} \tag{6.31}$$

where $y = (\sqrt{r}, 0, \ldots, 0)'$, $\alpha = (\sqrt{\delta}, 0, \ldots, 0)'$ are p-vectors and $r = t^2/(1+t^2)$. After a Taylor expansion of q, the numerator of (6.31) can be written as

$$\int_{G_l} q(\text{tr } gg')|gg'|^{\frac{1}{2}(n-p)} dg + \delta \int_{G_l} q^{(1)}(\text{tr } gg')|gg'|^{(n-p)/2} dg$$
$$+ 2\delta r \int_{G_l} q^{(2)}(\text{tr } gg') g_{11}^2 |gg'|^{(n-p)/2} dg + o(\delta)$$

where $g = (g_{ij})$. Using (6.18) we get the ratio in (6.13) as

$$1 + \delta(k + cr) + B(y, \eta, \delta) \tag{6.32}$$

where $B(y, \eta, \delta) = o(\delta)$ uniformly in y, η and $\eta = (\eta_1, \ldots, \eta_p)'$, with $\eta_i = \delta_i/\delta, i = 1, \ldots, p$ and k, c are positive constants. Hence for testing $H_{10} : \delta = 0$ against the alternative $H'_{11} : \delta = \lambda$ the Hotelling's T^2 test with critical region.

$$R^* = \{U(X) = R \geq c_\alpha\} \tag{6.33}$$

where c_α is a positive constant so that R^* has the size α, satisfies

$$P_{\lambda,\eta}(R^*) = \alpha + b\lambda + g(\lambda, \eta) \tag{6.34}$$

with $b > 0$ and $g(\lambda, \eta) = o(b\lambda)$. Thus we get the following theorem.

Theorem 6.3.1. *For testing $H_{10} : \mu = 0$ against the alternative $H'_{11} : \delta = \lambda$ specified > 0 Hotelling's T^2 test is locally minimax as $\lambda \to 0$ for the family of distribution* (6.14).

Remark 1. Since $G_l(p)$ satisfies the condition of Theorem 6.1.1., the ratio (6.31) is independent of q. From Corollary 6.3.1. and equations (6.31) and (6.34) it follows that Hotelling's T^2 test is locally best invariant for testing H_{10} against $H'_{11} : \delta = \lambda$ as $\lambda \to 0$ for the family of distributions (6.14). If q is a nonincreasing convex function then Hotelling's T^2 test is uniformly most powerful invariant.

Problem 2. Write $X = (X_{(1)}, X_{(2)})$ with $X_{(1)} : n \times p_1, X_{(2)} : n \times p_2$, and $p_1 + p_2 = p$.

Since the problem is invariant under translations of the last p_2 components of each $X_j, j = 1, \ldots, n$, this can be reduced to that of testing $H_{20} : \mu_{(1)} = 0$ against $H_{21} : \mu_{(1)} \neq 0$ in terms of $X_{(1)}$ only whose marginal probability density function is given by

$$f_{X_{(1)}}(x_{(1)}) = |\Sigma_{11}|^{-n/2} \tilde{q}(\text{tr } \Sigma_{11}^{-1}(x_{(1)} - e\mu'_{(1)})'(x_{(1)} - e\mu'_{(1)}))$$

where Σ_{11} is $p_1 \times p_1$ left-hand cornered submatrix of $\Sigma, e = (1, \ldots, 1)'$ n-vector and

$$\tilde{q}(\mathrm{tr}\, v) = \int_{R^{np_2}} q(\mathrm{tr}(v + ww'))dw\,,$$

and $\tilde{q}$ is a function on $[0, \infty)$ to $[0, \infty)$ satisfying (6.1a). Now, using the results of problem 1 with $p = p_1$, $R = \bar{R}_1 = \sum_{j=1}^{p_1} R_j$ we have the following theorem.

Theorm 6.3.2. *For testing $H_{20} : \mu_{(1)} = 0$ against $H'_{21} : \bar{\delta}_1 = n\mu'_{(1)} \cdot \Sigma_{11}^{-1}\mu_{(1)} = \lambda$ (specified) > 0 the test which rejects H_{20} for large values of*

$$\begin{aligned}\bar{R}_1 &= n\bar{X}'_{(1)}(S_{(11)} + n\bar{X}_{(1)}\bar{X}'_{(1)})^{-1}\bar{X}_{(1)} \\ &= n\bar{X}'_{(1)}S_{(11)}^{-1}\bar{X}_{(1)}/(1 + n\bar{X}'_{(1)}S_{(11)}^{-1}\bar{X}_{(1)})\end{aligned}$$

is locally minimax as $\lambda \to 0$ for the family of distributions (6.34).

Remark 2. The locally best invariant property and the uniformly most powerful invariant property under the assumption of nonincreasing convex property of $\tilde{q}$ follows from the results of problem 1.

Problem 3. The invariance property of this problem has been discussed in Chapter 4. The problem of testing H_{30} against H_{31} remains invariant under the multiplicative group of transformations G of $p \times p$ nonsingular matrices g,

$$g = \begin{pmatrix} g_{(11)} & 0 \\ g_{(21)} & g_{(22)} \end{pmatrix}$$

where $g_{(11)}$ is $p_1 \times p_1$, transforming $(\bar{X}, S) \to (g\bar{X}, gSg')$. A maximal invariant in the sample space under G is $(\bar{R}_1, \bar{R}_2)$ where

$$\bar{R}_1 = \sum_{i=1}^{p_1} R_i, \;\; \bar{R}_1 + \bar{R}_2 = R = \sum_{i=1}^{p} R_i\,.$$

A corresponding maximal invariant in the parametric space is $\bar{\delta}_1$, $\bar{\delta}_2$ where

$$\bar{\delta}_1 = \sum_{i=1}^{p_1} \delta_i = n\mu'_{(1)}\Sigma_{(11)}^{-1}\mu_{(1)},$$

$$\bar{\delta}_1 + \bar{\delta}_2 = \sum_{i=1}^{p} \delta_i = n\mu'\Sigma^{-1}\mu\,.$$

For invariant tests under G the problem is reduced to testing $H_{30} : \bar{\delta}_2 = 0$ against the alternatives $H_{31} : \bar{\delta}_2 > 0$, when it is given that $\bar{\delta}_1 = 0$, in terms of $\bar{R}_1$ and $\bar{R}_2$.

As stated in Chapter 5 the group G does not satisfy the conditions of the Hunt–Stein theorem. However this theorem does apply for the subgroup $G_T(p)$ of $p \times p$ nonsingular lower triangular matrices. From problem 1 the maximal invariant under $G_T(p)$ is $(R_1, \ldots, R_p)'$ whose distribution depends continuously on $\Delta = (\sqrt{\delta}_1, \ldots, \sqrt{\delta}_p)'$ with $\delta_i \geq 0$ for all i. Under $H_{30} : \delta_i = 0$, $i = 1, \ldots p$ and under $H_{31} : \delta_i = 0, i = 1, \ldots, p_1$. The ratio $\bar{R}$ of the probability densities of $(R_1, \ldots, R_p)'$ under $H'_{31} : \bar{\delta}_2 = \lambda > 0$ and under $H_{30} : \bar{\delta}_2 = 0$, when it is given that $\bar{\delta}_1 = 0$, is given by

$$\bar{R} = \frac{\displaystyle\int_{G_T} q(\mathrm{tr}(gg' - 2\Delta'_{(2)}(g_{(21)}y_{(1)} + g_{(22)}y_{(2)} + \lambda)) \prod_1^p (g_{ii}^2)^{(n-i)/2} dg}{\displaystyle\int_{G_T} q(\mathrm{tr}\; gg') \prod_1^p (g_{ii}^2)^{(n-i)/2} dg} \tag{6.35}$$

where $g \in G_T(p) = G_T, y = (\sqrt{R}_1, \ldots, \sqrt{R}_p)'$ and

$$g = (g_{ij}) = \begin{pmatrix} g_{(11)} & 0 \\ g_{(21)} & g_{(22)} \end{pmatrix}$$

with $g_{(11)}$ $g_{(22)}$ both lower triangular matrices.

As before we expand q in the numerator of (6.35) as

$$\begin{aligned} &q(\mathrm{tr}\; gg') + (-2\nu + \lambda)q^{(1)}(\mathrm{tr}\; gg') + (-2\nu + \lambda)^2 \\ &q^{(2)}(\mathrm{tr}\; gg') + (-2\nu + \lambda)^3 q^{(3)}(z) \end{aligned} \tag{6.36}$$

where

$$\begin{aligned} z &= \mathrm{tr}\; gg' + (1 - \alpha)(-2\nu + \lambda), \;\; 0 < \alpha < 1, \\ \nu &= \mathrm{tr}\; \Delta'_{(2)}(g_{(21)}y_{(1)} + g_{(22)}y_{(2)}) \,. \end{aligned}$$

Using (6.18) the integration of the second term in (6.36) with respect to the measure

$$\prod_1^p (g_{ii}^2)^{(n-i)/2} dg \tag{6.37}$$

gives $\lambda\alpha_1$ where

$$\alpha_1 = \int_{G_T} q^{(1)}(\mathrm{tr}\; gg') \prod_1^p (g_{ii}^2)^{(n-i)/2} dg \,.$$

To integrate the third term in (6.36) with respect to the measure given in (6.37) we first observe that (by (6.18))

$$\int_{G_T} \left(\operatorname{tr}(\Delta'_{(2)} g_{(21)} y_{(1)})\right)^2 q^{(2)}(\operatorname{tr} gg') \prod_1^p (g_{ii}^2)^{(n-i)/2} dg$$
$$= \int_{G_T} \left(\sum_{i=p_1+1}^{p} \delta_i \sum_{j=1}^{p_1} r_j g_{ij}^2 \right) q^{(2)}(\operatorname{tr} gg') \prod_1^p (g_{ii}^2)^{(n-i)/2} dg \tag{6.38}$$

and

$$\int_{G_T} \left(\operatorname{tr}(\Delta'_{(2)} g_{(22)} y_{(2)})\right)^2 q^{(2)}(\operatorname{tr} gg') \prod_1^p (g_{ii}^2)^{(n-i)/2} dg$$
$$= \int_{G_T} \left[\sum_{j=p_1+1}^{p} r_j \sum_{j \geq i} (\sigma_i g_{ij}^2 + \cdots + \sigma_j g_{jj}^2) \right]$$
$$\cdot\, q^{(2)}(\operatorname{tr} gg') \prod_1^p (g_{ii}^2)^{(n-i)/2} dg\,. \tag{6.39}$$

Denote

$$\begin{aligned}
K &= \int_{G_T} q^{(2)}(\operatorname{tr} gg') dg, \\
L &= \operatorname{tr} gg', \\
D &= \int_{G_T} q(\operatorname{tr} gg') \prod_1^p (g_{ii}^2)^{(n-i)/2} dg, \\
N &= \int_{G_T} L^{\Sigma_{i=1}^p (n-i)/2} q^{(2)}(L) dg, \\
M &= E\left(\prod_1^p (e_i)^{(n-i)/2} \right)
\end{aligned} \tag{6.40}$$

where $e_i' s$ are defined in (6.23). From (6.35)–(6.40) we get

$$\bar{R} = 1 + \frac{\lambda}{D} \left(\alpha_1 + \frac{2MN}{K} \left[\bar{r}_1 + \sum_{j=p_1+1}^{p} r_j \left(\sum_{i>j} \eta_i + (n-j+1)\eta_j \right) \right] \right)$$
$$+ B(y, \eta, \lambda) \tag{6.41}$$

where $B(y,\eta,\lambda) = o(\lambda)$ uniformly in y, η (independently of q by Theorem 6.1.1). The set $\{\lambda = 0\}$ is a simple point $\eta = 0$. So the ξ_0 in Theorem 6.2.1. assigns measure 1 to the single point $\eta = 0$. The set $\{\bar{\delta}_2 = \lambda\}$ is a convex $p_2 = (p - p_1)$ dimensional Euclidean set wherein each component $\eta_i = 0(h(\lambda))$.

Any probability measure ξ_λ can be replaced by degenerate measure ξ_λ^* which assigns measure 1 to the mean η_i^*, $i = p_1+1,\ldots,p$ of ξ_λ. Hence

$$\int \bar{R}\xi_\lambda^*(d\eta) = 1 + \frac{\lambda}{D}\left(\alpha_1 + \frac{2MN}{K}\left(\bar{r}_1 + \sum_{j=p_1+1}^{p} r_j\left(\sum_{i>j}\eta_i^* + (n-j+1)\right)\eta_j^*\right)\right) + B(y,\eta,\lambda) \tag{6.42}$$

where $B(y,\eta,\lambda) = o(h(\lambda))$ uniformly in y,η. Consider the rejection region

$$R_K = \{x : U(x) = \bar{r}_1 + K\bar{r}_2 \geq c_\alpha\} \tag{6.43}$$

where K is a constant such that (6.42) is reduced to yield (6.9) and C_α depends on the level of significance α of the test for the chosen K, independently of q (by Lemma 6.3.1.) Now choose

$$\begin{aligned}\eta_p^* &= \frac{n-p_1}{(n-p+1)p_2}, \\ \eta_j^* &= \frac{(n-j-1)\cdots(n-p)}{(n-j+1)\cdots(n-p+2)}\left(\frac{n-p_1}{(n-p+1)p_2}\right), \quad j = p_1+1,\ldots,p\end{aligned} \tag{6.44}$$

so that

$$\sum_{j<i}\eta_i^* + (n-j+1)\eta_j^* = \frac{n-p_1}{p_2}, \quad j = p_1+1,\ldots,p.$$

Hence we can conclude that the test with rejection region

$$R' = \left\{x : U(x) = \bar{r}_1 + \frac{n-p_1}{p_2}\bar{r}_2 \geq C_\alpha\right\} \tag{6.45}$$

with $P_{0,\eta}(R') = \alpha$ satisfies (6.9) as $\lambda \to 0$. Furthermore any region R_K of the form (6.43) must have $K = \frac{n-p_1}{p_2}$ to satisfy (6.9) as $\lambda \to 0$ for some ξ_λ.

It can be shown that (Giri (1988))

$$\begin{aligned} & f_{\bar{R}_1\bar{R}_2}(\bar{r}_1,\bar{r}_2|H_{31})/f_{\bar{R}_1,\bar{R}_2}(\bar{r}_1,\bar{r}_2|H_{30}) \\ &\quad = 1 + \frac{\bar{\delta}_2}{D_1}\left(\alpha_1 + \frac{2\alpha_2}{p_2}\left(\bar{r}_1 + \left(\frac{n-p_1}{p_2}\right)\bar{r}_2\right)\right) + B(\bar{r}_1,\bar{r}_2,\bar{\delta}_2)\end{aligned} \tag{6.46}$$

where D_1, α_1, α_2 are positive constants and $B(\bar{r}_1,\bar{r}_2,\bar{\delta}_2) = o(\bar{\delta}_2)$ uniformly in $\bar{r}_1$, $\bar{r}_2$. From this and Theorm 6.1.1 it follows that for testing H_{30} against H_{31} the test with critical region R' is locally best invariant as $\bar{\delta}_2 \to 0$. Hotelling's test which rejects H_{30} whenever $\bar{r}_1 + \bar{r}_2 \geq$ constant does not coincide with the LBI test and hence it is locally worse. From (6.32) it follows that Hotelling's test whose power depends only on $\delta = \bar{\delta}_2$, has positive derivative at $\bar{\delta}_2 = 0$. Thus, for the critical region of Hotelling's test the value of $h(\lambda)$ with $\bar{\delta}_2 = \lambda$ in (6.8) is positive. The first equation in (6.8) follows from Lemma 6.3.1. Hence we get the following theorem.

Theorem 6.2.2. *For testing H_{30} against $H'_{31} : \bar{\delta}_{(2)} = \lambda$ the test, which rejects H_{30} whenever $\bar{r}_1 + \frac{n-p_1}{p_2}\bar{r}_2 \geq$ constant, is locally minimax as $\lambda \to 0$ for the family of distributions* (6.2).

Example 6.3.2. Let $X = (X_{ij}) = (X_1, \ldots, X_n)'$ $X'_i = (X_{i1}, \ldots, X_{ip})', i = 1, \ldots, n$ be an $n \times p$ random matrix with probability density function given in (6.2). Let $\bar{X} = \frac{1}{n}\sum_1^n X_i$, $S = \sum_1^n (X_i - \bar{X})(X_i - \bar{X})'$. For any $p \times p$ matrix A we shall write

$$A = \begin{pmatrix} A_{(11)} & A_{(12)} & A_{(13)} \\ A_{(21)} & A_{(22)} & A_{(23)} \\ A_{(31)} & A_{(32)} & A_{(33)} \end{pmatrix} \tag{6.47}$$

with $A_{(11)} : 1 \times 1$, $A_{(22)} : p_1 \times p_1$, $A_{(33)} : p_2 \times p_2$ so that $p_1 + p_2 = p - 1$. Define

$$\bar{\rho}_1^2 = \Sigma_{(12)}\Sigma_{(22)}^{-1}\Sigma_{(21)}/\Sigma_{(11)},$$

$$\bar{\rho}^2 = \bar{\rho}_1^2 + \bar{\rho}_1^2 = (\Sigma_{(12)}, \Sigma_{(13)})' \begin{pmatrix} \Sigma_{(22)} & \Sigma_{(23)} \\ \Sigma_{(32)} & \Sigma_{(33)} \end{pmatrix}^{-1} (\Sigma_{(12)}, \Sigma_{(13)})'/\Sigma_{(11)}.$$

We shall consider the following three testing problems: (see Sec. 4.4)

Problem 4. To test $H_0 : \rho^2 = 0$ against $H_1 : \rho^2 > 0$ when μ, Σ are unkown.

Problem 5. To test $H_0 : \rho^2 = 0$ against $H_2 : \rho_1^2 > 0$, $\rho_2^2 = 0$ when μ, Σ are unknown.

Problem 6. To test $H_0 : \rho^2 = 0$ against $H_3 : \rho_1^2 = 0$, $\rho_2^2 > 0$ when μ, Σ are unknown.

These problems remain invariant under the group of translations transforming

$$(\bar{X}, S; \mu, \Sigma) \to (\bar{X} + b, S; \mu + b, \Sigma), \quad b \in \mathbb{R}^p.$$

The effiect of these transformations is to reduce these problems to corresponding ones where $\mu = 0$ and $S = \sum_{i=1}^{n} X_i X_i'$. Thus we reduce n by 1. We treat now the latter formulation with $\mu = 0$ and assume that $n \geq p \geq 2$ to assert that S is positive definite with probability one. Let G be the full linear group of $p \times p$ nonsingular matrices g of the form

$$g = \begin{pmatrix} g_{(11)} & 0 & 0 \\ 0 & g_{(22)} & 0 \\ 0 & g_{(32)} & g_{(33)} \end{pmatrix}. \tag{6.48}$$

For the first problem $p_2 = 0$, $p_1 = p - 1$ and for the other two problems $p_1 > 0, p_2 > 0$ satisfying $p_1 + p_2 = p - 1$. The problems above remain invariant under G operating as

$$(S; \Sigma) \to (gSg', g\Sigma g')$$

with $g \in G$. For Problem 6 a maximal invariant in the space of S under G (with $p_2 = 0$) is

$$R^2 = \frac{S_{(12)} S_{(22)}^{-1} S_{(21)}}{S_{(11)}}$$

and a corresponding maximal invariant in the parametric space of Σ under the induced group (with $p_2 = 0$) is

$$\rho^2 = \frac{\Sigma_{(12)} \Sigma_{(22)}^{-1} \Sigma_{(21)}}{\Sigma_{(11)}}. \tag{6.49}$$

For Problems 5, 6 a maximal invariant under G is $(\bar{R}_1^2, \bar{R}_2^2)$, where

$$\bar{R}_1^2 = \frac{S_{(12)} S_{(22)}^{-1} S_{(21)}}{S_{(11)}}, \tag{6.50}$$

$$\bar{R}_1^2 + \bar{R}_2^2 = (S_{(12)}, S_{(13)}) \begin{pmatrix} S_{(22)} & S_{(23)} \\ S_{(32)} & S_{(33)} \end{pmatrix}^{-1} (S_{(12)}, S_{(13)})' / S_{(11)}). \tag{6.51}$$

A corresponding maximal invariant in the parametric space of Σ is (ρ_1^2, ρ_2^2), where

$$\rho_1^2 = \frac{\Sigma_{(12)} \Sigma_{(22)}^{-1} \Sigma_{(21)}}{\Sigma_{(11)}},$$

$$\rho_1^2 + \rho_2^2 = \frac{(\Sigma_{(12)}, \Sigma_{(13)}) \begin{pmatrix} \Sigma_{(22)} \Sigma_{(23)} \\ \Sigma_{(32)} \Sigma_{(33)} \end{pmatrix} (\Sigma_{(12)}, \Sigma_{(13)})'}{\Sigma_{(11)}}. \tag{6.52}$$

Problem 4. The invariance of this problem has been discussed in Sec. 4.4. The group G does not satisfy the conditions of the Hunt–Stein theorem. The subgroup G_T of G, given by,

$$G_T = \left\{ g = \begin{pmatrix} g_{(11)} & 0 & 0 \\ 0 & g_{(22)} & 0 \\ 0 & g_{(32)} & g_{(33)} \end{pmatrix} \right\}$$

where $g_{(22)}(p_1 \times p_1)$, $g_{(33)}(p_2 \times p_2)$ are nonsingular lower triangular matrices, satisfies the conditions of the theorm. It may be remarked that we are using the same notation g for lower triangular matrices as g in G. In the place of R^2, the maximal invariant in the sample space under G_T is a $(p-1)$-dimensional vector $R = (R_2, \ldots, R_p)'$ defined in Example 5.1.5. with

$$\bar{R}_1 = \sum_{j=2}^{p_1+1} R_j, \bar{R}_1 + \bar{R}_2 = R^2 = \sum_{j=2}^{p} R_j.$$

The corresponding maximal invariant in the parametric space under G_T is $\Delta = (\delta_2^{1/2}, \ldots, \delta_p^{1/2})'$ as defined in Example 5.1.5. with

$$\rho_1^2 = \sum_{i=2}^{p_1+1} \delta_i, \ \ \rho^2 = \rho_1^2 + \rho_2^2 = \sum_{i=2}^{\rho} \delta_i$$

For testing $H_0 : \rho^2 = 0$ against the specified alternative $H_{1\lambda} : \rho^2 = \lambda$ the ratio $\bar{R}$ of the probability density function of R under $H_{1\lambda}$ to that under H_0 (using Theorem 2.7.1.) is

$$\begin{aligned} \bar{R} &= f_R(r|H_{1\lambda})/f_R(r|H_0) \\ &\times \frac{(1-\lambda)^{-n/2}}{B} \int_{G_T} q[(1-\lambda)^{-1}\mathrm{tr}(g_{(11)}^2 - 2g_{(11)}\Delta y' g'_{(22)} + g'_{(22)} g_{(22)}]\nu(dg) \end{aligned} \tag{6.53}$$

where

$$\begin{aligned} \nu(dg) &= (g_{(11)}^2)^{(n-1)/2} \prod_{i=2}^{p} (g_{ii}^2)^{(N+1-i)/2} dg \\ B &= \int_{G_T} q(\mathrm{tr}\ gg')\nu(dg). \\ y &= (\sqrt{r}_2, \ldots, \sqrt{r}_p)' \end{aligned}$$

Since the $(p-1)\times(p-1)$ matrix

$$C = (1-\rho^2)^{-1}(I - \Delta\Delta')$$

is positive definite there exists a $(p-1)\times(p-1)$ lower triangular nonsingular matrix T such that $TCT' = I$. Define $\gamma_i, \alpha_i, 2 \le i \le p$ by

$$\gamma_i = 1 - \sum_{j=2}^{i} \delta_j, \quad \alpha_i = \delta_i \gamma_p / \gamma_i \gamma_{i-1}.$$

Obviously $\gamma_p = 1 - \rho^2$. Let $\alpha = (\alpha_2, \ldots, \alpha_p)'$. Then $\alpha = T\Delta$. Since $C\Delta = \Delta$ and $TCT' = I$ we get

$$\begin{aligned} \alpha &= T\Delta = TC\Delta = (T')^{-1}\Delta, \\ \alpha'\alpha &= \Delta'(T'T)^{-1}\Delta = \Delta C \Delta' \\ &= (1-\rho^2)^{-1}(\Delta'\Delta) - (\Delta'\Delta)(\Delta'\Delta) = \rho^2 . \end{aligned}$$

Furthermore with $\alpha_{[i]} = (\alpha_2, \ldots, \alpha_i)'$

$$\alpha'_{[i]}\alpha_{[i]} = \sum_{j=2}^{i} \delta_j\,, \quad 2 \le i \le p.$$

Since

$$|C| = (1-\rho^2)^{2-p}$$

we can write

$$\bar{R} = \frac{(1-\lambda)^{-(n/2)}}{B} \int_{G_T} q(\mathrm{tr}(g_{(11)}^2 - 2g_{(11)}\alpha y' g'_{(22)} + g_{(22)}g'_{(22)}))\nu(dg)\,. \quad (6.54)$$

Expanding the integrand in (6.45) as

$$\begin{aligned} & q(\mathrm{tr}\; gg') + q^{(1)}(\mathrm{tr}\; gg')(-2\mathrm{tr}(g_{(11)}\alpha y' g'_{(22)})) \\ & \quad + \frac{q^{(2)}(\mathrm{tr}\; gg')}{2}(-2\mathrm{tr}(g_{(11)}\alpha y' g'_{(22)}))^2 \\ & \quad + \frac{q^{(3)}(z)}{6}(-2\mathrm{tr}\;(g_{(11)}\alpha y' g_{(22)}))^3 \end{aligned}$$

where

$$z = \mathrm{tr}\; gg' + (1-\delta)(-2\mathrm{tr}\; g_{(11)}\alpha y' g'_{(22)})'$$

with $0 < \delta < 1$.

Since, with $g = (g_{ij})$,

$$\text{tr } g_{(11)}\alpha y' g'_{(22)} = g_{(11)} \sum_{j=2}^{p} (r_j)^{1/2} \Big(\sum_{i>j} \alpha_i g_{ij} + \alpha_j g_{ij} \Big)$$

and the measure $q^{(1)}$ (tr gg') $\nu(dg)$ is invariant under the change of sign of g to $-g$ we conclude as in (6.18) that

$$\begin{aligned} &\int_{G_T} [\text{tr } g_{(11)}\alpha y' g'_{(22)}] q^{(1)}(\text{tr } gg')\nu(dg) = 0 \\ &\int g_{ij} g_{lk} q^{(1)}(\text{tr } gg')\nu(dg) = 0, \;\; i \neq l, \;\; j \neq k\,. \end{aligned} \tag{6.55}$$

Let

$$\begin{aligned} L = \text{tr } gg', \;\; Le_1 &= g_{(11)}^2, \;\; Le_i = g_{ii}^2, \;\; i = 2, \ldots, p, \\ Le_{2p-1+i} &= G_{i+2,i}^2, \;\; i = 2, \ldots, p, \\ &\vdots \\ Le_{1+p(p-1)/2} &= g_{p_1}^2\,. \end{aligned}$$

Now proceeding as in Problem 3 we get

$$\begin{aligned} \bar{R} = 1 + \frac{n\lambda}{B} \Bigg[-1 + \frac{NM}{K} \sum_{j=2}^{p} r_j \Bigg(\sum_{i>j} \eta_i + (n-j+2)\eta_j \Bigg) \Bigg] \\ + B(y, \lambda, \eta) \end{aligned} \tag{6.56}$$

where

$$\begin{aligned} \eta_j &= \delta_j/\lambda, \;\; j = 2, \ldots, p, \\ M &= E\left[e_1^{(n-1)/2} \prod_{j=2}^{p} e_i^{(n-i+1)/2} \right], \\ N &= \int_{G_T} L^{(n-1)/2} + \sum_{i=2}^{p} \frac{(n-i+1)}{2} q^{(2)}(L) dg, \end{aligned}$$

$B(y_i, \lambda, \eta) = o(\lambda)$ uniformly in y and η. Furthermore from Theorem 6.1.1 $\bar{R}$ does not depend on q.

Since $\lambda = 0$ is a single point $\eta = 0$, ξ_0 assigns measure 1 to the point $\eta = 0$. The set $\rho^2 = \lambda$ is a convex $(p-1)$-dimensional Euclidean set wherein each component is $0(\lambda)$. Thus any probability measure ξ_λ can be replaced by the degenerate measure which assigns measure 1 to the mean $\eta_i^*, i = 2, \ldots, p$ of ξ_λ. Choosing

$$\eta_j^* = (n-j+1)^{-1}(n-j+2)^{-1}(p-1)^{-1}n(n-p+1), \\ j = 2, \ldots, p, \tag{6.57}$$

we see that (6.9) is satisfied with $U = \sum_{j=2}^p R_j = R^2$. From Giri (1988) or using Theorem 2.7.1. we get

$$\begin{aligned} & f_{R^2}(r^2|H_1)/f_{R^2}(r^2|H_0) \\ =& \frac{(1-\rho^2)^{-n/2}}{B} \int_G q[(1-\rho^2)^{-1}(\mathrm{tr}[g_{(11)}^2 - 2g_{(11)}\Delta\mu' g'_{(22)} \\ & + (1-\rho^2)(I - \Delta\Delta')g_{(22)}g'_{(22)}])\mu(dg) \end{aligned}$$

where $g \in G$, as defined in (6.48), with $p_2 = 0$,

$$\mu(dg) = (g_{(11)}^2)^{(n-1)/2} |g_{(22)}g'_{(22)}|^{n(p-1)/2} dg, \\ B = \int_G q(\mathrm{tr}\; gg')\mu(dg),$$

and $u = (u_1, \ldots, u_{p-1})'$ satisfying $u'u = r^2$. If ρ^2 is small the power of the test against the alternative H_1 which rejects H_0 whenever $R^2 \geq C_\alpha$ is given by $\alpha + h(\rho^2) + o(\rho^2)$, where $h(\rho^2) > 0$ for $\rho^2 > 0$ and $h(\rho^2) = o(1)$. The null robustness of the test follows from the fact that $T(X) = R^2$ satisfies $T(X) = T(Xh)$ where $h \in G_T$ (Kariya (1981b)). Hence we get the following theorem:

Theorem 6.2.3. *For testing H_0 against $H_1' : \rho^2 = \lambda$, the level-α test which rejects H_0 whenever $R^2 \geq C$ is locally minimax as $\lambda \to 0$.*

Problem 5. Here

$$\eta_j = 0, \;\; j = p_1 + 2, \ldots, p.$$

For testing H_0 against $H_2 : \rho_1^2 = \lambda$ we can write (using 6.55)

$$\bar{R} = 1 + \frac{n\lambda}{2}\left[-1 + \frac{MN}{K}\left(\sum_{j=2}^{p_1+1} r_j \left(\sum_{i>j} n_i + (n-j+2)\eta_j\right)\right)\right] \\ + B(r, \lambda, \eta). \tag{6.58}$$

Now letting ξ_o give measure one to $\eta = 0$ and ξ_λ give measure one to $\eta_2^*, \ldots, \eta_{p_1+1}^*$, the mean of ξ_λ where

$$\begin{aligned} \eta_j^* &= (n-p_1)n(n-j+1)^{-1}(n-j+2)^{-1}p_1^{-1} \\ j &= 1, \ldots, p_1 . \end{aligned} \tag{6.59}$$

we conclude that (6.9) is satisfied with $U = \bar{R}_1^2 = \sum_{j=2}^{p_1} R_j$ and (6.8) follows from Theorem 6.2.3.

Theorem 6.2.4. *For testing H_0 against $H_2' : \rho_1^2 = \lambda$, the level-$\alpha$ test which rejects H_0 whenever $\bar{R}_1^2 \geq C_\alpha$, where the constant C_α depends on level α of the test, is locally minimax as $\lambda \to 0$.*

Problem 6. Here

$$\eta_j = 0, \;\; j = 2, \ldots, p_1 + 1.$$

For testing H_0 against $H_3 : \rho_2^2 = \lambda$ we can write $\bar{R}$ as $\lambda \to 0$ as

$$\begin{aligned} \bar{R} = 1 + \frac{n\lambda}{2B}\left(\bar{r}_1^2 + \sum_{j=p_1+2}^{p} r_j\left(\sum_{i>j}\eta_i + (n-j+2)\eta_j\right)\right) \\ + B(r, \lambda, \eta) \end{aligned} \tag{6.60}$$

where $B(r, \lambda, \eta) = o(\lambda)$ uniformly in r and η. Let us now consider the rejection region of the form

$$R_K = \{y : U(y) = \bar{r}_1^2 + K\bar{r}_2^2 \geq C_\alpha\} \tag{6.61}$$

where K is chosen such that (6.61) is reduced to yield (6.9) and C_α depends on the level α of the test for the chosen K. Let ξ_0 give measure one to $\eta = 0$ and ξ_λ give measure one to a single point $(0, \ldots, 0, \eta_{p_1+2}^*, \ldots, \eta_p^*)$ where

$$\eta_j^* = (n-p+1)n(n-j+1)^{-1}(n-j+2)^{-1}, (p-1)^{-1}, \;\; j = p_1 + 2, \ldots, p$$

and let

$$\tilde{R} = \left\{y : U(y) = \bar{r}_1^2 + \frac{n-p_1}{p_2}\bar{r}_2^2 \geq C_\alpha\right\}$$

with $p_1 + p_2 = p - 1$. Observe that the rejection region $\tilde{R}$ with $P_{0,\eta}\tilde{R} = \alpha$ satisfies (6.9). Furthermore, for the invariant region $\tilde{R}$, $P_{\lambda,\eta}(\tilde{R})$ depends only on λ. Hence, from (6.8) $r(\lambda, r) = 0$. Since $\tilde{R}$ is locally best invariant as $\lambda \to$

0, the test, which rejects H_0 whenever $R^2 \geq C$, does not coincide with $\tilde{R}$ and hence is locally worse. From Problem 4 the power of the test (which rejects H_0 whenever $R^2 \geq C$) which depends only on λ, has positive derivative at $\lambda = 0$. Thus from (6.8) with $R^* = \tilde{R}$ we conclude that $h(\lambda) > 0$. Hence we have

Theorem 6.2.5. *For testing H_0 against $H_3' : \rho_2^2 = \lambda > 0$, the test with critical region $\tilde{R}$ is locally minimax as $\lambda \to 0$.*

Exercises

1. Let $X = (X_i, \ldots, X_p)'$ be a random vector with peobability density function (6.1). Show that $E(X) = \mu$, $\text{cov}(X) = \alpha_\Sigma$ with $\alpha = p^{-1}E(X-\mu)'\Sigma^{-1}(X-\mu))$.
2. Let X be a $n \times p$ random matrix with probability density function (6.2). Find the maximum likelihood estimator of μ and Σ.

References

1. N. Giri, *Locally minimax tests for multiple correlations*, Canad. J. Statist., pp. 53–60 (1979).
2. N. Giri, *Locally minimax tests in symmetrical distributions*, Ann. Inst. Stat. Math., pp. 381–394 (1988).
3. N. Giri, *Multivariate Statistical Analysis*, Marcel Dekker, N.Y. (1996).
4. T. Kariya and M. Eaton, *Robust test of spherical symmetry*, Ann. Statist., pp. 208–215 (1977).
5. T. Kariya, *A robustness property of Hotelling's T^2-problem*, Ann. Statist., pp. 206–215 (1981).
6. T. Kariya, *Robustness of multivariate tests*, Ann. Statist., pp. 1267–1275 (1981b).
7. T. Kariya and B. Sinha, *Non-null and optimality robustness of some tests*, Ann. Statist., pp. 1182–1197 (1985).
8. T. Kariya and B. Sinha, *Robustness of Statistical Tests*, Academic Press, N.Y. (1988).

Chapter 7

TYPE D AND E REGIONS

The notion of a type D or E region is due to Isaacson (1951). Kiefer (1958) showed that the usual F-test of the univarial general linear hypotheses possesses this property. Lehmann (1959) showed that, in finding type D region, invariance could be invoked in the manner of the Hunt-stein theorem; and this could also be done for type E regions (if they exist) provided that one works with a group which operates as the identity on the nuisance parameter set H of the testing problem.

Suppose, for a parameter set $\Omega = \{(\theta, \eta) : \theta \in \Theta, \eta \in H$ with associated distributions, with Θ a Euclidean set, that every test function ϕ has a power function $\beta_\phi(\theta, \eta)$ which, for each η, is twice continuously differentiable in the components of θ at $\theta = 0$, an interior point of Θ. Let Q_α be the class of locally strictly unbiased level α test of $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$. Our assumption on β_ϕ implies that all tests in Q_α are similar and that $\partial\beta_\phi / \partial\theta_i|_{\theta=0} = 0$ for all ϕ in Q_α. Let $\Delta_\phi(\eta)$ be the determinant of the matrix $B_\phi(\eta)$ of second derivalives of $\beta_\phi(\theta, \eta)$ with respect to the components of θ (the Gaussian curvature) at $\theta = 0$. Suppose the parametrization be such that $\Delta_{\phi'}(\eta) > 0$ for all η for at least one ϕ' in Q_α.

Definition 7.1. Type E test. A test ϕ^* is said to be of type E if $\phi^* \in Q_\alpha$ and $\Delta_{\phi^*}(\eta) = \max_{\phi \in Q_\alpha} \Delta_\phi(\eta)$ for all η.

Definition 7.2. Type D test. A type E test ϕ^* is said to be of type D if the nuisance parameter set H is a single point. In the problems treated earlier in Chapter 4 it seems doubtful that type E regions exist (in terms of

Lehmann's development, H is not left invariant by many transformations). We introduce here two possible optimality criteria in the same sprit as the type D and E criteria which will always be fulfilled by some test under minimum regularity assumptions. Let

$$\bar{\triangle}(\eta) = \max_{\phi \in Q_\alpha} \triangle_\phi(\eta) . \tag{7.1}$$

Definition 7.3. Type D_A test. A test ϕ^* is of type D_A if

$$\max_\eta \left[\bar{\triangle}(\eta) - \triangle_{\phi^*}(\eta)\right] = \min_{\phi \in Q_\alpha} \max_\eta \left[\bar{\triangle}(\eta) - \triangle_\phi(\eta)\right] . \tag{7.2}$$

Definition 7.4. Type D_M test. A test ϕ^* is of type D_M if

$$\max_\eta [\bar{\triangle}(\eta) / \triangle_{\phi^*}(\eta)] = \min_{\phi \in Q_\alpha} \max_\eta [\bar{\triangle}(\eta) / \triangle_\phi(\eta)] . \tag{7.3}$$

These two criteria resemble stringency and regret criteria employed elsewhere in statistics; the subscripts "A" and "M" stand for "additive" and "multiplicative" regret principles. The possession of these properties is invariant under the product of any transformations on Θ (acting trivially on H) of the same general type as those for which type D regions retain their property, and an arbitrary 1-1 transformation on H (acting trivially on Θ), but, of course, not under more general transformations on Ω. Obviously, a type E test automatically satisfies these weaker criteria.

Let us now assume that a testing problem is invariant under a group of transformations G for which the Hunt-Stein theorem holds and which acts trivially on Θ; that is

$$g(\theta, \eta) = (\theta, g\eta), \quad g \in G . \tag{7.4}$$

If ϕg is the test function, defined by $\phi g(x) = \phi(gx)$, then a trivial computation shows that

$$\triangle_{\phi g}(\eta) = \triangle_\phi(g\eta) . \tag{7.5}$$

It is easy to verify that $\triangle_{\phi g}(\eta) = \triangle_\phi(g\eta)$ implies $\bar{\triangle}(\eta) = \bar{\triangle}(g\eta)$. Furthermore, if ϕ is better than ϕ' in the sense of either of the above criteria, then ϕg is better than $\phi' g$. All of the requirements of Lehmann (1959) are easily seen to be satisfied, so that we can conclude that there is an almost invariant (hence, in our problems, invariant) test which is of type D_A or D_M. This differs from the way in which invariance is used in page 883 of Lehmann (1959), where it is used to reduce Θ rather than H here.

If the group G is transitive on H, then $\bar{\Delta}(\eta)$ is constant as is $\Delta_\phi(\eta)$ for an invariant ϕ, which we therefore write simply as Δ_ϕ. In this case we conclude that if ϕ^* is invariant and if ϕ^* of type D among invariant ϕ (i.e. if Δ_{ϕ^*} maximizes Δ_ϕ over all invariant ϕ), then ϕ^* is of type D_A or D_M among all ϕ.

To verify these optimality properties we need the following lemma.

Lemma 7.1. Let L be a class of non-negative definite symmetric matrices of order m and suppose J is a fixed nousingular member of L. If tr $J^{-1}B$ is maximized (over B in L) by $B = J$, then det B is also maximized by J. Conversely, if L is convex and J maximizes det B, the tr $J^{-1}B$ is maximized by $B = J$.

Proof. Write $J^{-1}B = A$. If $A = I$ maximizes tr A, we get $(\det A)^{1/m} \leq \operatorname{tr} \frac{A}{m} \leq 1 = \det I$. Conversely, if I maximizes $\det A$, it also maximizes tr A, since tr $B >$ tr I implies $\det(\alpha B + (1-\alpha)I) > 1$ for α small and positive. $\square$

Remarks. The usefulness of this lemma lies in the fact that the generalized Neyman-Pearson lemma allows us to maximize tr QB_ϕ, for fixed Q more easily than to maximize $\Delta_\phi = \det B_\phi$ among similar level α tests. We can find, for each Q, a, ϕ_Q which maximizes tr QB_ϕ; a ϕ^* which maximizes Δ_ϕ is then obtained by finding a ϕ_Q for which $B_{\phi_Q} = Q^{-1}$.

In problems of the type which we consider here the reduction by invariance under a group G which is transitive on H often results in a reduced problem wherein the maximal invariant is a vector $R = (R_1, \ldots, R_m)'$ whose distribution depends only on $\Delta = (\delta_1, \ldots, \delta_m)'$ where $\delta_i = \theta_i^2$, $\theta = (\theta_i, \ldots, \theta_m)'$ and such that the density f_Δ of R with respect to a σ-finite measure μ is of the form

$$f_\Delta(r) = f_0(r)\left\{1 + \sum_1^m \delta_i \left(h_i + \sum_j a_{ij}\ r_j\right)\right\} + Q(r, \Delta) \tag{7.6}$$

where the h_i and a_{ij} are constants and $Q(r, \Delta) = o(\sum \delta_i)$ as $\Delta \to 0$ and we can differentiate under the integral sign to obtain, for invariant test ϕ of level α,

$$\beta_\phi(\theta, \eta) = \alpha\left(1 + \sum_i^m \delta_i\ h_i\right) + \sum_i \delta_i \sum_j a_{ij} \int r_j\ \phi(r)\ f_0(r)\ \mu(dr) + o\Big(\sum \delta_i\Big) \tag{7.7}$$

as $\theta \to 0$. According to Giri and Kiefer (1969) this is called the symmetric reduced regular case (SRR).

Theorem 7.1. *In the SRR case, an invariant test ϕ^* of level α is of type D among invariant ϕ (and, hence, of type D_A and D_M among all ϕ) if and only if ϕ^* is of the form*

$$\phi^*(r) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \quad \text{if} \quad \sum_{ij} a_{ij} q_i r_j \begin{Bmatrix} > \\ < \end{Bmatrix} c \tag{7.8}$$

where c is a constant and $q_i^{-1} = \text{const}\left[h_i\alpha + E_0\left\{\sum_j a_{ij} R_j \phi^*(R)\right\}\right]$.

Proof. In the SRR case, every invariant ϕ has a diagonal B_ϕ whose ith diagonal entry, by (6.7), is $2[h_i\alpha + E_0\{\sum_j a_{ij} R_j \phi(R)\}]$. By Neyman–Pearson lemma tr QB_ϕ is maximized over such a ϕ by a ϕ^* of the form (6.7). Hence we get the theorem. □

Example 7.1. **(T^2-test)** Consider once again. Problem 1 of Chapter 4. Let $\theta = N^{1/2} F^{-1} \xi$ where F is the unique member of G_T with positive diagonal elements satisfying $F\ F' = \Sigma$. Then Θ is Euclidean p-space, G_T operates transitively on $H = \{\text{positive definite } \Sigma\}$ but trivially on Θ and we have the SRR case. We thus have (6.7) with $h_i = -\frac{1}{2}$,

$$a_{ij} = \begin{cases} 1, & \text{if } i > j\,, \\ 0, & \text{if } i < j\,, \\ N - j + 1, & \text{if } i = j\,. \end{cases} \tag{7.9}$$

Hotelling's T^2 test has a power function which depends only $\sum_i \delta_i$, so that, with the above parametrization for θ, we have B_{T^2} a multiple of identity. Hotelling's critical region is of the form $\sum r_i > c$. But, when all q_i are equal, the critical region corresponding to (6.8) is of the form $\sum_j (N+p-1-2j) r_j > c$, which is not Hotelling's region if $p > 1$. Hence we get the following theorem.

Theorem 7.2. *For $0 < \alpha < 1 < p < N$, Hotelling's T^2 test is not of type D among G_T-invariant tests, and hence is not of type D_A or D_M (nor of type E) among all tests.*

Remarks. The actual computation of a ϕ^* of type D appears difficult due to the fact that we need to evaluate an integral of the form

$$E_0(R_i^h\phi(R)) = \int_{\sum_j c_j r_j > c} \pi^{-\frac{1}{2}p}\Gamma(N/2)\, r_i^{h-\frac{1}{2}}(1-\sum_j r_j)^{\frac{1}{2}(N-p-2)} \prod dr_i \tag{7.10}$$

for $h = 0$ or 1 for various choices of the c_i's and c. When α is close to zero or 1, we can carry out approximate computations as follows. As $\alpha \to 1$ the complement $\tilde{R}$ of the critical region becomes a simplex with one corner at 0. When $p = 2$, if we write $\rho = 1-\alpha$ and consider critical regions of level α of the form $by_1 + y_2 > c$ where $0 < L^{-1} \le b \le L$, L being fixed but large (this keeps $\tilde{R}$ close to the origin), we get from (6.10) that $\rho = E_0(1-\phi(R)) = (N-2)c/2b^{1/2} + o(c)$ as $c \to 0$. Similarly, $E_0\{(1-\phi(R)R_i\} = (N-2)c^2b^{i-5/2}/8 + o(c) = \rho^2 b^{i-3/2}/2(N-2) + o(\rho)$ as $\rho \to 0$ while $E_0(R_i) = 1/N$. From (5.18) we obtain for the power near H_0

$$\alpha + \frac{\rho}{2}\left\{\delta_i\left[1 - \frac{N\rho + o(\rho)}{2(N-2)b^{\frac{1}{2}}}\right] + \delta_2\left[1 - \frac{\rho + o(\rho)}{2(N-2)b^{\frac{1}{2}}} - \frac{(N-1)\rho b^{\frac{1}{2}} + o(\rho)}{2(N-2)}\right]\right\}$$
$$+\, o\left(\sum \delta_i\right)$$

where $o(\rho)$ and $o(\sum\delta_i)$ terms are uniform in Δ and ρ, respectively. The product Δ_ϕ of the coefficients of δ_1 and δ_2 is maximized when $b = (N+1)/(N-1) + o(1)$, as $\rho \to 0$; with more care, one can obtain further terms in an expansion in ρ for the type D choice of b. The argument is completed by showing that $b < L^{-1}$ implies that $\tilde{R}$ lies in a strip so close to the r_1-axis as to make $E_0(1-\phi(R))R_1$ too large and $E_0\{(1-\phi(R))R_2\}$ too small to yield a ϕ as good as that with $b = (N+1)/(N-1)$, with a similar argument if $b > L$. When ρ is very close to 0, all choices of $b > 0$ give substantially the same power, $\alpha + (\delta_1+\delta_2)/2 + o(\rho^2)$, so that the relative departure from being type D, of Hotelling's T^2 test or any other critical region of the form $bR_1 + R_2 > c$, with b fixed and positive, approaches 0 as $\alpha \to 1$. However we do not know how great the departure of Δ_{T^2} from $\bar{\Delta}$ can be for arbitrary α. We can similarly treat the case $p > 2$ and also the case $\alpha \to 0$.

One can treat similarly other problems considered in Chapter 4.

References

1. N. Giri, and J. Kiefer, *Local and Asymptotic Minimax Properties of Multivariate Tests*, Ann. Math. Statist. **35**, 21–35 (1964).

2. S. L. Isaacson, *On the Theory of Unbiased Tests of Simple Statistical Hypothesis Specifying the Values of Two or more Parameters*, Ann. Math. Statist. **22**, 217–234 (1951).
3. J. Kiefer, *On the Nonrandomized Optimality and Randomized Nonoptimality of Symmetrical Designs*, Ann. Math. Statist. **29**, 675–699 (1958).
4. E. L. Lehmann, *Optimum Invariant Tests.*, Ann. Math. Statist. **30**, 881–884 (1959).