STOCHASTIC MONOTONICITY OF A DISTRIBUTION FAMILY ASSOCIATED WITH MATRIX PROJECTIONS AND ITS APPLICATIONS

A Dissertation by

Yufei WANG

Master of Science, Wichita State University, 2013

Bachelor of Science, Tianjin Normal University, 2007

Submitted to the Department of Mathematics, Statistics, and Physics
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

December 2019

STOCHASTIC MONOTONICITY OF A DISTRIBUTION FAMILY ASSOCIATED WITH MATRIX PROJECTIONS AND ITS APPLICATIONS

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy with a major in Applied Mathematics.

Xiaomi Hu, Committee Chair

Chunsheng Ma, Committee Member

Adam Jarger, Committee Member

Lop-Hing Ho, Committee Member

Hyuck M. Kwon, Committee Member

Accepted for the College of Liberal Arts and Sciences

Andrew Hippisley, Dean

Accepted for the Graduate School

Coleen Pugh, Dean

DEDICATION

To my parents, my teachers, my friends

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study. My first debt of gratitude must go to my advisor, Professor Xiaomi Hu, who was truly an inspiration. Without his guidance, this dissertation would not be possible to be completed. I would like to thank all of my professors for years of instruction and guidance. I would also like to thank my parents, my cousin Na and other relatives for their love and continuous support.

ABSTRACT

This dissertation defines a distribution family based on the distribution of a random variable associated with the projections of a normally distributed random matrix with unknown mean onto a linear space and onto a closed convex cone. The stachastic monotonicity property of the distributions in the family is established. This property finds the application in the proving the unbiasedness of a likelihood ratio test in a multivariate order restricted model.

TABLE OF CONTENTS

Chapter Page

TABLE OF CONTENTS (continued)

CHAPTER 1

Introduction

Random matrix $Z \sim N_{p\times q}(M, \Sigma, \Psi)$, see Arnold (1981), has whole space $R^{p\times q}$ as its support so that $E(Z) = M$ could be anywhere in $R^{p\times q}$. When $E(Z) = A\Theta$ with unknown parameter matrix Θ, $E(Z)$ is confined in a subspace of $R^{p\times q}$ spanned by the columns of A. When $E(Z) = \Theta B$, $E(Z)$ is restricted in a subspace of $R^{p\times q}$ spanned by the rows of B. Those are two types of commonly encountered linear models. The univariate regression models, the ANOVA models, the multivariate regression models, and the MANOVA models are all within this framework. Model $E(Z) = A\Theta B$, called a bilinear model, is also a linear model since $E(Z)$ again is a linear function of Θ, and $E(Z)$ is confined in a subspace of $R^{p\times q}$ spanned by the products of the columns of A and the rows of B. Growth curve model is an example, see Pan and Fang (2002). Hu and Banerjee (2012) introduced multivariate order restriction on M. This restriction confines M in a closed convex cone in $R^{p\times q}$. Generally, based on the true situations one may consider the restriction that M is in a closed convex subset of $R^{p\times q}$.

With respect to the two matrices Σ and Ψ in the distribution of Z, an inner product system can be established in $R^{p\times q}$. With the norm induced from the inner product, the projection of a matrix onto a closed convex set in $R^{p\times q}$ is explored since the restricted MLE (RMLE) of M constrained to a closed convex set is shown to be a projection onto the restriction set. Therefore for linear models the RMLEs are the matrix projections onto linear spaces. For order restrictions the RMLEs are the projections onto closed convex cones. While the projections onto linear spaces do have closed forms, the projections onto general closed convex sets including closed convex cones do not have closed forms. Many efforts have been devoted to developing algorithms for computing such projections in particular cases, see Sasabuchi, Inutsuka, and Kulatunga (1992), Hansohm and Hu (2012), Hu et al. (2012).

For computation simplicity Hu (2018) proposed a pseudo RMLE that is the projection onto a subset of the restriction set.

A projection of Z is still a random matrix with distributions depending on the unknown parameter matrix M. In hypotheses testing problems one has to deal with the projections onto restriction sets defined by the null hypotheses and the projections onto the sets imposed by the model specifications. The likelihood ratio test statistics are most likely involved with the functions of the norms of such projections. These functions are random variables with distributions again depending on the parameter matrix M.

In this dissertation, a particular function of the projections of Z onto a linear space and a closed convex cone, and a positive random variable independent to the projections is examined. The distributions of this function form a distribution family with members indexed by M. A study is conducted to establish a stochastic monotonicity on some matrix rays for the distribution family. This result, in specific cases, verifies the well-known facts that the non-central χ^2- distributions and non-central F-distributions are stochastically increasing with their non-centrality parameters.

The study in this dissertation is motivated by a problem of unbiasedness of a LRT in a multivariate order restricted MANOVA where the common variance-covariance matrix of response vectors is unknown, but is proportional to a known matrix with unknown proportion factor. The null hypothesis of the homogeneity of response vectors is $H_0 : M = \theta 1_q'$ which specifies a linear space for M while the model order restriction imposes a closed convex cone for M. The study derives a likelihood ratio test statistic through a function of the projections onto the linear space and the cone, and a positive random variable independent to the projections. The stochastic monotonicity for the family established earlier is successfully applied to the distribution of the test statistic to show that the test is unbiased. This work is a compliment to that of Mustafa Hamdan (2017). In his Ph.D dissertation a similar test with variance-covariance matrix known was proved to be unbiased.

In Chapter 2 random matrix $Z \sim N_{p\times q}(M, \Sigma, \Psi)$ is defined. The convenient properties, especially the ones associated with our function of the projections, are presented. In Chapter 3 the inner product system is established for $R^{p\times q}$, and the preliminaries on the matrix projections in $R^{p\times q}$ with respect to the norm induced from the inner product are discussed. In Chapter 4 with a particular function of the projections and a positive random variable, a distribution family is introduced. Chapter 5 is devoted to the proof of the stochastic monotonicity for the distribution family. In Chapter 6 the order restricted MANOVA model is introduced. The likelihood ratio test procedure is developed for the test on the homogeneity of response means in Chapter 7. By applying the result established in previous chapters the test is shown to be unbiased.

CHAPTER 2

A random matrix

2.1 Definition

Let Z be a $p \times q$ random matrix with a normal distribution denoted by

$$Z \sim N_{p\times q}(M, \Sigma, \Psi), \text{ i.e., } \mathrm{vec}(Z) \sim N_{pq}(\mathrm{vec}(M), \Psi \otimes \Sigma) \tag{2.1.1}$$

where $E(Z) = M \in R^{p\times q}; \Sigma \in R^{p\times p}$ and $\Psi \in R^{q\times q}$ are two positive definite matrices; and the operators $\mathrm{vec}(\cdot)$ and $\otimes$ are vectorization and Kronecker product respectively.

Specifying the structure of $E(Z) = M$ leads to many different models. For example the linear model with $M = A\Theta$ or $M = \Theta B$ where Θ is an unknown parameter matrix; bilinear model with $M = A\Theta B$; and multivariate order restricted model with $M \in \mathcal{C}$ where $\mathcal{C}$ is an order restricted cone in $R^{p\times q}$. The last model will be discussed in detail in Chapter 6.

2.2 Properties

Random matrix Z in (2.1.1) possesses many useful and convenient properties. We list some below.

Lemma 2.2.1. *Suppose $Z \sim N_{p\times q}(M, \Sigma, \Psi)$. Then $Z' \sim N_{q\times p}(M', \Psi, \Sigma)$.*

Proof: Note that $\mathrm{vec}(Z') = K_{pq}\mathrm{vec}(Z), \mathrm{vec}(M') = K_{pq}\mathrm{vec}(M)$ and $K_{pq}(\Psi\otimes\Sigma)K'_{pq} = \Sigma\otimes\Psi$ where K_{pq} is the $pq \times pq$ commutation matrix, see Magnus and Neudecker (1988). Thus

$$\mathrm{vec}(Z') = K_{pq}\mathrm{vec}(Z) \sim N_{qp}(K_{pq}\mathrm{vec}(M), K_{pq}(\Psi \otimes \Sigma)K'_{pq}) = N_{qp}(\mathrm{vec}(M'), \Sigma \otimes \Psi).$$

Hence $Z' \sim N_{q\times p}(M', \Psi, \Sigma)$. $\square$

Lemma 2.2.2. *Suppose* $Z \sim N_{p\times q}(M, \Sigma, \Psi)$. *For* $A \in R^{m\times p}$ *with* $rank(A) = m$, $B \in R^{q\times n}$ *with* $rank(B) = n$, *and* $C \in R^{m\times n}$, $AZB + C \sim N_{m\times n}(AMB + C, A\Sigma A', B'\Psi B)$.

Proof: Since $Z \sim N_{p\times q}(M, \Sigma, \Psi)$, i.e., $\text{vec}(Z) \sim N_{pq}(\text{vec}(M), \Psi \otimes \Sigma)$, we only need to check the distribution of $\text{vec}(AZB + C)$. Then with $E[\text{vec}(AZB + C)] = \text{vec}(AMB + C)$ and $cov[\text{vec}(AZB + C)] = (B' \otimes A)(\Psi \otimes \Sigma)(B' \otimes A)' = (B'\Psi B) \otimes (A\Sigma A')$, $\text{vec}(AZB + C) \sim N_{mn}(\text{vec}(AZB + C), (B'\Psi B) \otimes (A\Sigma A'))$. By (2.1.1), $AZB + C \sim N_{m\times n}(AMB + C, A\Sigma A', B'\Psi B)$. □

Lemma 2.2.3. *Suppose* $Z \sim N_{p\times q}(M, \Sigma, \Psi)$. *If* $A \in R^{q\times m}, B' = B \in R^{q\times q}$ *and* $A'\Psi B = 0$, *then* ZA *and* ZBZ' *are independent.*

Proof: Suppose $A \in R^{q\times m}$ and $B' = B \in R^{q\times q}$. By the compact form of eigenvalue decomposition of $B, B = P\Lambda P'$ where $P \in R^{q\times r}$ has orthnormal columns; $\Lambda \in R^{r\times r}$ is a non-singular diagonal matrix; and $r = rank(B)$. Therefore $0 = A'\Psi B = A'\Psi P\Lambda P'$ is equivalent to $A'\Psi P = 0$. By (2.1.1), $\text{vec}(Z) \sim N_{pq}(\text{vec}(M), \Psi \otimes \Sigma)$. But $\text{vec}(ZA) = (A' \otimes I_p)\text{vec}(Z)$ and $\text{vec}(ZP) = (P' \otimes I_p)vec(Z)$. Thus $A'\Psi P = 0$, which is equivalent to $(A' \otimes I_p)(\Psi \otimes \Sigma)(P' \otimes I_p)' = 0$ since $(A' \otimes I_p)(\Psi \otimes \Sigma)(P' \otimes I_p)'$ and $(A'\Psi P) \otimes \Sigma$ are equal, implies the independence of ZA and ZP. But $ZBZ' = (ZP)\Lambda(ZP)'$ is a function of ZP. Therefore, the independence of ZA and ZBZ' is proved. □

2.3 Wishart distribution

Z in (2.1.1) can also be used to define other distributions. Wishart distribution for random matrices is an example. With $\Psi = I_q$, suppose $Z \sim N_{p\times q}(M, \Sigma, I_q)$. Then the distribution of $ZZ' \in R^{p\times p}$ is called a Wishart distribution with degrees of freedom q, non-centrality parameter MM', and a parameter matrix Σ, which is denoted as

$$Z \sim N_{p\times q}(M, \Sigma, I_q) \Rightarrow ZZ' \sim W_p(q, MM', \Sigma).$$

When the non-centrality parameter is zero, $M = 0$, the Wishart distribution $W_p(q, 0, \Sigma)$ simplified as $W_p(q, \Sigma)$ is called a central Wishart distribution. If $\Sigma = I_p, W_p(q, I_p)$ is called a standardized Wishart distribution since it is generated from the standardized Z, $\Sigma^{-\frac{1}{2}}(Z - M) \sim N_{p\times q}(0, I_p, I_q)$.

The following property will be used for later proofs.

Lemma 2.3.1. *Suppose* $W \in R^{p\times p} \sim W_p(q, MM', \Sigma)$. *If* $A \in R^{m\times p}$ *and* $rank(A) = m$, *then* $AWA' \sim W_m(q, AMM'A', A\Sigma A')$.

Proof: By the definition of Wishart distribution, $W = ZZ'$ where $Z \sim N_{p\times q}(M, \Sigma, I_q)$. But by Lemma 2.2.2, $AZ \sim N_{m\times q}(AM, A\Sigma A', I_q)$. Hence

$$(AZ)(AZ)' = AWA' \sim W_m(q, AM(AM)', A\Sigma A'). \ \square$$

Lemma 2.3.2. *Suppose* $Z \sim N_{p\times q}(M, \Sigma, I_q), A \in R^{q\times q}, A' = A, A^2 = A$ *and* $rank(A) = r$, *then* $ZAZ' \sim W_p(r, MAM', \Sigma)$.

Proof: $A' = A, A^2 = A$ and $rank(A) = r$ imply that the compact form of the eigenvalue decomposition of A is $A = PP'$ where $P \in R^{q\times r}$ and $P'P = I_r$. By Lemma 2.2.2, $ZP \sim N_{p\times r}(MP, \Sigma, I_r)$. By the definition of Wishart distribution,

$$(ZP)(ZP)' = ZAZ' \sim W_p(r, MP(MP)', \Sigma) = W_p(r, MAM', \Sigma). \ \square$$

CHAPTER 3

Inner product and projections

3.1 The inner product

By (2.1.1), the probability density function of Z is

$$\phi(Z) = \frac{1}{(2\pi)^{\frac{pq}{2}}|\Psi \otimes \Sigma|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}[\text{vec}(Z) - \text{vec}(M)]'(\Psi \otimes \Sigma)^{-1}[\text{vec}(Z) - \text{vec}(M)] \right\}.$$

With Σ and Ψ in the distribution of Z, for A and B in $R^{p\times q}$ define

$$\langle A, B\rangle = [\text{vec}(B)]'(\Psi \otimes \Sigma)^{-1}[\text{vec}(A)]. \tag{3.1.1}$$

Now we have the following lemma.

Lemma 3.1.1. *$\langle A, B\rangle$ in (3.1.1) satisfies the axioms for an inner product in $R^{p\times q}$ and $\langle A, B\rangle = tr[(\Sigma^{-\frac{1}{2}}B\Psi^{-\frac{1}{2}})'(\Sigma^{-\frac{1}{2}}A\Psi^{-\frac{1}{2}})]$.*

Proof: Note that both Σ and Ψ are positive definite matrices. So are $\Psi\otimes\Sigma$ and $(\Psi\otimes\Sigma)^{-1}$. Therefore,

(i) $\langle A, A\rangle = [\text{vec}(A)]'(\Psi \otimes \Sigma)^{-1}[\text{vec}(A)] \geq 0$ and $\langle A, A\rangle = 0$ if and only if $\text{vec}(A) = 0$, i.e. $A = 0$.

(ii) $\langle A, B\rangle = \langle B, A\rangle$ since

$$\begin{aligned} \langle A, B\rangle &= [\text{vec}(B)]'(\Psi \otimes \Sigma)^{-1}[\text{vec}(A)] \\ &= \{[\text{vec}(B)]'(\Psi \otimes \Sigma)^{-1}[\text{vec}(A)]\}' \\ &= [\text{vec}(A)]'(\Psi \otimes \Sigma)^{-1}[\text{vec}(B)] = \langle B, A\rangle. \end{aligned}$$

(iii) For C in $R^{p\times q}$ and all scalars α and β,

$$\begin{aligned}\langle \alpha A+\beta C, B\rangle &= [\mathrm{vec}(B)]'(\Psi\otimes\Sigma)^{-1}[\mathrm{vec}(\alpha A+\beta C)]\\ &= \alpha[\mathrm{vec}(B)]'(\Psi\otimes\Sigma)^{-1}[\mathrm{vec}(A)]+\beta[\mathrm{vec}(B)]'(\Psi\otimes\Sigma)^{-1}[\mathrm{vec}(C)]\\ &= \alpha\langle A,B\rangle+\beta\langle C,B\rangle.\end{aligned}$$

Therefore, $\langle A,B\rangle$ in (3.1.1) is an inner product. Note that

$$\begin{aligned}[\mathrm{vec}(B)]'(\Psi\otimes\Sigma)^{-1}[\mathrm{vec}(A)] &= [\mathrm{vec}(B)]'(\Psi^{-1}\otimes\Sigma^{-1})[\mathrm{vec}(A)]\\ &= [(\Psi^{-\frac{1}{2}}\otimes\Sigma^{-\frac{1}{2}})\mathrm{vec}(B)]'[(\Psi^{-\frac{1}{2}}\otimes\Sigma^{-\frac{1}{2}})\mathrm{vec}(A)]\\ &= [\mathrm{vec}(\Sigma^{-\frac{1}{2}}B\Psi^{-\frac{1}{2}})]'[\mathrm{vec}(\Sigma^{-\frac{1}{2}}A\Psi^{-\frac{1}{2}})]\\ &= \langle \Sigma^{-\frac{1}{2}}A\Psi^{-\frac{1}{2}},\Sigma^{-\frac{1}{2}}B\Psi^{-\frac{1}{2}}\rangle_F\end{aligned}$$

where $\langle\cdot,\cdot\rangle_F$ is Frobenius inner product in $R^{p\times q}$. But

$$\langle \Sigma^{-\frac{1}{2}}A\Psi^{-\frac{1}{2}},\Sigma^{-\frac{1}{2}}B\Psi^{-\frac{1}{2}}\rangle_F=\mathrm{tr}[(\Sigma^{-\frac{1}{2}}B\Psi^{-\frac{1}{2}})'(\Sigma^{-\frac{1}{2}}A\Psi^{-\frac{1}{2}})].$$

Therefore the inner product in (3.1.1) can also be expressed as

$$\langle A,B\rangle=\mathrm{tr}[(\Sigma^{-\frac{1}{2}}B\Psi^{-\frac{1}{2}})'(\Sigma^{-\frac{1}{2}}A\Psi^{-\frac{1}{2}})]. \tag{3.1.2}$$

□

With the inner product in Lemma 3.1.1, a norm of a matrix in $R^{p\times q}$, a distance between A and B in $R^{p\times q}$, and an angle formed by A and B in $R^{p\times q}$ can be induced. With $\|\cdot\|=\sqrt{\langle\cdot,\cdot\rangle}$, the immediate benefit is that the probability density function of Z can be expressed as

$$\phi(Z)=\frac{1}{(2\pi)^{\frac{pq}{2}}|\Psi\otimes\Sigma|^{\frac{1}{2}}}\exp\bigg(-\frac{1}{2}\|Z-M\|^2\bigg), Z\in R^{p\times q}. \tag{3.1.3}$$

3.2 Projections

Treating the function in (3.1.3) as a function of M, when M is confined in $\mathcal{B} \subset R^{p\times q}$, the function is maximized over $M \in \mathcal{B}$ at $\widehat{M} \in \mathcal{B}$ such that $\|Z - \widehat{M}\| \leq \|Z - M\|$ for all $M \in \mathcal{B}$. This $\widehat{M}$ is called the minimum distance projection of Z onto $\mathcal{B}$ and is denoted by $\pi(Z|\mathcal{B})$.

Definition 3.2.1. $\widehat{X}$ is the minimum distance projection of $X \in R^{p\times q}$ onto $\mathcal{B} \subset R^{p\times q}$ and is denoted as $\pi(X|\mathcal{B})$ if $\widehat{X} \in \mathcal{B}$ and $\|X - \widehat{X}\| \leq \|X - Y\|$ for all $Y \in \mathcal{B}$.

$\pi(X|\mathcal{B})$ may not exist for all $\mathcal{B}$. But generally, if $\mathcal{B}$ is a closed convex set, then $\pi(X|\mathcal{B})$ exists, is unique, and

$$\widehat{X} = \pi(X|\mathcal{B}) \Leftrightarrow \widehat{X} \in \mathcal{B} \text{ and } \langle X - \widehat{X}, \widehat{X} - Y\rangle \geq 0 \text{ for all } Y \in \mathcal{B} \tag{3.2.1}$$

see Zarantonello (1971).

If $\mathcal{B}$ is a linear space $\mathcal{L}$, then $\pi(X|\mathcal{L})$ exists, is unique, and is also the orthogonal projection in the sense that

$$\widehat{X} = \pi(X|\mathcal{L}) \Leftrightarrow \widehat{X} \in \mathcal{L} \text{ and } \langle X - \widehat{X}, Y\rangle = 0 \text{ for all } Y \in \mathcal{L}. \tag{3.2.2}$$

Clearly (3.2.2) is a special case of (3.2.1).

3.3 Closed convex cones

In between the concepts of a closed convex set and a finite dimensional linear space, there is a concept of a closed convex cone.

Definition 3.3.1. A subset $\mathcal{C}$ of $R^{p\times q}$ is a cone if $X \in \mathcal{C}$ implies $\alpha X \in \mathcal{C}$ for all $\alpha \geq 0$. If a cone is also a convex set, then it is called a convex cone. If a convex cone is also a closed set, then it is called a closed convex cone.

A linear space of finite dimension is a closed convex cone. A closed convex cone is a closed convex set. Therefore, the following lemma is presented.

Lemma 3.3.2. *For $X \in R^{p\times q}$ and a closed convex cone $\mathcal{C}, \pi(X|\mathcal{C})$ exists and is unique. The sufficient and necessary condition is*

$$\widehat{X} = \pi(X|\mathcal{C}) \Leftrightarrow \widehat{X} \in \mathcal{C}, \langle X - \widehat{X}, \widehat{X}\rangle = 0 \text{ and} \langle X - \widehat{X}, Y\rangle \leq 0 \text{ for all } Y \in \mathcal{C}. \tag{3.3.1}$$

Proof: Closed convex cone $\mathcal{C}$ is a closed convex set. So $\pi(X|\mathcal{C})$ exists and is unique.
" $\Rightarrow$ " : Suppose $\widehat{X} = \pi(X|\mathcal{C})$. By (3.2.1) $\widehat{X} \in \mathcal{C}$ and $\langle X - \widehat{X}, \widehat{X} - Y\rangle \geq 0$ for all $Y \in \mathcal{C}$. With $Y = 0 \in \mathcal{C}$, $\langle X - \widehat{X}, \widehat{X}\rangle \geq 0$. With $Y = 2\widehat{X} \in \mathcal{C}, \langle X - \widehat{X}, \widehat{X}\rangle \leq 0$. So $\langle X - \widehat{X}, \widehat{X}\rangle = 0$ and $\langle X - \widehat{X}, Y\rangle \leq 0$ for all $Y \in \mathcal{C}$.
" $\Leftarrow$ " : Suppose $\widehat{X} \in \mathcal{C}, \langle X - \widehat{X}, \widehat{X}\rangle = 0$ and $\langle X - \widehat{X}, Y\rangle \leq 0$ for all $Y \in \mathcal{C}$. Then $\widehat{X} \in \mathcal{C}$ and $\langle X - \widehat{X}, \widehat{X} - Y\rangle = \langle X - \widehat{X}, \widehat{X}\rangle - \langle X - \widehat{X}, Y\rangle \geq 0$ for all $Y \in \mathcal{C}$. By (3.2.1), $\widehat{X} = \pi(X|\mathcal{C})$. □

Obviously, (3.3.1) is a special case of (3.2.1), and (3.2.2) is a special case of (3.3.1).

Suppose $\mathcal{C}$ is a closed convex cone in $R^{p\times q}$ equipped with inner product $\langle\cdot,\cdot\rangle$ and its induced norm $\|\cdot\|$.

Lemma 3.3.3. *For closed convex cone $\mathcal{C} \subset R^{p\times q}$, define*

$$\mathcal{C}^p = \{X \in R^{p\times q} : \langle X, Y\rangle \leq 0 \text{ for all } Y \in \mathcal{C}\}.$$

Then $\mathcal{C}^p$ is also a closed convex cone in $R^{p\times q}$ called the polar cone of $\mathcal{C}$.

Proof: Suppose X_1 and X_2 are in $\mathcal{C}^p$. $\langle \alpha X_1 + \beta X_2, Y\rangle = \alpha\langle X_1, Y\rangle + \beta\langle X_2, Y\rangle \leq 0$ for all $\alpha, \beta \geq 0$ and $Y \in \mathcal{C}$. Thus $\alpha X_1 + \beta X_2 \in \mathcal{C}^p$. By Definition 3.3.1, $\mathcal{C}^p$ is a convex cone. Suppose $X_n \in \mathcal{C}^p$ and $X_n \to X$ as $n \to \infty$. With $Y \in \mathcal{C}$,
$|\langle X_n, Y\rangle - \langle X, Y\rangle| = |\langle X_n - X, Y\rangle| \leq \|X_n - X\|\cdot\|Y\| \to 0$ as $n \to \infty$. Thus $\langle X_n, Y\rangle \to \langle X, Y\rangle$. But for $Y \in \mathcal{C}$, $\langle X_n, Y\rangle \leq 0$. So $\langle X, Y\rangle \leq 0$. Hence, $X \in \mathcal{C}^p$, which implies that $\mathcal{C}^p$ is a closed set. Therefore, $\mathcal{C}^p$ is a closed convex cone in $R^{p\times q}$. □

Lemma 3.3.4. *Let $\mathcal{C}^p$ be the polar cone of $\mathcal{C}$ in $R^{p\times q}$. Then $\pi(X|\mathcal{C}^p) = X - \pi(X|\mathcal{C})$ and $\|X\|^2 = \|\pi(X|\mathcal{C})\|^2 + \|\pi(X|\mathcal{C}^p)\|^2$.*

Proof: First by (3.3.1), $\langle X - \pi(X|\mathcal{C}), Y\rangle \leq 0$ for all $Y \in \mathcal{C}$. Thus $X - \pi(X|\mathcal{C}) \in \mathcal{C}^p$. Secondly, $\langle X - [X - \pi(X|\mathcal{C})], X - \pi(X|\mathcal{C})\rangle = \langle \pi(X|\mathcal{C}), X - \pi(X|\mathcal{C})\rangle = 0$. Finally with $Y \in \mathcal{C}^p$, $\langle X - [X - \pi(X|\mathcal{C})], Y\rangle = \langle \pi(X|\mathcal{C}), Y\rangle \leq 0$. Therefore, by (3.3.1), $X - \pi(X|\mathcal{C}) = \pi(X|\mathcal{C}^p)$ is true.

By $\langle X - \pi(X|\mathcal{C}), \pi(X|\mathcal{C})\rangle = 0$ in (3.3.1), and the Pythagorean theorem, $\|X\|^2 = \|X - \pi(X|\mathcal{C})\|^2 + \|\pi(X|\mathcal{C})\|^2$. But $\pi(X|\mathcal{C}^p) = X - \pi(X|\mathcal{C})$. So $\|X\|^2 = \|\pi(X|\mathcal{C})\|^2 + \|\pi(X|\mathcal{C}^p)\|^2$ holds. $\square$

Definition 3.3.5. For a linear space $\mathcal{L}$, let $\mathcal{L}^\perp = \{X \in R^{p\times q} : \langle X, Y\rangle = 0 \text{ for all } Y \in \mathcal{L}\}$ be the orthogonal compliment of $\mathcal{L}$.

As a closed convex cone, $\mathcal{L}^\perp = \mathcal{L}^p$. By Lemma 3.3.4,

$$X = \pi(X|\mathcal{L}) + \pi(X|\mathcal{L}^\perp). \tag{3.3.2}$$

Restriction $X \in \mathcal{C}$ contains many interesting cases for the specifications on X. Multivariate order restriction cone is one of them which will be discussed later in Chapter 6.

CHAPTER 4

A distribution family and expected result

4.1 A distribution family

Suppose $\mathcal{L} \subset \mathcal{C} \subset R^{p\times q}$ where $\mathcal{L}$ is a linear space and $\mathcal{C}$ is a closed convex cone. Let Z be from (2.1.1), and U be a positive random variable that is independent to Z and is distributed free of the unknown parameter matrix M. Define

$$T(Z,U) = \frac{\|\pi(Z|\mathcal{C}) - \pi(Z|\mathcal{L})\|^2}{U + \|Z - \pi(Z|\mathcal{C})\|^2}. \tag{4.1.1}$$

$T(Z,U)$ is a random variable with a distribution depending on M. Denote the distribution by $T(M)$, i.e.,

$$T(Z,U) \sim T(M). \tag{4.1.2}$$

Then $T(M)$ gives a distribution family whose members are indexed by parameter matrix $M \in R^{p\times q}$. By selecting p, q, Σ and Ψ in (2.1.1), and selecting $\mathcal{L}, \mathcal{C}$ and U in (4.1.1), $T(M)$ yields many different distribution families. The next example shows that χ^2-distribution family and F-distribution family can be presented as $T(M)$.

Example 4.1.1. Suppose in (2.1.1), $q = 1$ and $\Psi = 1$. Then $M \in R^{p\times 1}$ becomes a vector in R^p. For convenience denote M by $\mu \in R^p$. Hence Z is now an ordinary normal vector $Z \sim N_{p\times 1}(\mu, \Sigma, 1) = N_p(\mu, \Sigma)$. If $\Sigma = I_p$, then the components of Z are independent normal variables that share the variance 1 but with possible different mean $\mu_i, i = 1, \cdots, p$. Note that the inner product in (3.1.1) becomes $\langle x, y\rangle = [\text{vec}(y)]'(1 \otimes I_p)^{-1}[\text{vec}(x)] = y'x$, the Frobenius inner product for vectors in R^p. Consequently $\|Z\|^2 = Z'Z = Z_1^2 + \cdots + Z_p^2$ where $Z_i, i = 1, \cdots, p$, are the components of Z. Clearly $\|Z\|^2 \sim \chi^2(\mu'\mu, p)$.

In (4.1.1), let $\mathcal{C} = R^p$. Then $\pi(Z|\mathcal{C}) = \pi(Z|R^p) = Z$. Let $\mathcal{L} = \{0\} \subset R^p$. Then $\pi(Z|\mathcal{L}) = \pi(Z|0) = 0$. Therefore, $T(Z,U) = \frac{\|Z-0\|^2}{U+\|Z-Z\|^2} = \frac{\|Z\|^2}{U}$.

In the case of $U \equiv 1$, $T(Z, U) = T(Z, 1) = \|Z\|^2 \sim \chi^2(\mu'\mu, p)$ where $\mu'\mu$ is the non-centrality parameter. Thus non-central χ^2-distribution family is the family $T(\mu)$.

In the case of $U \sim \frac{p}{k}X^2$ where the positive random variable $X^2 \sim \chi^2(k)$ is independent to Z, $T(Z, U) = T(Z, \frac{p}{k}X^2) = \frac{\|Z\|^2/p}{X^2/k}$. Because $\|Z\|^2 \sim \chi^2(\mu'\mu, p)$ and $X^2 \sim \chi^2(k)$ are independent, $T(Z, U) \sim F(\mu'\mu, p, k)$ where $\mu'\mu$ is the non-centrality parameter. Therefore the non-central F-distribution is also $T(\mu)$.

4.2 Stochastic order

Suppose X and Y are two random variables. We say that the distribution of X is stochastically less than or equal to the distribution of Y, or simply X is stochastically less than or equal to Y denoted by $X \leq_{st} Y$ if $P(X \leq t) \geq P(Y \leq t)$ for all t, or equivalently $P(X > t) \leq P(Y > t)$ for all t.

One can read $P(X \leq t) \geq P(Y \leq t)$ as "The probability that X is small is greater than or equal to the probability that Y is small." Here small X and small Y are measured by the small scale. Similarly one can read $P(X > t) \leq P(Y > t)$ as "The probability that X is large is less than or equal to the probability that Y is large." Again here large X and large Y are measured by the same scale.

As a relation for univariate distributions $\leq_{st}$ is reflexive, i.e., $X \leq_{st} X$ for all X. The relation $\leq_{st}$ is also transitive, i.e., $X \leq_{st} Y$ and $Y \leq_{st} W$ imply $X \leq_{st} W$. Therefore the relation $\leq_{st}$ is a pre-order, or quasi order for distributions. $X \sim U(1, 2)$ and $Y \sim U(3, 4)$ by definition provide an example of $X \leq_{st} Y$ since $P(X > t) \leq P(Y > t)$ for all $t \in R$.

4.3 Stochastic monotonicity

Suppose $V_0 \in R^{p\times q}$ and $0 \neq V_1 \in R^{p\times q}$. Then $\{V_0 + V_1 t : t \geq 0\}$ is a subset of $R^{p\times q}$. The matrices in this set form a ray in $R^{p\times q}$ with one end V_0. Let

$$M(t) = V_0 + V_1 t,\ t \geq 0 \tag{4.3.1}$$

be a moving matrix on the line segment. When t goes from 0 to ∞, $M(t)$ is moving from V_0 in the direction of V_1 further and further. Thus this set is a matrix ray originated at V_0, in the direction of V_1.

For the distribution $T(M)$ in (4.1.2), when M is on the ray in (4.3.1) we identify the conditions on V_0 and V_1 such that the distribution $T(M(t)) = T(V_0 + V_1 t)$ is stochastically non-decreasing on the ray with increasing t, i.e., $T(V_0 + V_1 t_1) \leq_{st} T(V_0 + V_1 t_2)$ for all $0 \leq t_1 < t_2$. The expected condition is that $V_0 \in \mathcal{L}$ and $V_1 \in \mathcal{C}$ where $\mathcal{L}$ and $\mathcal{C}$ are from (4.1.1). This result is formally stated in the Theorem below.

Theorem 4.3.1. $T(V_0+V_1 t_1) \leq_{st} T(V_0+V_1 t_2)$ *for all* $V_0 \in \mathcal{L}$, *all* $V_1 \in \mathcal{C}$ *and all* $0 \leq t_1 < t_2$.

While the proof of the theorem is placed in the next chapter, a corollary of the theorem is presented and proved here.

Corollary 4.3.2. $T(V_0) \leq_{st} T(V_1)$ for all $V_0 \in \mathcal{L}$ and all $V_1 \in \mathcal{C}$.

Proof: For $V_0 \in \mathcal{L}$ and $V_1 \in \mathcal{C}, V_1 - V_0 \in \mathcal{C}$. By Theorem 4.3.1,

$$T(V_0 + (V_1 - V_0)t_1) \leq_{st} T(V_0 + (V_1 - V_0)t_2) \text{ for all } 0 \leq t_1 < t_2.$$

Taking $t_1 = 0$ and $t_2 = 1$ leads to $T(V_0) \leq_{st} T(V_1)$. $\square$

Example 4.3.3. For $T(\mu)$ in Example 4.1.1 with $\mu_0 \in \mathcal{L} = \{0\}$ and $\mu_1 \in \mathcal{C} = R^p$, the ray $M(t)$ in (4.3.1) is $\mu(t) = 0 + \mu_1 t = \mu_1 t,\ t \geq 0$. On this ray for $T(\mu(t)) = \chi^2([\mu(t)]'\mu(t),\, p) = \chi^2(\|\mu_1\|^2 t^2,\, p)$, $\|\mu_1\|^2 t^2$ is non-decreasing as $t \to \infty$ from $t = 0$. Thus we conclude that the non-central χ^2-distributions are stochastically monotone non-decreasing with its increasing non-centrality parameter. For $T(\mu(t)) = F([\mu(t)]'\mu(t), p, k) = F(\|\mu_1\|^2 t^2, p, k)$, we conclude that the non-central $F-$distributions are stochastically monotone non-decreasing with its increasing non-centrality parameter.

4.4 An application of stochastic monotonicity property

Stochastic monotonicity for a random variable or a statistic if established is very useful in the study of related statistical inferences. For example let $X_1, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$, the standard $Z-$test on $H_0 : \mu = \mu_0$ versus $H_\alpha : \mu > \mu_0$ uses test statistic $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ and rejects H_0 when $Z > Z_\alpha$. Note that $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$. Thus $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(\frac{\mu-\mu_0}{\sigma/\sqrt{n}}, 1^2)$. On $\mu(t) = \mu_0 + 1 \cdot t, t \geq 0, Z = Z(\mu(t)) \sim N(\frac{t}{\sigma/\sqrt{n}}, 1^2)$. It can be shown that $Z(\mu)$ is stochastically non-decreasing on the ray when t goes from 0 to ∞, i.e., $P(Z(\mu(t_1)) > c) \leq P(Z(\mu(t_2)) > c)$ for all $c > 0$ and $0 \leq t_1 < t_2$. With $c = Z_\alpha, P(Z(\mu(t_1)) > Z_\alpha)$ is the power of the test at $\mu = \mu(t)$. Therefore one can conclude that the power of the test is non-decreasing on the ray. Other properties for the test including the unbiasedness of the test follow.

CHAPTER 5

Establishing the stochastic monotonicity

5.1 A simplified sufficient condition

In this section it is shown that the conclusion of Theorem 4.3.1,

$$T(V_0 + V_1 t_1) \leq_{st} T(V_0 + V_1 t_2) \text{ for all } V_0 \in \mathcal{L}, \text{ all } V_1 \in \mathcal{C} \text{ and all } 0 \leq t_1 < t_2$$

is implied by

$$T(Vt_1) \leq_{st} T(Vt_2) \text{ for all } 0 \neq V \in \mathcal{C} \cap \mathcal{L}^{\perp} \text{ and } 0 \leq t_1 < t_2. \tag{5.1.1}$$

This sufficient condition is achieved in two steps in Lemma 5.1.1 and Lemma 5.1.2.

Lemma 5.1.1. $T(M) = T(\pi(M|\mathcal{L}^{\perp}))$.

Proof: For Z in (2.1.1) let $Y = Z - \pi(M|\mathcal{L})$. By (3.3.1) and (3.2.2) it can be shown that $\pi(Y|\mathcal{C}) = \pi(Z|\mathcal{C}) - \pi(M|\mathcal{L})$ and $\pi(Y|\mathcal{L}) = \pi(Z|\mathcal{L}) - \pi(M|\mathcal{L})$. Consequently by (4.1.1), $T(Y, U) = T(Z, U)$.

For M in (2.1.1), by (3.3.2) $M - \pi(M|\mathcal{L}) = \pi(M|\mathcal{L}^{\perp})$. Thus while $Z \sim N_{p\times q}(M, \Sigma, \Psi)$, by Lemma 2.2.2, $Y \sim N_{p\times q}(\pi(M|\mathcal{L}^{\perp}), \Sigma, \Psi)$. Therefore while $T(Z, U) \sim T(M)$, $T(Y, U) \sim T(\pi(M|\mathcal{L}^{\perp}))$. The conclusion on the equality of the distribution $T(M)$ and the distribution $T(\pi(M|\mathcal{L}^{\perp}))$ follows. □

By Lemma 5.1.1, with $V_0 \in \mathcal{L}$ and $V_1 \in \mathcal{C}$, $\pi(V_0 + V_1 t|\mathcal{L}^{\perp}) = 0 + \pi(V_1|\mathcal{L}^{\perp})t$. Thus the inequality in Theorem 4.3.1 becomes $T(Vt_1) \leq_{st} T(Vt_2)$ where $V = \pi(V_1|\mathcal{L}^{\perp})$.

Lemma 5.1.2. *If* $V_1 \in \mathcal{C}$, $\pi(V_1|\mathcal{L}^{\perp}) \in \mathcal{C} \cap \mathcal{L}^{\perp}$.

Proof: Suppose $V_1 \in \mathcal{C}$, by (3.3.2) $\pi(V_1|\mathcal{L}^\perp) = V_1 - \pi(V_1|\mathcal{L})$ where both V_1 and $-\pi(V_1|\mathcal{L})$ are in $\mathcal{C}$. Thus $\pi(V_1|\mathcal{L}^\perp) \in \mathcal{C}$. This projection is clearly in $\mathcal{L}^\perp$. So $\pi(V_1|\mathcal{L}^\perp) \in \mathcal{C} \cap \mathcal{L}^\perp$. $\square$

Based on the result of Lemma 5.1.2, $T(Vt_1) \leq_{st} T(Vt_2)$ for $V = \pi(V_1|\mathcal{L}^\perp)$ is implied by the same inequality for all $V \in \mathcal{C} \cap \mathcal{L}^\perp$. But $T(Vt_1) = T(Vt_2)$ is trivially true when $V = 0$. Therefore, the sufficient condition in (5.1.1) for the conclusion of Theorem 4.3.1 is established.

5.2 Introducing Z_1, Z_2 and $\mathcal{A}$

$T(Vt_1) \leq_{st} T(Vt_2)$ in (5.1.1) means $P(T(Z_1, U) \leq c) \geq P(T(Z_2, U) \leq c)$ for all $c > 0$ where

$$Z_1 \sim N_{p\times q}(Vt_1, \Sigma, \Psi) \text{ and } Z_2 \sim N_{p\times q}(Vt_2, \Sigma, \Psi). \tag{5.2.1}$$

With $u > 0$ and $c > 0$. Define

$$\mathcal{A} = \{X \in R^{p\times q} : T(X, u) \leq c\}. \tag{5.2.2}$$

Then $P(Z_1 \in \mathcal{A}) \geq P(Z_2 \in \mathcal{A})$ becomes a sufficient condition for (5.1.1), and in turn a sufficient condition for the conclusion of Theorem 4.3.1. This result is studied in Theorem 5.2.1 below.

Theorem 5.2.1. *If $P(Z_1 \in \mathcal{A}) \geq P(Z_2 \in \mathcal{A})$, then Theorem 4.3.1 holds.*

Proof: By the independence of Z_i and $U, i = 1, 2$, for all $u \geq 0$,

$$P(T(Z_i, U) \leq c|U = u) = P(T(Z_i, u) \leq c|U = u) = P(T(Z_i, u) \leq c) = P(Z_i \in \mathcal{A}).$$

So $P(Z_1 \in \mathcal{A}) \geq P(Z_2 \in \mathcal{A})$ implies that $P(T(Z_1, u) \leq c|u) \geq P(T(Z_2, u) \leq c|u)$. Consequently,

$$E_u[P(T(Z_1, u) \leq c|u)] \geq E_u[P(T(Z_2, u) \leq c|u)].$$

But

$$E_U[P(T(Z_i, U) \leq c|U)] = P(T(Z_i, U) \leq c) = P(T(Vt_i) \leq c).$$

Thus $P(T(Vt_1) \leq c) \geq P(T(Vt_2) \leq c)$ which is the sufficient condition in (5.1.1) for the conclusion of Theorem 4.3.1. Therefore, the theorem is eatablished. $\square$

$\mathcal{A}$ in (5.2.2) depends on $u > 0$ and $c > 0$. It is only for the simplicity notation $\mathcal{A}(u, c)$ was not used. Thus $P(Z_1 \in \mathcal{A}) \geq P(Z_2 \in \mathcal{A})$ actually means for all $u > 0$ and all $c > 0$.

To establish Theorem 4.3.1 in the rest of this chapter, the effort will be made to establish the inequality in Theorem 5.2.1.

5.3 A mirror image transformation

With $0 \neq V \in \mathcal{C} \cap \mathcal{L}^{\perp}$ and $0 \leq t_1 < t_2$, for $X \in R^{p\times q}$, define

$$Y = f(X) = X - \frac{2\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V. \tag{5.3.1}$$

This transformation is a mapping from $R^{p\times q}$ to $R^{p\times q}$. For $X \in R^{p\times q}$ and $Y = f(X) \in R^{p\times q}$, there is a simple and useful relation

$$\left\langle Y - \frac{t_1+t_2}{2}V, V\right\rangle = -\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle \tag{5.3.2}$$

which can be proved by direct computations

$$\begin{aligned}
\left\langle Y - \frac{t_1+t_2}{2}V, V\right\rangle &= \left\langle X - \frac{2\langle X - \frac{t_1+t_2}{2}V, V\rangle}{\|V\|^2}V - \frac{t_1+t_2}{2}V, V\right\rangle \\
&= \left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle - 2\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle \\
&= -\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle.
\end{aligned}$$

Based on the relation in (5.3.2), Lemma 5.3.1 can be established.

Lemma 5.3.1. *For $Y = f(X)$ in (5.3.1), $f(Y) = f[f(X)] = X$.*

Proof: By (5.3.1),

$$\begin{aligned} f[f(X)] &= f(Y) \\ &= Y - \frac{2\left\langle Y - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V \\ &= f(X) - \frac{2\left\langle Y - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V \\ &= X - \frac{2\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V - \frac{2\left\langle Y - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V. \end{aligned}$$

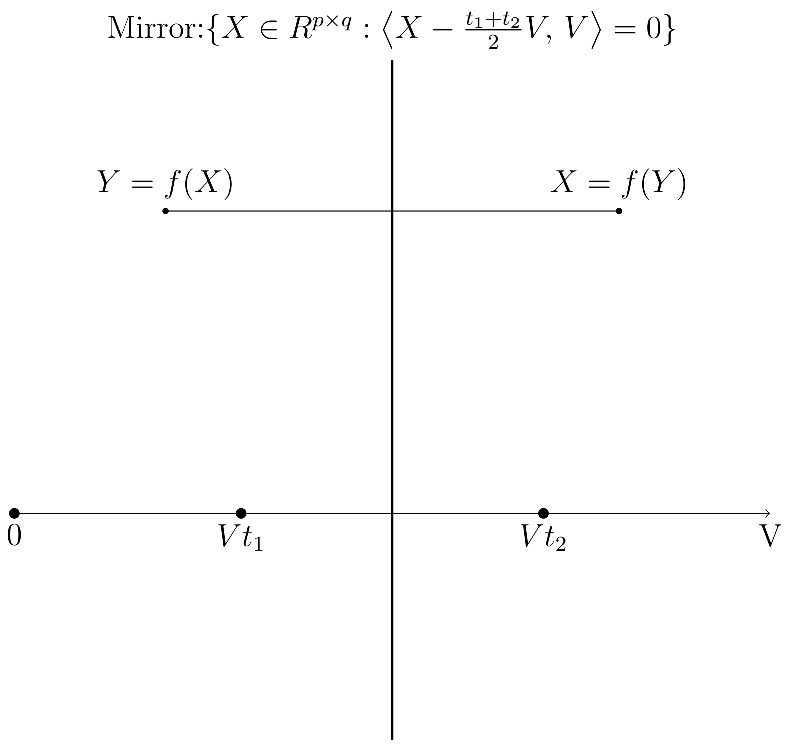
By (5.3.2),

$$\begin{aligned} f[f(X)] &= X - \frac{2\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V - \frac{2\left\langle Y - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V \\ &= X - \frac{2\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V + \frac{2\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V \\ &= X. \ \square \end{aligned}$$

Let $f(X_1) = f(X_2)$, then $f[f(X_1)] = f[f(X_2)]$. By Lemma 5.3.1, $X_1 = X_2$. So $Y = f(X)$ is an injection. For $Y \in R^{p\times q}$, let $X = f(Y)$. By Lemma 5.3.1, $f(X) = f(f(Y)) = Y$. Thus $Y = f(X)$ is a surjection. Therefore $Y = f(X)$ is a one-to-one transformation in $R^{p\times q}$ with an inverse $f^{-1}(\cdot)$.

Suppose $X = f^{-1}(Y)$. Then $f(X) = Y$. Again by Lemma 5.3.1, $X = f[f(X)] = f(Y)$. So $f^{-1}(Y) = f(Y)$, i.e., $f^{-1} = f$.

One can visualize the transformation by imagining a two-sided mirror in $R^{p\times q}$ at $\frac{t_1+t_2}{2}V$ perpendicular to V that divided the space into two parts. For $X \in R^{p\times q}$ in one part of the space, its image is in the other part. This mirror image of X is $Y = f(X)$.



5.4 Distribution of transformed Z

Set $\mathcal{A} \subset R^{p\times q}$ in (5.2.2), just like the space $R^{p\times q}$, is also divided into two parts by the mirror. The part in the direction of ray Vt, $t \geq 0$, in (5.1.1) is defined as $\mathcal{A}_1$,

$$\mathcal{A}_1 = \left\{ X \in R^{p\times q} : X \in \mathcal{A} \text{ and } \left\langle X - \frac{t_1 + t_2}{2} V, V \right\rangle > 0 \right\}. \tag{5.4.1}$$

The image of $\mathcal{A}_1$ is defined as $\mathcal{A}_2$,

$$\mathcal{A}_2 = \{Y \in R^{p\times q} : Y = f(X) \text{ for some } X \in \mathcal{A}_1\}. \tag{5.4.2}$$

Since $\mathcal{A}_1$ is part of $\mathcal{A}$, $P(Z_i \in \mathcal{A}_1)$ is part of $P(Z_i \in \mathcal{A})$ in Theorem 5.2.1. With the one-to-one transformation $Y = f(X)$, $P(Z_i \in \mathcal{A}_1) = P(f(Z_i) \in f(\mathcal{A}_1)) = P(f(Z_i) \in \mathcal{A}_2)$. To study the impact of this transformation on the probability we need the distribution of $f(Z_i)$. The distribution of Z in (2.1.1) is described by that of vec(Z). Therefore vec$[f(X)]$ is needed and the result is presented in the lemma below.

Lemma 5.4.1. For $Y = f(X)$ in (5.3.1), $\text{vec}(Y) = G\text{vec}(X) + (t_1 + t_2)\text{vec}(V)$ where $G = I_{pq} - \frac{2\text{vec}(V)[\text{vec}(V)]'(\Psi\otimes\Sigma)^{-1}}{\|V\|^2}$.

Proof: By (5.3.1),

$$Y = f(X) = X - \frac{2\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle}{\|V\|^2}V = X - \frac{2\langle X, V\rangle}{\|V\|^2}V + (t_1+t_2)V.$$

So

$$\text{vec}(Y) = \text{vec}(X) - \frac{2\langle X, V\rangle}{\|V\|^2}\text{vec}(V) + (t_1+t_2)\text{vec}(V)$$

in which by (3.1.1)

$$\begin{aligned} \frac{2\langle X, V\rangle}{\|V\|^2}\text{vec}(V) &= \frac{2[\text{vec}(V)]'(\Psi\otimes\Sigma)^{-1}\text{vec}(X)}{\|V\|^2}\text{vec}(V) \\ &= \frac{2\text{vec}(V)[\text{vec}(V)]'(\Psi\otimes\Sigma)^{-1}}{\|V\|^2}\text{vec}(X). \end{aligned}$$

Therefore $\text{vec}(Y) = G\text{vec}(X) + (t_1+t_2)\text{vec}(V)$ where $G = I_{pq} - \frac{2\text{vec}(V)[\text{vec}(V)]'(\Psi\otimes\Sigma)^{-1}}{\|V\|^2}$. □

Matrix G has two useful properties

$$G[\text{vec}(V)] = -\text{vec}(V) \text{ and } G(\Psi\otimes\Sigma)G' = \Psi\otimes\Sigma. \tag{5.4.3}$$

Write $G = I - 2D$ where $D = \frac{[\text{vec}](V)[\text{vec}(V)]'(\Psi\otimes\Sigma)^{-1}}{\|V\|^2}$. Then

$$D[\text{vec}(V)] = \frac{[\text{vec}(V)][\text{vec}(V)]'(\Psi\otimes\Sigma)^{-1}}{\|V\|^2}[\text{vec}(V)] = \frac{[\text{vec}(V)]\|V\|^2}{\|V\|^2} = \text{vec}(V).$$

So $G[\text{vec}(V)] = (I-2D)[\text{vec}(V)] = -\text{vec}(V)$. The first equation in (5.4.3) holds.

Note that

$$G(\Psi\otimes\Sigma)G' = \Psi\otimes\Sigma - 2D(\Psi\otimes\Sigma) - 2(\Psi\otimes\Sigma)D' + 4D(\Psi\otimes\Sigma)D'.$$

For the second equation in (5.4.3) it suffices to show that

$$D(\Psi\otimes\Sigma) = (\Psi\otimes\Sigma)D' = D(\Psi\otimes\Sigma)D'.$$

By direct computation, we see that $D(\Psi\otimes\Sigma) = \frac{[\text{vec}(V)][\text{vec}(V)]'}{\|V\|^2}$ is symmetric. Therefore

$$D(\Psi\otimes\Sigma) = [D(\Psi\otimes\Sigma)]' = (\Psi\otimes\Sigma)'D' = (\Psi'\otimes\Sigma')D' = (\Psi\otimes\Sigma)D'.$$

On the other hand,

$$\begin{aligned}
D(\Psi\otimes\Sigma)D' &= \frac{[\mathrm{vec}(V)][\mathrm{vec}(V)]'(\Psi\otimes\Sigma)^{-1}}{\|V\|^2}(\Psi\otimes\Sigma)\frac{(\Psi\otimes\Sigma)^{-1}[\mathrm{vec}(V)][\mathrm{vec}(V)]'}{\|V\|^2} \\
&= \frac{[\mathrm{vec}(V)]}{\|V\|^2}[\mathrm{vec}(V)]'(\Psi\otimes\Sigma)^{-1}[\mathrm{vec}(V)]\frac{[\mathrm{vec}(V)]'}{\|V\|^2} \\
&= \frac{[\mathrm{vec}(V)]}{\|V\|^2}\|V\|^2\frac{[\mathrm{vec}(V)]'}{\|V\|^2} \\
&= \frac{[\mathrm{vec}(V)][\mathrm{vec}(V)]'}{\|V\|^2} \\
&= D(\Psi\otimes\Sigma).
\end{aligned}$$

So $D(\Psi\otimes\Sigma) = (\Psi\otimes\Sigma)D' = D(\Psi\otimes\Sigma)D'$ holds. Thus the second equation in (5.4.3) is established.

Theorem 5.4.2. *For random matrices Z_1 and Z_2 in (5.2.1) and transformation $Y = f(X)$ in (5.3.1), $f(Z_1) = Z_2$ in distributions and $f(Z_2) = Z_1$ in distributions.*

Proof: By (5.2.1) and (2.1.1), $Z_1 \sim N_{p\times q}(Vt_1, \Sigma, \Psi) \Leftrightarrow \mathrm{vec}(Z_1) \sim N_{pq}(\mathrm{vec}(V)t_1, \Psi\otimes\Sigma)$ and $Z_2 \sim N_{p\times q}(Vt_2, \Sigma, \Psi) \Leftrightarrow \mathrm{vec}(Z_2) \sim N_{pq}(\mathrm{vec}(V)t_2, \Psi\otimes\Sigma)$.

But by Lemma 5.4.1 and the two properties in (5.4.3),

$$\begin{aligned}
\mathrm{vec}[f(Z_1)] &= G[\mathrm{vec}(Z_1)] + (t_1+t_2)[\mathrm{vec}(V)] \\
&\sim N_{pq}(G[\mathrm{vec}(V)]t_1 + (t_1+t_2)[\mathrm{vec}(V)], G(\Psi\otimes\Sigma)G') \\
&= N_{pq}(-\mathrm{vec}(V)t_1 + (t_1+t_2)[\mathrm{vec}(V)], \Psi\otimes\Sigma) \\
&= N_{pq}(\mathrm{vec}(V)t_2, \Psi\otimes\Sigma).
\end{aligned}$$

Therefore, $\mathrm{vec}[f(Z_1)]$ and $\mathrm{vec}(Z_2)$ are equal in distributions. Consequently, $f(Z_1)$ and Z_2 are equal in distributions. In a similar way one can show that $f(Z_2)$ and Z_1 are equal in distributions. □

5.5 Image inclusive property

By Theorem 5.4.2, $P(Z_1 \in \mathcal{A}_1) = P(f(Z_1) \in f(\mathcal{A}_1)) = P(Z_2 \in \mathcal{A}_2)$. Here $P(Z_1 \in \mathcal{A}_1)$ is part of $P(Z_1 \in \mathcal{A})$ in Theorem 5.2.1. But whether or not $P(Z_2 \in \mathcal{A}_2)$ is part of $P(Z_2 \in \mathcal{A})$

in Theorem 5.4.2 depends on if $\mathcal{A}_2$ is in $\mathcal{A}$. The property $\mathcal{A}_2 \subset \mathcal{A}$, if established, is called the image inclusive property since $\mathcal{A}_2$ is the image of $\mathcal{A}_1$ by the mapping $f(\cdot)$ in (5.3.1). This section is devoted to establishing this property.

Lemma 5.5.1. *$T(X, u)$ in (5.2.2) can be rewritten as*

$$T(X,u) = \frac{\|X\|^2 - \|X - \pi(X|\mathcal{C} \cap \mathcal{L}^{\perp})\|^2}{u + \|X - \pi(X|\mathcal{C})\|^2}.$$

Proof: By (4.1.1), to establish the lemma

$$\|\pi(X|\mathcal{C}) - \pi(X|\mathcal{L})\|^2 = \|X\|^2 - \|X - \pi(X|\mathcal{C} \cap \mathcal{L}^{\perp})\|^2$$

is needed.

As $\mathcal{L}^{\perp} = \mathcal{L}^p$ and $\mathcal{L}^p$ is a closed convex cone, $\mathcal{L}^{\perp}$ is a closed convex cone. Then the intersection of two closed convex cones, $\mathcal{C} \cap \mathcal{L}^{\perp}$, is also a closed convex cone.

By Lemma 3.3.4, $X - \pi(X|\mathcal{C} \cap \mathcal{L}^{\perp}) = \pi(X|(\mathcal{C} \cap \mathcal{L}^{\perp})^p)$. By Lemma 3.3.4 again, $\|X\|^2 - \|\pi(X|(\mathcal{C} \cap \mathcal{L}^{\perp})^p)\|^2 = \|\pi(X|\mathcal{C} \cap \mathcal{L}^{\perp})\|^2$. Therefore one only needs to show $\pi(X|\mathcal{C}) - \pi(X|\mathcal{L}) = \pi(X|\mathcal{C} \cap \mathcal{L}^{\perp})$ for which we check the conditions in (3.3.1).

First of all, $\pi(X|\mathcal{C}) - \pi(X|\mathcal{L}) \in \mathcal{C}$ since both $\pi(X|\mathcal{C})$ and $-\pi(X|\mathcal{L})$ are in $\mathcal{C}$; and $\pi(X|\mathcal{C}) - \pi(X|\mathcal{L}) \in \mathcal{L}^{\perp}$ since for $Y \in \mathcal{L}$,

$$\begin{aligned} \langle \pi(X|\mathcal{C}) - \pi(X|\mathcal{L}), Y \rangle &= \langle \pi(X|\mathcal{C}) - X + X - \pi(X|\mathcal{L}), Y \rangle \\ &= \langle \pi(X|\mathcal{C}) - X, Y \rangle + \langle X - \pi(X|\mathcal{L}), Y \rangle \\ &= 0 + 0 = 0. \end{aligned}$$

We conclude that $\pi(X|\mathcal{C}) - \pi(X|\mathcal{L}) \in \mathcal{C} \cap \mathcal{L}^{\perp}$.

Secondly,

$$\begin{aligned}
& \langle X - [\pi(X|\mathcal{C}) - \pi(X|\mathcal{L})], \pi(X|\mathcal{C}) - \pi(X|\mathcal{L})\rangle \\
= & \langle X - \pi(X|\mathcal{C}), \pi(X|\mathcal{C}) - \pi(X|\mathcal{L})\rangle + \langle \pi(X|\mathcal{L}), \pi(X|\mathcal{C}) - \pi(X|\mathcal{L})\rangle \\
= & 0 + \langle \pi(X|\mathcal{L}), \pi(X|\mathcal{C}) - \pi(X|\mathcal{L})\rangle \\
= & \langle \pi(X|\mathcal{L}), \pi(X|\mathcal{C}) - X + X - \pi(X|\mathcal{L})\rangle \\
= & \langle \pi(X|\mathcal{L}), \pi(X|\mathcal{C}) - X\rangle + \langle \pi(X|\mathcal{L}), X - \pi(X|\mathcal{L})\rangle \\
= & 0 + 0 = 0.
\end{aligned}$$

Finally, for $Z \in \mathcal{C} \cap \mathcal{L}^{\perp}$,

$$\langle X - [\pi(X|\mathcal{C}) - \pi(X|\mathcal{L})], Z\rangle = \langle X - \pi(X|\mathcal{C}),\ Z\rangle + \langle \pi(X|\mathcal{L}), Z\rangle$$

where $\langle X - \pi(X|\mathcal{C}), Z\rangle \leq 0$ since $Z \in \mathcal{C}$, and $\langle \pi(X|\mathcal{L}), Z\rangle = 0$ since $Z \in \mathcal{L}^{\perp}$. Therefore, $\langle X - [\pi(X|\mathcal{C}) - \pi(X|\mathcal{L})], Z\rangle \leq 0$ is true. The conclusion is proved. □

Corollary 5.5.2. For X and Y in $R^{p\times q}$, if $|Y - \pi(Y|\mathcal{C})\|^2 \geq \|X - \pi(X|\mathcal{C})\|^2$, $\|Y\|^2 \leq \|X\|^2$ and $\|Y - \pi(Y|\mathcal{C} \cap \mathcal{L}^{\perp})\|^2 \geq \|X - \pi(X|\mathcal{C} \cap \mathcal{L}^{\perp})\|^2$, then $T(Y, u) \leq T(X, u)$.

The corollary is a direct result of Lemma 5.5.1. The proof is skipped.

We are now ready to establish the image inclusive property.

Lemma 5.5.3. *Let $\mathcal{A}$ and $\mathcal{A}_2$ be as defined as in (5.2.2) and (5.4.2). Then $\mathcal{A}_2 \subset \mathcal{A}$.*

Proof: Let $Y \in \mathcal{A}_2$. One needs to show $Y \in \mathcal{A}$, i.e., $T(Y, u) \leq c$. Note that by (5.4.2) $Y \in \mathcal{A}_2$ implies $Y = f(X)$ where $X \in \mathcal{A}_1$, i.e., $\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle > 0$ and $T(X, u) \leq c$. Therefore it suffices to show $T(Y, u) \leq T(X, u)$. By Corollary 5.5.2 one needs to establish the three inequalities in the conditions of that corollary.

By (5.3.1), in

$$\begin{aligned}
\|Y-\pi(Y|\mathcal{C})\|^2 &= \|f(X)-\pi(Y|\mathcal{C})\|^2 \\
&= \left\| X - \frac{2\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}V - \pi(Y|\mathcal{C}) \right\|^2 \\
&= \left\| X - \left[\frac{2\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}V + \pi(Y|\mathcal{C})\right] \right\|^2
\end{aligned}$$

$\frac{2\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}V + \pi(Y|\mathcal{C})$ is in $\mathcal{C}$ since $V \in \mathcal{C}\cap\mathcal{L}^\perp \subset \mathcal{C}$ and $\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle > 0$. Thus by the definition of $\pi(X|\mathcal{C})$, $\|Y-\pi(Y|\mathcal{C})\|^2 \geq \|X-\pi(X|\mathcal{C})\|^2$. So the first inequality in the conditions for Corollary 5.5.2 is true.

By (5.3.1) and direct computations

$$\begin{aligned}
\|Y\|^2 &= \|f(X)\|^2 \\
&= \left\| X - \frac{2\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}V \right\|^2 \\
&= \|X\|^2 + \frac{4\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle^2}{\|V\|^2} - \frac{4\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}\left\langle X-\frac{t_1+t_2}{2}V+\frac{t_1+t_2}{2}V,V\right\rangle \\
&= \|X\|^2 - 2(t_1+t_2)\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle \\
&< \|X\|^2
\end{aligned}$$

since $\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle > 0$. Thus the second inequality in the conditions for Corollary 5.5.2 holds.

Now in

$$\begin{aligned}
\|Y-\pi(Y|\mathcal{C}\cap\mathcal{L}^\perp)\|^2 &= \|f(X)-\pi(Y|\mathcal{C}\cap\mathcal{L}^\perp)\|^2 \\
&= \left\| X - \frac{2\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}V - \pi(Y|\mathcal{C}\cap\mathcal{L}^\perp) \right\|^2 \\
&= \left\| X - \left[\frac{2\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}V + \pi(Y|\mathcal{C}\cap\mathcal{L}^\perp)\right] \right\|^2
\end{aligned}$$

$\frac{2\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle}{\|V\|^2}V + \pi(Y|\mathcal{C}\cap\mathcal{L}^\perp)$ is in $\mathcal{C}\cap\mathcal{L}^\perp$ since $V \in \mathcal{C}\cap\mathcal{L}^\perp$ and $\left\langle X-\frac{t_1+t_2}{2}V,V\right\rangle > 0$. Thus by the definition of $\pi(X|\mathcal{C}\cap\mathcal{L}^\perp)$, $\|Y-\pi(Y|\mathcal{C}\cap\mathcal{L}^\perp)\|^2 \geq \|X-\pi(X|\mathcal{C}\cap\mathcal{L}^\perp)\|^2$. So the last inequality in the conditions for Corollary 5.5.2 is also true.

Thus $T(Y,u) \leq T(X,u) \leq c$. Hence $Y \in \mathcal{A}$. $\square$

5.6 Establishing the stochastic monotonicity

In this section the stochastic monotonicity for the distribution family stated in Theorem 4.3.1 is to be proved by showing the sufficient condition $P(Z_1 \in \mathcal{A}) \geq P(Z_2 \in \mathcal{A})$ in Theorem 5.2.1. The proof will make use of the results in Lemma 5.6.1.

Lemma 5.6.1. *For $\mathcal{A}, \mathcal{A}_1$ and $\mathcal{A}_2$ in (5.2.2), (5.4.1) and (5.4.2) define*

$$\mathcal{A}_3 = \mathcal{A} \cap (\mathcal{A}_1 \cup \mathcal{A}_2)^c.$$

Then

(i) $\mathcal{A}$ is partitioned by $\mathcal{A}_1, \mathcal{A}_2$ and $\mathcal{A}_3$.

(ii) $\|X - Vt_1\|^2 \leq \|X - Vt_2\|^2$ for all $X \in \mathcal{A}_3$.

Proof:

(i) $\mathcal{A}$ is partitioned by $\mathcal{A}_1, \mathcal{A}_2$ and $\mathcal{A}_3$ if $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$, $\mathcal{A}_1 \cap \mathcal{A}_3 = \emptyset$, $\mathcal{A}_2 \cap \mathcal{A}_3 = \emptyset$ and $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$. The first three equations hold by the definition of $\mathcal{A}_3$. To show $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$, let $Y \in A_2$. Then there is $X \in \mathcal{A}_1$ such that $Y = f(X)$. Thus $\left\langle X - \frac{t_1+t_2}{2}V, V \right\rangle > 0$ since $X \in \mathcal{A}_1$ and $\left\langle Y - \frac{t_1+t_2}{2}V, V \right\rangle = -\left\langle X - \frac{t_1+t_2}{2}V, V \right\rangle$ by (5.3.2). Clearly $\left\langle Y - \frac{t_1+t_2}{2}V, V \right\rangle < 0$ which means $Y \notin \mathcal{A}_1$. Therefore $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$.

(ii) Write

$$\begin{aligned}
\|X - Vt_1\|^2 &= \left\| X - \frac{t_1+t_2}{2}V + \frac{t_2-t_1}{2}V \right\|^2 \\
&= \left\| X - \frac{t_1+t_2}{2}V \right\|^2 + \left\| \frac{t_2-t_1}{2}V \right\|^2 + 2\left\langle X - \frac{t_1+t_2}{2}V, \frac{t_2-t_1}{2}V \right\rangle \\
&= \left\| X - \frac{t_1+t_2}{2}V \right\|^2 + \left\| \frac{t_2-t_1}{2}V \right\|^2 + (t_2-t_1)\left\langle X - \frac{t_1+t_2}{2}V, V \right\rangle
\end{aligned}$$

and

$$\begin{aligned}\|X - Vt_2\|^2 &= \left\|X - \frac{t_1+t_2}{2}V - \frac{t_2-t_1}{2}V\right\|^2 \\ &= \left\|X - \frac{t_1+t_2}{2}V\right\|^2 + \left\|\frac{t_2-t_1}{2}V\right\|^2 - 2\left\langle X - \frac{t_1+t_2}{2}V, \frac{t_2-t_1}{2}V\right\rangle \\ &= \left\|X - \frac{t_1+t_2}{2}V\right\|^2 + \left\|\frac{t_2-t_1}{2}V\right\|^2 - (t_2-t_1)\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle.\end{aligned}$$

When $X \in \mathcal{A}_3$, $X \in \mathcal{A}$ but $X \notin \mathcal{A}_1$. Thus $\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle \leq 0$. Hence

$$\|X - Vt_1\|^2 - \|X - Vt_2\|^2 = 2(t_2-t_1)\left\langle X - \frac{t_1+t_2}{2}V, V\right\rangle \leq 0.$$

Conclusion $\|X - Vt_1\|^2 \leq \|X - Vt_2\|^2$ follows. $\square$

We are now ready to establish Theorem 4.3.1.

By (i) of Lemma 5.6.1 and the one-to-one transformation $f(\cdot)$ in (5.3.1),

$$\begin{aligned}P(Z_1 \in \mathcal{A}) &= P(Z_1 \in \mathcal{A}_1) + P(Z_1 \in \mathcal{A}_2) + P(Z_1 \in \mathcal{A}_3) \\ &= P(f(Z_1) \in f(\mathcal{A}_1)) + P(f(Z_1) \in f(\mathcal{A}_2)) + P(Z_1 \in \mathcal{A}_3).\end{aligned}$$

By Theorem 5.4.2, $f(Z_1)$ and Z_2 are equal in distributions. Note that $f(\mathcal{A}_1) = \mathcal{A}_2$ and $f(\mathcal{A}_2) = \mathcal{A}_1$. So

$$P(Z_1 \in \mathcal{A}) = P(Z_2 \in \mathcal{A}_2) + P(Z_2 \in \mathcal{A}_1) + P(Z_1 \in \mathcal{A}_3).$$

But

$$P(Z_2 \in \mathcal{A}) = P(Z_2 \in \mathcal{A}_1) + P(Z_2 \in \mathcal{A}_2) + P(Z_2 \in \mathcal{A}_3).$$

Thus $P(Z_1 \in \mathcal{A}) \geq P(Z_2 \in \mathcal{A})$ is equivalent to $P(Z_1 \in \mathcal{A}_3) \geq P(Z_2 \in \mathcal{A}_3)$. Let $\phi_{Z_1}(X)$ and $\phi_{Z_2}(X)$ are the probability density functions of Z_1 and Z_2 respectively. By (3.1.3),

$$\phi_{Z_1}(X) = \frac{1}{(2\pi)^{\frac{pq}{2}}|\Psi \otimes \Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\|X - Vt_1\|^2\right)$$

and

$$\phi_{Z_2}(X) = \frac{1}{(2\pi)^{\frac{pq}{2}}|\Psi \otimes \Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\|X - Vt_2\|^2\right).$$

By (ii) of Lemma 5.6.1, $\phi_{Z_1}(X) \geq \phi_{Z_2}(X)$ on $X \in \mathcal{A}_3$. Therefore $P(Z_1 \in \mathcal{A}_3) \geq P(Z_2 \in \mathcal{A}_3)$. Consequently, $P(Z_1 \in \mathcal{A}) \geq P(Z_2 \in \mathcal{A})$. By Theorem 5.2.1, the stochastic monotonicity on $T(M)$ in Theorem 4.3.1 is confirmed.

CHAPTER 6

A multivariate order restricted model

6.1 An MANOVA model

Consider an MANOVA model with q p-dimensional normal populations $N_p(\mu_i, \sigma^2\Sigma)$, $i = 1, \cdots, q$, where $\Sigma \in R^{p\times p}$ is a known positive definite matrix. But $\mu_i \in R^p$, $i = 1, \cdots, q$, and the proportional factor σ^2 are unknown. The integrated expression for these q populations is

$$N_{p\times q}(M, \sigma^2\Sigma, I_q) \text{ where } M = (\mu_1, \cdots, \mu_q) \in R^{p\times q}. \tag{6.1.1}$$

The common interpretation for the model is that the q populations are the responses to q treatments for q levels of a factor. Since the purpose is to determine if the factor is effective in affecting the responses, the hypothesis $H_0 : \mu_1 = \cdots = \mu_q$ is of interest. This H_0 is called a null hypothesis on the homogeneity of response means. The rejection of H_0 implies that the factor is effective while the failure in rejecting H_0 will be interpreted as the useless of the model.

Let $\mathcal{L}$ be the collection of all matrices whose columns satisfy the homogeneity assumption in H_0, i.e.,

$$\mathcal{L} = \{(\mu_1, \cdots, \mu_q) \in R^{p\times q} : \mu_1 = \cdots = \mu_q\}. \tag{6.1.2}$$

Then $\mathcal{L}$ is a linear space in $R^{p\times q}$ that can be written as $\mathcal{L} = \{\mu 1_q', \ \mu \in R^p\}$ where $1_q = (1, \cdots, 1)' \in R^q$. With this linear space the null hypotheses can now be expressed as

$$H_0 : M \in \mathcal{L}.$$

In traditional MANOVA without any restrictions, M could be anywhere in $R^{p\times q}$. The test on the homogeneity of the response vectors is a test on $H_0 : M \in \mathcal{L}$ versus $H_\alpha : M \in R^{p\times q}$

which means H_0 against $H_\alpha - H_0 : M \in \mathcal{L}^c$. However in this project we consider the model under multivariate order restrictions on M.

6.2 Multivariate order

Sasabuchi et al. (2003) studied an MANOVA model. Its response mean vectors satisfy the relations $\mu_1 \leq \cdots \leq \mu_q$ where "$\mu_i \leq \mu_j$" means the components of $\mu_j - \mu_i$ are all non-negative. This relation "$\leq$" for vectors, just like the relation $\leq$ for real numbers , are reflexive and transitive, and hence is a quasi order. With respect to the two linear operations, the relation "$\leq$" in R^p, just like the relation $\leq$ in R, is preserved under linear combinations with non-negative coefficients, i.e., $x \leq y$ and $u \leq v$ imply $\alpha x + \beta u \leq \alpha y + \beta v$ for all $\alpha, \beta \geq 0$. Moreover, with respect to the norms induced from inner products, the relation "$\leq$" in R^p, just like the relation $\leq$ in R, is preserved under the limits, i.e., $x_n \leq y_n, x_n \to x$ and $y_n \to y$ imply $x \leq y$. Hu (2018) referred to such "$\leq$" in R^p as a multivariate order.

Definition 6.2.1. A relation of vectors in R^p, $\preceq$, is called a multivariate order if

(i) $\preceq$ is reflexive and transitive;

(ii) $\preceq$ is preservable under linear combinations with non-negative coefficients;

(iii) $\preceq$ is preservable under limits.

Multivariate order covers many interesting relations of vectors in R^p. For example if $x = (x_1, x_2, x_3, x_4)'$ and $y = (y_1, y_2, y_3, y_4)'$ are in R^4, then $x \preceq y$ defined by $x_i \leq y_i$ for all $i = 1, \cdots, 4$ is a multivariate order, see Sasabuchi et al. (2003); $x \preceq y$ defined by $x_1 \leq y_1, x_2 \geq y_2, x_3 = y_3$ and no restrictions on x_4 and y_4 is also a multivariate order. Hu and Banerjee (2012) pointed out that the relation generated from a closed convex cone in R^p called the cone association by Cohen and Sackrowitz (1996) is a multivariate order.

6.3 Multivariate order restrictions

For model (6.1.1), suppose that based on prior knowledge we knew that $\mu_i \preceq \mu_j$ for all (i, j) in a pre-specified $H \subset \{1, \cdots, q\} \times \{1, \cdots, q\}$ where $\preceq$ is a defined multivariate order. Then we say that the MANOVA model is under a multivariate order restriction.

Let $\mathcal{C}$ be the collection of all matrices whose columns satisfy the multivariate order restrictions, i.e.,

$$\mathcal{C} = \{(\mu_1, ..., \mu_q) \in R^{p\times q} : \mu_i \preceq \mu_j \text{ for all } (i, j) \in H\}. \tag{6.3.1}$$

Then the model restriction can be expressed as $M \in \mathcal{C}$. This restriction in essence is a closed convex cone restriction since as shown in the next lemma $\mathcal{C}$ is a closed convex cone.

Lemma 6.3.1. *Let $\mathcal{C}$ be defined as in (6.3.1). Then $\mathcal{C}$ is a closed convex cone in $R^{p\times q}$.*

Proof: Suppose $A = (A_1, \cdots, A_q)$ and $B = (B_1, \cdots, B_q)$ are both in $\mathcal{C}$. Then $A_i \preceq A_j$ and $B_i \preceq B_j$ for all $(i, j) \in H$. With $\alpha, \beta \geq 0$, $\alpha A + \beta B = (\alpha A_1 + \beta B_1, \cdots, \alpha A_q + \beta B_q)$ in which $\alpha A_i + \beta B_i, \leq \alpha A_j + \beta B_j$ since $\preceq$ is preservable under linear combinations with non-negative coefficients. Therefore $\alpha A + \beta B \in \mathcal{C}$. With $\beta = 0$ we see that $\alpha A \in \mathcal{C}$. Thus $\mathcal{C}$ is a cone. With $\alpha \in (0, 1)$ and $\beta = 1 - \alpha$, we see that $\alpha A + (1 - \alpha)B \in \mathcal{C}$. Hence $\mathcal{C}$ is a convex set.

Suppose $A^{(n)} = (A_1^{(n)}, \cdots, A_q^{(n)}) \in \mathcal{C}$ and $A^{(n)} \to A = (A_1, \cdots, A_q)$ with respect to a norm induced from an inner product in $R^{p\times q}$. Then $A_i^{(n)} \preceq A_j^{(n)}$, for all $(i, j) \in H$ and $A_i^{(n)} \to A_i$ for all $i = 1, \cdots, q$ with respect to the norms induced from the inner products in R^p since the convergence in $R^{p\times q}$ and the convergence in R^p are all componentwise convergence. Therefore $A_i \preceq A_j$ for all $(i, j) \in H$ since $\preceq$ is preserved by limits. Thus $A \in \mathcal{C}$. Hence $\mathcal{C}$ is a closed set. Therefore $\mathcal{C}$ is a closed convex cone in $R^{p\times q}$. □

Thus the model in (6.1.1) is an MANOVA model with order restriction $M \in \mathcal{C}$ where $\mathcal{C}$ is an order restricted cone in (6.3.1). To determine the usefulness, a test on $H_0 : M \in \mathcal{L}$ with $\mathcal{L}$ in (6.1.2) is needed.

6.4 Samples and basic statistics

Let $Y_i = (Y_{i1}, \cdots, Y_{in_i}) \in R^{p\times n_i}$ be a random sample from the ith population $N(\mu_i, \sigma^2\Sigma)$. Then

$$Y_i \sim N_{p\times n_i}(\mu_i 1'_{n_i}, \sigma^2\Sigma, I_{n_i})$$

with sample size n_i; simple mean $\overline{Y}_i = \frac{Y_i 1_{n_i}}{n_i} \in R^p$; and sample CSSCP matrix

$$\begin{aligned}
\text{CSSCP}_i &= \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)(Y_{ij} - \overline{Y}_i)' \\
&= (Y_i - \overline{Y}_i 1'_{n_i})(Y_i - \overline{Y}_i 1'_{n_i})' \\
&= Y_i(I_{n_i} - \frac{1_{n_i}1'_{n_i}}{n_i})Y_i' \in R^{p\times p}.
\end{aligned}$$

With $i = 1, \cdots, q, Y = (Y_1, \cdots Y_q) \in R^{p\times n}$ is the pooled sample of size n. Clearly

$$Y \sim N_{p\times n}(MJ', \sigma^2\Sigma, I_n) \text{ where } J = \begin{pmatrix} 1_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1_{n_q} \end{pmatrix} \in R^{n\times q}. \tag{6.4.1}$$

Define $\overline{Y} = (\overline{Y}_1, \cdots, \overline{Y}_q) \in R^{p\times q}$ and CSSCP=$\text{CSSCP}_1 + \cdots + \text{CSSCP}_q \in R^{p\times p}$. Then $\overline{Y} = YJ(J'J)^{-1}$ and $\text{CSSCP} = Y[I_n - J(J'J)^{-1}J']Y'$.

By (6.4.1) and Lemma 2.2.2,

$$\overline{Y} \sim N_{p\times q}(M, \sigma^2\Sigma, (J'J)^{-1})$$

and

$$\overline{Y}/\sigma \sim N_{p\times q}(M/\sigma, \Sigma, (J'J)^{-1}).$$

By (6.4.1) and Lemma 2.3.2,

$$\text{CSSCP} \sim W_p(n - q, 0, \sigma^2\Sigma)$$

and further by Lemma 2.3.1

$$\text{CSSCP}/\sigma^2 \sim W_p(n - q, 0, \Sigma).$$

By Lemma 2.2.3,

$$\overline{Y}/\sigma \text{ and CSSCP}/\sigma^2 \text{ are independent.}$$

We therefore obtain the Lemma below that summarizes the above results.

Lemma 6.4.1. *For the statistic matrices $\overline{Y}$ and CSSCP defined in this section, $\overline{Y}/\sigma \sim N_{p\times q}(M/\sigma, \Sigma, (J'J)^{-1})$; CSSCP/$\sigma^2$ has Wishart distribution $W_p(n-q, \Sigma)$ that is free of both M and σ, and CSSCP/σ^2 is independent to $\overline{Y}/\sigma$.*

CHAPTER 7

A likelihood ratio test

In this Chapter for MANOVA model $N_{p\times q}(M,\sigma^2\Sigma,I_q)$ in (6.1.1) under the multivariate order restriction $M\in\mathcal{C}$ where $\mathcal{C}$ is in (6.3.1) we develop LRT procedure on the null hypothesis $H_0: M\in\mathcal{L}$ with $\mathcal{L}$ in (6.1.2) and show that the test is unbiased.

7.1 Likelihood function

Based on sample $Y\sim N_{p\times n}(MJ',\sigma^2\Sigma,I_n)$ in (6.4.1), the likelihood function of M and σ^2 is

$$L(M,\sigma^2)=\frac{1}{(2\pi)^{\frac{pn}{2}}|\Sigma|^{\frac{n}{2}}}\cdot\frac{1}{(\sigma^2)^{\frac{pn}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)'\Sigma^{-1}(Y_{ij}-\mu_i)\right]$$

where

$$\begin{aligned}
&\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\mu_i)'\Sigma^{-1}(Y_{ij}-\mu_i)\\
=\;&\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i+\overline{Y}_i-\mu_i)'\Sigma^{-1}(Y_{ij}-\overline{Y}_i+\overline{Y}_i-\mu_i)\\
=\;&\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)'\Sigma^{-1}(Y_{ij}-\overline{Y}_i)+\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)'\Sigma^{-1}(\overline{Y}_i-\mu_i)\\
&+\sum_{i=1}^{q}\sum_{j=1}^{n_i}(\overline{Y}_i-\mu_i)'\Sigma^{-1}(Y_{ij}-\overline{Y}_i)+\sum_{i=1}^{q}\sum_{j=1}^{n_i}(\overline{Y}_i-\mu_i)'\Sigma^{-1}(\overline{Y}_i-\mu_i).
\end{aligned}$$

But

$$\begin{aligned}
\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)'\Sigma^{-1}(\overline{Y}_i-\mu_i) &= \sum_{i=1}^{q}\left[\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)'\right]\Sigma^{-1}(\overline{Y}_i-\mu_i)\\
&= \sum_{i=1}^{q}0\Sigma^{-1}(Y_{ij}-\overline{Y}_i)=0,\\
\sum_{i=1}^{q}\sum_{j=1}^{n_i}(\overline{Y}_i-\mu_i)'\Sigma^{-1}(Y_{ij}-\overline{Y}_i) &= \left[\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)'\Sigma^{-1}(\overline{Y}_i-\mu_i)\right]'=0,
\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)'\Sigma^{-1}(Y_{ij}-\overline{Y}_i) &= \operatorname{tr}\left[\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)'\Sigma^{-1}(Y_{ij}-\overline{Y}_i)\right] \\ &= \sum_{i=1}^{q}\sum_{j=1}^{n_i}\operatorname{tr}[(Y_{ij}-\overline{Y}_i)'\Sigma^{-1}(Y_{ij}-\overline{Y}_i)] \\ &= \sum_{i=1}^{q}\sum_{j=1}^{n_i}\operatorname{tr}[\Sigma^{-1}(Y_{ij}-\overline{Y}_i)(Y_{ij}-\overline{Y}_i)'] \\ &= \operatorname{tr}\left[\Sigma^{-1}\sum_{i=1}^{q}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)(Y_{ij}-\overline{Y}_i)'\right] \\ &= \operatorname{tr}\left[\Sigma^{-1}\left(\sum_{i=1}^{q}\text{CSSCP}_i\right)\right] \\ &= \operatorname{tr}[\Sigma^{-1}(\text{CSSCP})],\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{q}\sum_{j=1}^{n_i}(\overline{Y}_i-\mu_i)'\Sigma^{-1}(\overline{Y}_i-\mu_i) &= \sum_{i=1}^{q}(\overline{Y}_i-\mu_i)'n_i\Sigma^{-1}(\overline{Y}_i-\mu_i) \\ &= [\operatorname{vec}(\overline{Y}-M)]'(N\otimes\Sigma^{-1})[\operatorname{vec}(\overline{Y}-M)]\end{aligned}$$

with $N = J'J = \operatorname{diag}(n_1, \cdots, n_q)$.

Then the likelihood function can be expressed as

$$L(M,\sigma^2) = \frac{1}{(2\pi)^{\frac{pn}{2}}|\Sigma|^{\frac{n}{2}}}\cdot\frac{1}{(\sigma^2)^{\frac{pn}{2}}}\exp\left(-\frac{d}{2\sigma^2}\right)$$

with $d = \operatorname{tr}[\Sigma^{-1}(\text{CSSCP})] + [\operatorname{vec}(\overline{Y}-M)]'(N\otimes\Sigma^{-1})[\operatorname{vec}(\overline{Y}-M)]$.

To further simplify the expression, in $R^{p\times q}$ define the inner product

$$\langle A, B\rangle = [vec(B)]'(N^{-1}\otimes\Sigma)^{-1}[vec(A)]. \tag{7.1.1}$$

With respect the norm induced from this inner product,

$$[\operatorname{vec}(\overline{Y}-M)]'(N\otimes\Sigma^{-1})[\operatorname{vec}(\overline{Y}-M)] = \|\overline{Y}-M\|^2.$$

We therefore obtain

$$L(M,\sigma^2) = \frac{1}{(2\pi)^{\frac{pn}{2}}|\Sigma|^{\frac{n}{2}}}\cdot\frac{1}{(\sigma^2)^{\frac{pn}{2}}}\exp\left[-\frac{\operatorname{tr}[\Sigma^{-1}(\text{CSSCP})]+\|\overline{Y}-M\|^2}{2\sigma^2}\right]. \tag{7.1.2}$$

The inner product in (7.1.1) is actually the inner product in (3.1.1) with Z in (2.1.1) replaced by $\overline{Y}/\sigma$ in Lemma 6.4.1. Consequently, Ψ is replaced by N^{-1}.

7.2 Maximized likelihood function

When M is restricted in a closed convex set $\mathcal{B} \subset R^{p\times q}$,

$$\begin{aligned} L(M,\sigma^2) &= \frac{1}{(2\pi)^{\frac{pn}{2}}|\Sigma|^{\frac{n}{2}}}\cdot\frac{1}{(\sigma^2)^{\frac{pn}{2}}}\exp\left[-\frac{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})]+\|\overline{Y}-M\|^2}{2\sigma^2}\right] \\ &\le \frac{1}{(2\pi)^{\frac{pn}{2}}|\Sigma|^{\frac{n}{2}}}\cdot\frac{1}{(\sigma^2)^{\frac{pn}{2}}}\exp\left[-\frac{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})]+\|\overline{Y}-\pi(\overline{Y}|\mathcal{B})\|^2}{2\sigma^2}\right] \\ &= L(\pi(\overline{Y}|\mathcal{B}),\sigma^2). \end{aligned}$$

Here $L(\pi(\overline{Y}|\mathcal{B}),\sigma^2)$ is a function of σ^2. Consider $\ln[L(\pi(\overline{Y}|\mathcal{B}),\sigma^2)]$. By simple computations one can obtain the stationary point, the point at which the first derivative of the function is zero, $\widehat{\sigma}^2 = \widehat{d}/(pn)$ with

$$\widehat{d} = \mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y}-\pi(\overline{Y}|\mathcal{B})\|^2.$$

The second order derivative at this point is $-\frac{(pn)^3}{(\sqrt{2}\widehat{d})^2} < 0$. Thus by the second order derivative test,

$$L(\pi(\overline{Y}|\mathcal{B}),\sigma^2) \le L(\pi(\overline{Y}|\mathcal{B}),\widehat{\sigma}^2) = h\widehat{d}^{-\frac{pn}{2}}$$

with $h = \left(\frac{np}{2\pi e|\Sigma|^{\frac{1}{p}}}\right)^{\frac{np}{2}}$.

Therefore under the model restriction $M \in \mathcal{C}$, the maximized value of the likelihood function is

$$\max[L(M,\sigma^2): M\in\mathcal{C}] = h\{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y}-\pi(\overline{Y}|\mathcal{C})\|^2\}^{-\frac{np}{2}}. \tag{7.2.1}$$

Under the restriction imposed by $H_0 : M \in \mathcal{L}$, the maximized value of the likelihood function is

$$\max[L(M,\sigma^2): M\in\mathcal{L}] = h\{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y}-\pi(\overline{Y}|\mathcal{L})\|^2\}^{-\frac{np}{2}}. \tag{7.2.2}$$

7.3 A likelihood ratio test

The test on the homogeneity of response mean vectors in the MANOVA under a multivariate order restriction is that on $H_0 : M \in \mathcal{L}$ versus $H_\alpha : M \in \mathcal{C}$. For this test the likelihood ratio is

$$\Lambda = \frac{\max[L(M, \sigma^2) : M \in \mathcal{C}]}{\max[L(M, \sigma^2) : M \in \mathcal{L}]}.$$

By (7.2.1) and (7.2.2),

$$\Lambda = \left\{ \frac{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y} - \pi(\overline{Y}|\mathcal{L})\|^2}{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2} \right\}^{\frac{np}{2}}.$$

This is an increasing function of

$$T = \frac{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y} - \pi(\overline{Y}|\mathcal{L})\|^2}{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2} - 1 = \frac{\|\overline{Y} - \pi(\overline{Y}|\mathcal{L})\|^2 - \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2}{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2}.$$

But by (3.2.2) and (3.3.1),

$$[\overline{Y} - \pi(\overline{Y}|\mathcal{C})] \perp [\pi(\overline{Y}|\mathcal{C}) - \pi(\overline{Y}|\mathcal{L})]$$

since

$$\begin{aligned} & \langle \overline{Y} - \pi(\overline{Y}|\mathcal{C}), \pi(\overline{Y}|\mathcal{C}) - \pi(\overline{Y}|\mathcal{L}) \rangle \\ = \; & \langle \overline{Y} - \pi(\overline{Y}|\mathcal{C}), \pi(\overline{Y}|\mathcal{C}) \rangle - \langle \overline{Y} - \pi(\overline{Y}|\mathcal{C}), \pi(\overline{Y}|\mathcal{L}) \rangle \\ = \; & 0 - 0 = 0. \end{aligned}$$

Thus, by Pythagorean theorem, $\|\pi(\overline{Y}|\mathcal{C}) - \pi(\overline{Y}|\mathcal{L})\|^2 = \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2 - \|\overline{Y} - \pi(\overline{Y}|\mathcal{L})\|^2$.

Therefore,

$$T = \frac{\|\pi(\overline{Y}|\mathcal{C}) - \pi(\overline{Y}|\mathcal{L})\|^2}{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2}. \qquad (7.3.1)$$

Thus T is a LRT statistic and H_0 is rejected for the large values of T, say for $T > c_0$.

7.4 Unbiasedness of the test

For the likelihood ratio test on $H_0 : M \in \mathcal{L}$ versus $H_\alpha : M \in \mathcal{C}$ using T in (7.3.1) as the test statistic with rejection rule, $T > c_0$, the distribution of T depends on both M and σ^2. Thus the probability of the rejection of H_0 is a function of M and σ^2 denoted as $P(T > c_0|M, \sigma^2)$. This test is unbiased if

$$P(T > c_0|M = V_0, \sigma^2) \leq P(T > c_0|M = V_1, \sigma^2)$$

for all $V_0 \in \mathcal{L}$, all $V_1 \in \mathcal{C}$ and all σ^2. In this section, we apply the result in Corollary 4.3.2 to establish the unbiasedness for the test.

7.4.1 Distribution of the test statistic

Divide the numerator and denominator of the test statistic T derived in last section by σ^2. Note that by (7.1.1), $\|A\|^2/\sigma^2 = \|A/\sigma\|^2$; by (3.2.2), $\pi(A|\mathcal{L})/\sigma = \pi(A/\sigma|\mathcal{L})$; by (3.3.1), $\pi(A|\mathcal{C})/\sigma = \pi(A/\sigma|\mathcal{C})$.

$$\begin{aligned} T &= \frac{\|\pi(\overline{Y}|\mathcal{C}) - \pi(\overline{Y}|\mathcal{L})\|^2}{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})] + \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2} \\ &= \frac{\|\pi(\overline{Y}|\mathcal{C}) - \pi(\overline{Y}|\mathcal{L})\|^2/\sigma^2}{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})]/\sigma^2 + \|\overline{Y} - \pi(\overline{Y}|\mathcal{C})\|^2/\sigma^2} \\ &= \frac{\|\pi(\overline{Y}/\sigma|\mathcal{C}) - \pi(\overline{Y}/\sigma|\mathcal{L})\|^2}{\mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})/\sigma^2] + \|\overline{Y}/\sigma - \pi(\overline{Y}/\sigma|\mathcal{C})\|^2}. \qquad (7.4.1) \end{aligned}$$

By Lemma 6.4.1, the distribution of CSSCP/σ^2 is $W_p(n-q, \Sigma)$ free of M and σ. Therefore

$$U = \mathrm{tr}[\Sigma^{-1}(\mathrm{CSSCP})/\sigma^2]$$

is a non-negative random variable with distribution free of M and σ. By Lemma 6.4.1 again,

$$\overline{Y}/\sigma \sim N_{p\times q}(M/\sigma, \Sigma, N^{-1})$$

is independent to CSSCP/σ^2, and hence is independent to U. Hence by (7.4.1) and (4.1.2), the distribution of the test statistic T fits the distribution family discussed in Chapter 4. Specifically,

$$T \sim T(M/\sigma).$$

This distribution depends on M and σ only through M/σ. Define

$$\Theta = M/\sigma. \tag{7.4.2}$$

Then the distribution of the test statistic can be written as $T \sim T(\Theta)$.

7.4.2 The unbiasedness of the test

Because $\sigma > 0$, by (7.4.2) it is easy to see that $M \in \mathcal{C}$ if and only if $\Theta \in \mathcal{C}$; and $M \in \mathcal{L}$ if and only if $\Theta \in \mathcal{L}$. Thus the hypothesis $H_0 : M \in \mathcal{L}$ and $H_\alpha : M \in \mathcal{C}$ become $H_0 : \Theta \in \mathcal{L}$ and $H_\alpha : \Theta \in \mathcal{C}$. For the LRT on

$$H_0 : \Theta \in \mathcal{L} \text{ versus } H_\alpha : \Theta \in \mathcal{C}$$

the probability of the rejection of the null hypothesis is

$$P(T > c_0|M, \sigma) = P(T > c_0|\Theta) = P(T(\Theta) > c_0|\Theta).$$

The unbiasedness of the test can now be characterized by

$$P(T(\Theta) > c_0|\Theta = V_0) \leq P(T(\Theta) > c_0|\Theta = V_1) \text{ for all } V_0 \in \mathcal{L} \text{ and all } V_1 \in \mathcal{C}.$$

By Corollary 4.3.2, T at V_0 is stochastically less than or equal to T at V_1, i.e.,

$$P(T(\Theta) > c|\Theta = V_0) \leq P(T(\Theta) > c|\Theta = V_1) \text{ for all } c.$$

With $c = c_0$, $P(T(\Theta) > c_0|\Theta = V_0) \leq P(T(\Theta) > c_0|\Theta = V_1)$ for all $V_0 \in \mathcal{L}$ and all $V_1 \in \mathcal{C}$. Thus the unbiasedness of the LRT in last section is established.

CHAPTER 8

Conclusion

In this research, a normal random matrix

$$Z \sim N_{p\times q}(M, \Sigma, \Psi)$$

is considered. Using the positive definite matrices Σ and Ψ in the distribution of Z, an inner product

$$\langle A, B\rangle = \mathrm{tr}[(\Sigma^{-\frac{1}{2}} B \Psi^{-\frac{1}{2}})'(\Sigma^{-\frac{1}{2}} A \Psi^{-\frac{1}{2}})]$$

is defined in $R^{p\times q}$. With respect to the norm induced from the inner product, the projection of matrix $X \in R^{p\times q}$ onto a closed convex set $\mathcal{B} \subset R^{p\times q}$, $\pi(X|\mathcal{B})$, is explored. As special closed convex sets, for closed convex cone $\mathcal{C}$ and a linear space $\mathcal{L}$, $\pi(X|\mathcal{C})$ and $\pi(X|\mathcal{L})$ are also introduced.

The distribution of a random variable

$$T(Z, U) = \frac{\|\pi(Z|\mathcal{C}) - \pi(Z|\mathcal{L})\|^2}{U + \|Z - \pi(Z|\mathcal{C})\|^2}$$

form a distribution family $T(M)$ with members indexed by M. It is shown in the study that $T(V_0)$ is stochastically less than or equal to $T(V_1)$ for all $V_0 \in \mathcal{L}$ and all $V_1 \in \mathcal{C}$.

The null hypothesis on the homogeneity of treatment mean vectors in MANOVA can be expressed as

$$H_0 : M \in \mathcal{L}$$

where $\mathcal{L}$ is a linear space. If the mean vectors are constrained by a multivariate order restriction, this model specification can be expressed as

$$H_\alpha : M \in \mathcal{C}$$

where $\mathcal{C}$ is the order restricted cone. For the likelihood ratio test on H_0 versus H_α, the monotonicity of the distributions in the family $T(M)$ is successfully applied to establish the unbiasedness.

REFERENCES

REFERENCES

[1] Arnold, S (1981). *The Theory of Linear Models and Multivariate Analysis*. Wiley Series in Probability and Mathematical Statistics.

[2] Cohen, A. and Sackrowitz, H. B. (1996). *Cone order association and stochastic cone ordering with applications to order-restricted testing*. Ann. Statist., 24, 2036-2048.

[3] Conaway, M., Pillers, C., Rovertson, T. and Sconing J. (1990). *The poser of the circular cone test: A noncentral chi-bar-squared distribution*. Canana. J. Statist., 18, 63-70.

[4] Conaway, M., Pillers, C., Rovertson, T. and Sconing J. (1991). *A circular-cone test for testing homogeneity against a simple tree order*. Canana. J. Statist., 19, 283-296.

[5] Gupta, A. K. and Varga, T. (1993), *Elliptically Contoured Models in Statistics*. Kluwer Academic Publishers.

[6] Hansohm, J. and Hu, X. (2012) *A convergence algorithm for a generalized multivariate isotonic regression problem* . Statistical Papers, 53 (1), 107-115.

[7] Hog, R., Mckean, J., Craig, A. (2005), *Introduction to Mathematical Statistics*, Pearson-Prentice Hall.

[8] Hu, X.(2018) *A pseudo restricted MLE under multivariate order restrictions and its algorithm*. Commun. Stat. Theory, https//doi.org/10.1080/03610926.2018.1535072

[9] Hu, X. and Banerjee, A. (2012) *On the test for the homogeneity of a parameter matrix with some rows constrained by synchronized order restrictions*. Journal of Multivariate Analysis, Vol. 107, 64-70.

[10] Hu, X., Hansohm, J., Hoffmann, L. and Zohner, Y. E. (2012) *On the convergence of row modification algorithm for matrix projections*. Journal of Multivariate Analysis, Vol. 105, 216-221.

[11] Hu, X. and Wright, F. T. (1994) *Monotonicity properties of the power functions of likelihood ratio tests for normal mean hypotheses constrained by a linear space and a cone*. The annual of statistics, Vol. 22, No. 3, 1547-1554.

[12] Johnson, R. A. and Wichern, D. W. (2007) *Applied Multivariate Statistical Analysis*, Pearson-Prentice Hall .

[13] Luenberger, D. G. (1969) *Optimization by Vector Space Methods*. Wiley, New York.

[14] Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York.

[15] Mukerjee, H., Robertson, T. and Wright, F. T. (1986) *A probability inequality for elliptically contoured densities with applications in order restricted inference*. The annual of statistics, Vol. 14, No. 4, 1544-1554.

[16] Pan, J. and Fang, K (2002) *Growth Curve Models and Statistical Diagnostics*. Springer, New York.

[17] Sasabuchi, S., Inutsuka, M. and Kulatunga, D. D. S. (1992), *An algorithm for computing multivariate isotonic regression*. Hiroshima Mathematical Journal, 22, 551-560.

[18] Sasabuchi, S., Tanaka, K. and Tsukamoto, T. (2003), *Testing homogeneity of multivariate normal mean vectors under an order restriction when the covariance matrices are common but unknown*. Ann. Statisti., 31, 1517-1536.

[19] Shiryaev, A. N. (2016), *Probability-1*. Springer, New York.

[20] Zarantonello, E. H. (1971), *Projections on convex sets in Hilbert space and spectral theory . Contributions to Nonlinear Functional Analysis (E. h. Zarantonello, Ed)*. 237-427. Academic Press, New York.