**THE LIKELIHOOD RATIO TEST OF NORMAL
MEAN WITH HYPOTHESIS DETERMINED
BY A CONVEX POLYHEDRAL CONE
AND THE MONOTONICITY OF
ITS POWER FUNCTION**

By

Kentaro NOMAKUCHI

(Received April 12, 1983)

§1. Introduction

Let X be normally distributed with the mean vector $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$ and the known variance matrix Σ . Let L be a subspace of R^p , which may consist of one point only, and let C be a convex polyhedral cone containing L as its proper subset. In this paper we consider a problem of the likelihood ratio test of the null hypothesis $H: \mu \in L$ against the alternative $AH: \mu \in C - L$. The special types of this test are treated in various papers. The ordinary one-sided or two-sided tests are regarded as the one. So is the test with the order restricted alternative which is treated in Barlow et al. [1] and Robertson et al. [9]. Also, Kudô [4] and Neusch [7] studied independently the case that C is the positive orthant. As far as X is normally distributed, these results essentially depend only on the fact that C is a convex polyhedral cone and contains L as its proper subset.

On this viewpoint, Kudô and Choi [5] and Wynn [13] solved this problem in the case when Σ is non-singular and $L = \{0\}$. Wynn did not give the concrete form of the test statistic. Kudô and Choi treated the problem under restrictive conditions on C . Our purpose in this paper is to generalize their results after removing the restrictions.

We will, in §2, prepare ourselves with preliminaries and notations, in §3, treat the case of $\Sigma = I$ which coincides with the Wynn's result, in §4, generalize the results of §3 to the case when Σ is not necessarily non-singular, in §5, show the monotonicity of the power function, and in §6, give some examples.

§ 2. Preliminaries and notations

Let $N = \{a_1, a_2, \dots, a_n\}$, a set of non-zero vectors in R^p . For a non-empty subset $M \subset N$, let A_M be a matrix with $a_i \in M$ as its column vectors and let $L_M = \{z: A'_M z = 0\}$. The convex cone determined by N is $C_N = \{z: A'_N z \geq 0\}$. A face of C_N can be expressed as $F_M = \{z: A'_M z = 0, A'_{M^c} z > 0\}$ for some $M \subset N$, where $M^c = N - M$, and when $M = \phi$ or $M^c = \phi$, the equality or inequality involving them is understood to be vacuous. F_M may be empty for some M . When $F_M \neq \phi$, L_M is the smallest subspace containing F_M . For a convex cone $C_M = \{z: A'_M z \geq 0\}$, its interior is given by $C_M^o = \{z: A'_M z > 0\}$, which may be empty, and its dual cone is defined as $C_M^\perp = \{z: y'z \leq 0 \text{ for all } y \in C_M\}$. We use the conventions $L_\phi = C_\phi = C_\phi^o = R^p$, and note the relations

$$(1) \quad F_M = L_M \cap C_{M^c}^o,$$

$$(2) \quad C = \bigcup_{M \subset N} F_M, \quad \text{and}$$

$$(3) \quad F_{M_1} \cap F_{M_2} = \phi \quad \text{if } M_1 \neq M_2.$$

The orthogonal projection operator to $L_M^\perp$ ($M \neq \phi$), the orthogonal space of L_M , is given by the matrix

$$P_M = A_M(A'_M A_M)^- A'_M$$

where A^- is a generalized inverse of A . Also, we use the convention $P_\phi = 0$. Let I be the identity matrix. We recall

$$(4) \quad P_M^2 = P_M,$$

$$(5) \quad P'_M = P_M, \quad \text{and}$$

$$(6) \quad (I - P_M)x \text{ uniquely attains the minimum of } (x - \mu)'(x - \mu) \text{ for } \mu \in L_M.$$

For a symmetric non-negative definite matrix Σ , let $R(\Sigma)$ be the range space of Σ , i.e. the space generated by the columns of Σ , and let P_Σ be the orthogonal projection matrix to $R(\Sigma)$.

Let $N_p(\mu, \Sigma)$ be the p -variate normal distribution with the mean μ and the variance Σ . Let χ_p^2 denote the chi-square variable with $p(\geq 1)$ degrees of freedom and $Pr(\chi_0^2 = 0) = 1$. $X \sim N_p(\mu, \Sigma)$ means X is distributed in $N_p(\mu, \Sigma)$, and $Z \sim \chi_p^2$ means the distribution of Z is the same as that of χ_p^2 . Let " $A \Rightarrow B$ " and " $A \Leftrightarrow B$ "

mean "if A then B " and " A is equivalent to B " respectively.

The following trio specifies a testing problem,

- (i) random variable and its type of distribution,
- (ii) assumptions on parameters, and
- (iii) the parameter spaces of the null hypothesis H and the alternative AH .

Throughout this paper, we assume that (i) $X \sim N_p(\mu, \Sigma)$, where Σ is known. We will specify (ii) and (iii) whenever necessary.

§3. The test in the case of $\Sigma = I$

Let $\Sigma = I$, the null hypothesis be $H: \mu \in L_{N_0}$ and the alternative be $AH: \mu \in C_N - L_{N_0}$, where N_0 is a finite set of non-zero vectors containing N , i.e. L_{N_0} is a subspace contained in L_N . We assume $C_N - L_{N_0} \neq \emptyset$. Let x be a sample and let $\tilde{x}$ and $\hat{x}$ be the MLE's under H and AH , which are obviously orthogonal projections of x on L_{N_0} and C_N respectively. As for $\tilde{x}$ we have $\tilde{x} = (I - P_{N_0})x$, and in view of (2) and (3), for any x there is uniquely a subset M with $\hat{x} \in F_M$, for which we have

LEMMA 1. When $\hat{x} \in F_M$, we have $(x - \hat{x})'z = 0$ for all $z \in L_M$ i.e. $\hat{x} = (I - P_M)x$. And particularly $(x - \hat{x})'\hat{x} = 0$.

PROOF. In view of (1), for any $z \in L_M$, there exists $t_0 > 0$ such that $\hat{x} + tz \in F_M$ for $|t| < t_0$. (cf. for instance Stoer and Witzgall [12] page 39) From the definition of $\hat{x}$, the minimum of $f(t) = (x - \hat{x} - tz)'(x - \hat{x} - tz)$ in $t(|t| < t_0)$ is attained at $t = 0$. Therefore $f'(0) = -2(x - \hat{x})'z = 0$. Since $(x - z)'(x - z) = (\hat{x} - x)'(\hat{x} - x) + (\hat{x} - z)'(\hat{x} - z)$, we have $\hat{x} = (I - P_M)x$.

LEMMA 2. For any $x \in R^p$ and $z \in C_N$ we have $(x - \hat{x})'z \leq 0$.

PROOF. Suppose that $(x - \hat{x})'z_0 > 0$ for some $x \in R^p$ and $z_0 \in C_N$. It follows, from $\hat{x} \in C_N$ and the convexity of C_N , that $\hat{x} + t(z_0 - \hat{x}) \in C_N$ for $0 \leq t \leq 1$. Clearly the minimum of $f(t) = (x - \hat{x} - t(z_0 - \hat{x}))'(x - \hat{x} - t(z_0 - \hat{x}))$ is attained at $t_0 = ((z_0 - \hat{x})'(z_0 - \hat{x}))^{-1}(x - \hat{x})'z_0$. From the supposition we have $t_0 > 0$. This implies that $f(t)$ is strictly decreasing in a neighbourhood of $t = 0$, and hence contradicts the definition of $\hat{x}$.

LEMMA 3. If M satisfies the conditions

$$(7) \quad P_M x \in C_M^\perp \quad \text{and}$$

$$(8) \quad (I - P_M)x \in F_M$$

then $(I - P_M)x = \hat{x}$ i.e. $\hat{x} \in F_M$

PROOF.

$$\begin{aligned} (x - (I - P_M)x)'(x - (I - P_M)x) &= \min_{\mu' P_M x \leq 0} (x - \mu)'(x - \mu) \\ &\leq \min_{\mu \in C_N} (x - \mu)'(x - \mu) \text{ (by (7))} \\ &= (x - \hat{x})'(x - \hat{x}). \end{aligned}$$

Therefore, from (8) and the uniqueness of $\hat{x}$, we have $\hat{x} = (I - P_M)x$.

THEOREM 1. A necessary and sufficient condition for $\hat{x} \in F_M$ is that

$$(9) \quad (I - P_M)x \in C_{M^c}^o \quad \text{and}$$

$$(10) \quad P_M x \in C_M^\perp,$$

where $M^c = N - M$. And in this case $\hat{x} = (I - P_M)x$.

PROOF. Necessity: From Lemma 1 we have $\hat{x} = (I - P_M)x$. Therefore, from $(I - P_M)x \in F_M$ and the definition of F_M , it follows that

$$(11) \quad A'_M(I - P_M)x > 0 \quad \text{and}$$

$$(12) \quad A'_M(I - P_M)x = 0.$$

(11) implies (9). If (10) does not hold true, there exists z_0 such that

$$(13) \quad z_0' P_M x > 0 \quad \text{and}$$

$$(14) \quad A'_M z_0 \geq 0.$$

Let $z_t = (I - P_M)x + t(z_0 - (I - P_M)x)$ ($0 < t < 1$). (12) and (14) imply

$$(15) \quad A'_M z_t \geq 0.$$

For sufficiently small t , (11) implies

$$(16) \quad A'_{M^c} z_t > 0.$$

Therefore, (15) and (16) mean that $z_t \in C_N$ for sufficiently small t . But it follows, from (13), that

$$\begin{aligned} z_t' P_M x &= ((I - P_M)x + t(z_0 - (I - P_M)x))' P_M x \\ &= t z_0' P_M x > 0. \end{aligned}$$

This contradicts Lemma 2 and hence (10) should hold true. Sufficiency: The relation $A'_M(I - P_M)x = 0$ and (9) imply $(I - P_M)x \in F_M$. Also, (10) and $C_M^\perp \subset C_N^\perp$ imply $P_M x \in C_N^\perp$. Hence the conditions of Lemma 3 are satisfied and we have $\hat{x} \in F_M$.

The likelihood ratio test for our present problem is based on the two projections of x , $\hat{x}$ and $\tilde{x}$, defined in the beginning of this section, and rejects H when

$$(17) \quad \bar{\chi}^2 = (\hat{X} - \tilde{X})'(\hat{X} - \tilde{X})$$

is too large. In order to derive its distribution under H , we note that, when $X \sim N_p(\mu, I)$, $\mu \in L_{N_0}$, and $M \subset N \subset N_0$,

$$(18) \quad (I - P_M)\mu = (I - P_{N_0})\mu = \mu,$$

$$(19) \quad (I - P_M)(I - P_{N_0}) = I - P_{N_0},$$

$$(20) \quad (I - P_M)X \in C_{M^c}^0 \quad \text{if and only if} \quad (P_{N_0} - P_M)X \in C_{M^c}^0,$$

$$(21) \quad P_M X \sim N_p(0, P_M),$$

$$(22) \quad (I - P_{N_0})X \sim N_p(\mu, I - P_{N_0}),$$

$$(23) \quad (P_{N_0} - P_M)X \sim N_p(0, P_{N_0} - P_M),$$

$$(24) \quad P_M X \perp (P_{N_0} - P_M)X, \quad \text{and}$$

$$(25) \quad X'(P_{N_0} - P_M)X \sim \chi_{r(A_{N_0}) - r(A_M)}^2,$$

where $r(A)$ is the rank of A and $M^c = N - M$. (See Rao [8] page 187 for (25))

As for the distribution of $\bar{\chi}^2$ under H we have

THEOREM 2. *When H is true, then for $c \geq 0$ we have for all $\mu \in L_{N_0}$*

$$(26) \quad \Pr\{\bar{\chi}^2 \leq c\} = \sum_{M \subset N} \Pr\{P_M X \in C_M^\perp\} \cdot \Pr\{(I - P_M)X \in C_{M^c}^0\} \cdot \Pr\{\chi_{r(A_{N_0}) - r(A_M)}^2 \leq c\}.$$

where the summation is for all subsets of N .

PROOF.

$$\begin{aligned} & \Pr\{(\hat{X} - \tilde{X})'(\hat{X} - \tilde{X}) \leq c\} \\ &= \sum_{M \subset N} \Pr\{\hat{X} \in F_M, (\hat{X} - \tilde{X})'(\hat{X} - \tilde{X}) \leq c\} \quad (\text{by (2) and (3)}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{M \subset N} \Pr\{P_M X \in C_M^\perp, (I - P_M)X \in C_{M^c}^o, X'(P_{N_0} - P_M)X \leq c\} \\
&\quad \text{(by Theorem 1 and (19))} \\
&= \sum_{M \subset N} \Pr\{P_M X \in C_M^\perp\} \cdot \Pr\{(P_{N_0} - P_M)X \in C_{M^c}^o, X'(P_{N_0} - P_M)X \leq c\} \\
&\quad \text{(by (20) and (24))} \\
&= \sum_{M \subset N} \Pr\{P_M X \in C_M^\perp\} \cdot \Pr\{(P_{N_0} - P_M)X \in C_{M^c}^o\} \cdot \Pr\{X'(P_{N_0} - P_M)X \leq c\} \\
&= \sum_{M \subset N} \Pr\{P_M X \in C_M^\perp\} \cdot \Pr\{(I - P_M)X \in C_{M^c}^o\} \cdot \Pr\{\chi_{r(A_{N_0}) - r(A_M)}^2 \leq c\} \cdot \\
&\quad \text{(by (20) and (25))}
\end{aligned}$$

The fourth equation follows from the fact that, when $X \sim N_p(0, I)$, the distance and the direction from the origin are mutually independent.

It is clear from (21), (23) and (25), that this distribution does not depend on $\mu \in L_{N_0}$.

REMARK 1. Let O_M be a ball in $L_M^\perp$ with its center at the origin and its volume 1 with respect to the Lebesgue measure induced in $L_M^\perp$. Then $\Pr(P_M X \in C_M^\perp)$ is equal to the volume of $C_M^\perp \cap O_M$ in $L_M^\perp$. For $\Pr(P_M X \in C_M^\perp)$ is determined by the shape of $C_M^\perp \cap L_M^\perp$ in $L_M^\perp$ and hence invariant under the rotations of the cone $C_M^\perp \cap L_M^\perp$ in the linear space $L_M^\perp$. Similarly as for $\Pr((I - P_M)X \in C_{M^c}^o)$.

REMARK 2. If $F_M \neq \phi$, then we have $\Pr((I - P_M)X \in C_{M^c}^o) = \Pr((I - P_M)X \in C_{M^c})$. To see this, it suffices to show that the Lebesgue measure in L_M of $L_M \cap (C_{M^c} - C_{M^c}^o)$ is 0. Let $a \in L_M \cap C_{M^c}^o$ and l_a be any semisegment in L_M from a . Then, $l_a \cap (L_M \cap (C_{M^c} - C_{M^c}^o))$ has at most one point ([12] page 39). Therefore the Lebesgue measure in l_a of $l_a \cap (L_M \cap (C_{M^c} - C_{M^c}^o))$ is 0, i.e. the Lebesgue measure in L_M of $L_M \cap (C_{M^c} - C_{M^c}^o)$ is 0.

§ 4. General case

In this section, the condition on Σ is relaxed, but still assumed to be known. Those on μ under H and AH remain the same ones as in §3. Let the rank of the known matrix Σ be r ($\leq p$). When $r < p$, there does not exist the density function with respect to the Lebesgue measure on R^p , and its probability is concentrated on a r -dimensional hyperplane which is called the support set. Let $S_\mu = \mu + R(\Sigma)$, then S_μ is the support set of $N_p(\mu, \Sigma)$, namely $\Pr(X \in S_\mu) = 1$, and the explicit representation of the density function on S_μ is given by Khatri [3] (or see page

527 of Rao [8]). On the other hand, given an observation x of X , we can infer $\mu \in S_x = x + R(\Sigma)$.

The null hypothesis H determines a set $S_H = \{x: x \in S_\mu \text{ for some } \mu \text{ in } H\}$ and the alternative AH a set $S_{AH} = \{x: x \in S_\mu \text{ for some } \mu \text{ in } AH\}$. If $X \in S_H - S_{AH}$ or $S_{AH} - S_H$, there arises no testing problem, and we can "accept H " or "reject H " respectively without any danger of error. We call $T = S_H \cap S_{AH}$ a testable set, where we have to determine a critical region satisfying the size condition.

LEMMA 4.

- i) When $T \neq \phi$, we have $T = L_{N_0} + R(\Sigma) = S_H$.
- ii) When $L_{N_0}^\perp \subset R(\Sigma)$, we have $T = R^p$.
- iii) Generally, for $\mu \in C_N$, $Pr(X \in T) = 0$ or 1, and in the latter case, there exists a vector $\mu_0 \in L_{N_0}$ such that $Pr((I - P_\Sigma)(X - \mu_0) = 0) = 1$.

PROOF.

i): If $T \neq \phi$, clearly $T \subset L_{N_0} + R(\Sigma)$. Therefore we show $T \supset L_{N_0} + R(\Sigma)$. Since T has a point, say z_0 , we have $z_0 = \mu_0 + \sigma_0 = \mu_1 + \sigma_1$, where $\mu_0 \in L_{N_0}$, $\mu_1 \in C_N - L_{N_0}$, and $\sigma_1, \sigma_2 \in R(\Sigma)$. Let $z \in L_{N_0} + R(\Sigma)$ i.e. $z = \mu_2 + \sigma_2 = \mu_1 + (\mu_2 - \mu_0) + (\mu_0 - \mu_1) + \sigma_2$, where $\mu_2 \in L_{N_0}$ and $\sigma_2 \in R(\Sigma)$. $\mu_1 + (\mu_2 - \mu_0) \in \mu_1 + L_{N_0} \subset C_N - L_{N_0}$ and $(\mu_0 - \mu_1) + \sigma_2 = \sigma_1 - \sigma_0 + \sigma_2 \in R(\Sigma)$ imply $z \in T$.

ii): From $C_N - L_{N_0} \neq \phi$ and the supposition, we have $\mu \in L_{N_0} + R(\Sigma)$ for some $\mu \in C_N - L_{N_0}$. This implies $\mu \in T$ i.e. $T \neq \phi$. From (i), (ii) is proved.

iii): From the definition, it is obvious that $Pr(X \in T) = 0$ or 1. When $Pr\{X \in T\} = 1$, we have $T \neq \phi$ and $\mu = E(X) \in T = L_{N_0} + R(\Sigma)$ i.e. there exist $\mu_0 \in L_{N_0}$ and $\sigma_0 \in R(\Sigma)$ with $\mu = \mu_0 + \sigma_0$. It follows that $(I - P_\Sigma)X = (I - P_\Sigma)\mu = (I - P_\Sigma)\mu_0$.

Since Σ is a symmetric non-negative definite matrix with rank r , there exists an $r \times p$ matrix Q such that $Q\Sigma Q' = I$ and $QQ' = \text{Diag}(\lambda_1^{-1}, \dots, \lambda_r^{-1})$, where λ_i 's are non-zero eigenvalues of Σ . Let $Q^+ = Q' \text{Diag}\{\lambda_1, \dots, \lambda_r\}$ and $\Sigma^+ = Q'Q$, which are the Moore-Penrose inverses of Q and Σ respectively. After routine calculations we have $Q\Sigma Q^+ = \text{Diag}\{\lambda_1, \dots, \lambda_r\}$, $QQ^+ = I$, $Q^+(Q^+)' = \Sigma$, and $Q^+Q = P_\Sigma$. As in §2, we need to define the following; $L_{N_0:\Sigma} = \{z: A'_{N_0}Q^+z = 0\}$, $P_{N_0:\Sigma} = A_{N_0}(A'_{N_0}\Sigma A_{N_0})^- A'_{N_0}$, $C_{\phi:\Sigma} = R^p$, $P_{\phi:\Sigma} = 0$. For a nonempty subset $M \subset N$, set $C_{M:\Sigma} = \{z: A'_M Q^+ z \geq 0\}$ and $P_{M:\Sigma} = A_M(A'_M \Sigma A_M)^- A'_M$. For Σ , we define a semi-norm on R^p

$$\|x\|_\Sigma = \begin{cases} (x' \Sigma^+ x)^{1/2} & x \in R(\Sigma) \\ \infty & x \notin R(\Sigma). \end{cases}$$

LEMMA 5.

- i) $z \in L_{N_0}$ if and only if $Q^+z \in L_{N_0}$.
- ii) For any subset $M \subset N$,
 $z \in C_{M:\Sigma}$ if and only if $Q^+z \in C_M$, and
 $z \in C_{M:\Sigma}^0$ if and only if $Q^+z \in C_M^0$.
- iii) For any subset $M \subset N$,
 $z \in C_{M:\Sigma}^\perp$ if and only if $Q'z \in (R(\Sigma) \cap C_M)^\perp$.

PROOF.

- i): $z \in L_{N_0:\Sigma} \iff A'_{N_0}Q^+z = 0$
 $\iff Q^+z \in L_{N_0}$.
- ii): The proof is similar to that of (i), and hence omitted.
- iii): $z \in C_{M:\Sigma}^\perp \iff z'x \leq 0$ for any $x \in C_{M:\Sigma}$
 $\iff (Q'z)'(Q^+x) \leq 0$ for any x s.t. $A'_M Q^+x \geq 0$
 $\iff (Q'z)'x \leq 0$ for any $x \in R(Q^+) = R(\Sigma)$ s.t. $A'_M x \geq 0$
 $\iff (Q'z)'x \leq 0$ for any $x \in R(\Sigma) \cap C_M$
 $\iff Q'z \in (R(\Sigma) \cap C_M)^\perp$.

THEOREM 3. If the testing problem: $X \sim N_p(\mu, \Sigma)$ where Σ is known and $r(\Sigma) = r < p$, $H: \mu \in L_{N_0}$ against $AH: \mu \in C_N - L_{N_0}$ has a nonempty testable set T , the testing problem on T is equivalent to the one: $Q(X - \mu_0) \sim N_r(Q(\mu - \mu_0), I)$, $H': Q(\mu - \mu_0) \in L_{N_0:\Sigma}$ against $AH': Q(\mu - \mu_0) \in C_{N:\Sigma} - L_{N_0:\Sigma}$, where μ_0 is defined in Lemma 4.

PROOF.

$$\begin{aligned}
 \mu \in L_{N_0} &\iff \mu - \mu_0 \in L_{N_0} && (\text{by } \mu_0 \in L_{N_0}) \\
 &\iff Q^+Q(\mu - \mu_0) \in L_{N_0} && (\text{by } P_\Sigma = Q^+Q \text{ and } \mu - \mu_0 \in R(\Sigma)) \\
 &\iff Q(\mu - \mu_0) \in L_{N_0:\Sigma} && (\text{by Lemma 5}).
 \end{aligned}$$

Similarly $\mu \in C_N - L_{N_0} \iff Q(\mu - \mu_0) \in C_{N:\Sigma} - L_{N_0:\Sigma}$ can be proved. This completes the proof.

Theorem 3 and the functional invariance property of a MLE imply

COROLLARY. Given $x \in T$, the MLE's $\hat{x}$ and $\tilde{x}$ under AH and H exist uniquely and satisfy $\widehat{Q(X - \mu_0)} = Q(\hat{x} - \mu_0)$ and $\widetilde{Q(X - \mu_0)} = Q(\tilde{x} - \mu_0)$, where those in the left hand sides are MLE's calculated from $Q(X - \mu_0)$ under H' and AH' respectively.

The following characterization of $\hat{x}$ and $\tilde{x}$ is the generalization of Theorem 1.

THEOREM 4. For $x \in T$, a necessary and sufficient condition for $\hat{x} \in F_M$ is that

$$(30) \quad (I - \Sigma P_{M:\Sigma})x \in C_{M^c}^0 \quad \text{and}$$

$$(31) \quad P_{\Sigma} P_{M:\Sigma} x \in (R(\Sigma) \cap C_M)^\perp, \quad \text{where } M^c = N - M.$$

And in this case $\hat{x} = (I - \Sigma P_{M:\Sigma})x$. We also have $\tilde{x} = (I - \Sigma P_{N_0:\Sigma})x$.

REMARK. It is clear $\hat{x}$ and $\tilde{x}$ are the projections of x on C_N and L_{N_0} respectively with respect to the semi-norm defined by $\|\cdot\|_{\Sigma}$. Without danger of confusion, we use hereafter the terminology "projection with respect to Σ " in the above sense.

PROOF. We prove the first assertion. The proof of the second is parallel to the first and hence omitted.

$$\begin{aligned} \hat{x} \in F_M &\iff \hat{x} - \mu_0 \in F_M, \quad (\text{by } \mu_0 \in L_{N_0}) \\ &\iff Q^+ Q(\hat{x} - \mu_0) \in F_M, \quad (\text{by } P_{\Sigma} = Q^+ Q) \\ &\iff Q^+ \widehat{Q(x - \mu_0)} \in F_M, \\ &\iff \begin{cases} (I - (Q^+)' A_M (A_M' Q^+ (Q^+)' A_M)^- A_M' Q^+) Q(x - \mu_0) \in C_{M^c:\Sigma}^0 & \text{and} \\ (Q^+)' A_M (A_M' Q^+ (Q^+)' A_M)^- A_M' Q^+ Q(x - \mu_0) \in C_{M:\Sigma}^\perp \end{cases} \\ &\quad (\text{by Theorem 1 and Theorem 3}) \\ &\iff \begin{cases} (I - \Sigma A_M (A_M' \Sigma A_M)^- A_M') (x - \mu_0) \in C_{M^c}^0 & \text{and} \\ P_{\Sigma} A_M (A_M' \Sigma A_M)^- A_M' (x - \mu_0) \in (R(\Sigma) \cap C_M)^\perp, \end{cases} \\ &\quad (\text{by Lemma 5}) \\ &\iff \begin{cases} (I - \Sigma P_{M:\Sigma})x \in C_{M^c}^0 & \text{and} \\ P_{\Sigma} P_{M:\Sigma} x \in (R(\Sigma) \cap C_M)^\perp. \end{cases} \quad (\text{by } \mu_0 \in L_{N_0} \text{ i. e. } A_{N_0}' \mu_0 = 0) \end{aligned}$$

Similarly as (17) the likelihood ratio test statistic in our present case is

$$(32) \quad \tilde{\chi}^2 = \begin{cases} \|\hat{X} - \tilde{X}\|_{\Sigma}^2 & \text{if } X \in T \\ \infty & \text{if } X \notin T. \end{cases}$$

The proof of the following theorem is parallel to that of Theorem 2 by using Theorem 4 instead of Theorem 1, and hence omitted.

THEOREM 5. Assume $T \neq \phi$. For $c \geq 0$ we have, for all $\mu \in L_{N_0}$,

$$(33) \quad \Pr\{\bar{\chi}^2 \leq c\} = \sum_{M \in N} \Pr\{P_{\Sigma} P_{M:\Sigma} X \in (R(\Sigma) \cap C_M)^\perp\} \cdot \Pr\{(I - \Sigma P_{M:\Sigma})X \in C_{M^c}^0\} \cdot \Pr\{\chi_{r(P_{\Sigma} A_{N_0}) - r(P_{\Sigma} A_M)}^2 \leq c\}.$$

Let $\Pr\{A | X \sim F\}$ for event A and distribution F denote $\Pr\{X \in A\}$ when $X \sim F$. Then, the following theorem is of some interest.

THEOREM 6. If every vector $a_j \in N_0$ is an eigenvector of Σ whose corresponding eigenvalue is non-zero, (30) and (31) in Theorem 4 can be replaced by (9) and (10) in Theorem 1, and (33) in Theorem 5 by (26) in Theorem 2. And then,

$$(34) \quad \Pr\{P_M X \in C_M^\perp | X \sim N_p(0, \Sigma)\} = \Pr\{P_M X \in C_M^\perp | X \sim N_p(0, I)\},$$

$$(35) \quad \Pr\{(I - P_M)X \in C_{M^c}^0 | X \sim N_p(0, \Sigma)\} = \Pr\{(I - P_M)X \in C_{M^c}^0 | X \sim N_p(0, I)\}.$$

PROOF. To show the first assertion, it suffices to prove (i) $T = R^p$, (ii) $\Sigma P_{M:\Sigma} = P_M$ and $P_{\Sigma} A_M = A_M$, and (iii) $P_{\Sigma} P_{M:\Sigma} x \in (R(\Sigma) \cap C_M)^\perp \Leftrightarrow P_M x \in C_M^\perp$.

i) Since $L_{N_0}^\perp = R(A_{N_0}) \subset R(\Sigma)$, it follows, from (ii) of Lemma 4, that R^p is the testable set.

ii) From the supposition, we have $\Sigma A_M = A_M I_\lambda$ where I_λ is a diagonal matrix with the eigenvalues corresponding to the column vectors of A_M . $I_\lambda^{-1}(A'_M A_M)^-$ is a generalized inverse of $A'_M \Sigma A_M$. Hence we have

$$\Sigma P_{M:\Sigma} = \Sigma A_M (A'_M \Sigma A_M)^- A'_M = A_M I_\lambda I_\lambda^{-1} (A'_M A_M)^- A'_M = P_M,$$

$$P_{\Sigma} A_M = P_{\Sigma} \Sigma A_M I_\lambda^{-1} = \Sigma A_M I_\lambda^{-1} = A_M.$$

iii) First we show $R(\Sigma) \cap (R(\Sigma) \cap C_M)^\perp = R(\Sigma) \cap C_M^\perp$, i.e., if $x \in R(\Sigma)$, then $x \in (R(\Sigma) \cap C_M)^\perp \Leftrightarrow x \in C_M$. Clearly it suffices to show $R(\Sigma) \cap (R(\Sigma) \cap C_M)^\perp \subset R(\Sigma) \cap C_M^\perp$. Let $y \in R(\Sigma) \cap (R(\Sigma) \cap C_M)^\perp$. For any $z = z_0 + z_1 \in C_M$ s.t. $z_0 \in R(\Sigma)$ and $z_1 \in R(\Sigma)^\perp$, it follows, from the supposition, that $A'_M z_0 = A'_M z \geq 0$, i.e., $z_0 \in C_M \cap R(\Sigma)$. Therefore we have $z'y = z'_0 y \leq 0$, i.e., $y \in R(\Sigma) \cap C_M^\perp$. Also, from $\Sigma = Q^+(Q^+)'$ and $QQ^+ = I$, we have $A'_M Q^+ = I_\lambda A'_M Q'$. These two facts lead us to the followings:

$$\begin{aligned}
P_{\Sigma}P_{M:\Sigma}X \in (R(\Sigma) \cap C_M)^\perp &\iff z'(Q^+)P_{M:\Sigma}X \leq 0 \quad \text{for any } z \text{ s.t. } A'_M Q^+ z \geq 0, \\
&\quad \text{(by Lemma 5)} \\
&\iff z'(Q^+)P_{M:\Sigma}X \leq 0 \quad \text{for any } z \text{ s.t. } I_\lambda A'_M Q^+ z \geq 0. \\
&\iff y'Q^+(Q^+)P_{M:\Sigma}X \leq 0 \quad \text{for any } y \in R(\Sigma) \cap C_M. \\
&\iff \Sigma P_{M:\Sigma}X \in (R(\Sigma) \cap C_M)^\perp \\
&\iff P_M X \in C_M^\perp.
\end{aligned}$$

To show (34) and (35), let λ_i , $i=0, 1, \dots, m$, be the mutually different eigenvalues of Σ and $\lambda_0=0$. Q_i denotes the orthogonal projection to the eigenspace corresponding to λ_i . Then, from the supposition, $Q_0 P_M = P_M Q_0 = 0$ and $Q_i P_M = P_M Q_i$ for any $M \subset N$ and $1 \leq i \leq m$. Let $Y_0 \sim N_p(0, Q_0)$ and independent to X . When we define $Y = \sum_{i=1}^n \lambda_i^{-\frac{1}{2}} Q_i X + Y_0$, we have $Y \sim N_p(0, I)$. Appropos of (34),

$$\begin{aligned}
P_M X \in C_M^\perp &\iff P_M \sum_{i=0}^n Q_i X \in C_M^\perp \\
&\iff \sum_{i=1}^n Q_i P_M X = A_M t \quad \text{for some } t \leq 0. \\
&\quad \text{(Farkas' Lemma ([9], page 20))} \\
&\iff Q_i P_M X = A_M t_i \quad \text{for some } t_i \leq 0, 1 \leq i \leq m, t = \sum_{i=1}^n t_i \\
&\quad \text{(by the uniqueness of decomposition)} \\
&\iff P_M Y = A_M \left(\sum_{i=1}^n \lambda_i^{-\frac{1}{2}} t_i \right) \\
&\iff P_M Y \in C_M^\perp.
\end{aligned}$$

Similary for (35) we have

$$(I - P_M)X \in C_M^o \iff (I - P_M)Y \in C_M^o.$$

Therefore, (34) and (35) hold true.

§ 5. The monotonicity of the power function

In this section, we will prove a kind of monotonicity of power of our test. For any points $\mu_0 \in L_{N_0}$ and $\mu_1 \in C_N - L_{N_0}$, we will prove that the power, i.e.,

the probability of rejecting H_0 , is monotone increasing on the segment starting from μ_0 and passing through μ_1 . Let $\mu(t) = \mu_0 + (\mu_1 - \mu_0)t$, $t \in [0, \infty)$. It is clear that $\mu(0) \in L_{N_0}$ and $\mu(t) \in C_N - L_{N_0}$ for all $t > 0$.

THEOREM 7. *$Pr\{H_0 \text{ is rejected} | X \sim N_p(\mu(t), \Sigma)\}$ is strictly increasing in $t \in [0, \infty)$ if it does not reach 1.*

PROOF. When $T = \phi$, the power is 0 when $t = 0$ and 1 when $t > 0$. The problem is when $T \neq \phi$, and in this case the acceptance region is given by

$$K_\alpha = \{x: \|\hat{x} - \tilde{x}\|_2^2 \leq t_\alpha, x \in T\},$$

where t_α is determined by the relation $Pr(X \in K_\alpha | X \sim N_p(\mu, \Sigma)) = 1 - \alpha$ for all $\mu \in L_{N_0}$. As T is a linear subspace, $\mu_1 = \mu(1) \in T$ or $\notin T$ implies $\mu(t) \in T$ or $\notin T$ for all $t > 0$, and in the latter case the power is α for $t = 0$ and 1 for $t > 0$. In the former case, due to Theorem 3, the problem can be transformed to a testing problem with $\Sigma = I$, where $K_\alpha = \{x: \|\hat{x} - \tilde{x}\|^2 \leq t_\alpha\}$. Let the power function be

$$\varphi(\mu) = Pr\{X \notin K_\alpha | X \sim N_p(\mu, I)\}.$$

Since $K_\alpha + \mu = K_\alpha$ for $\mu \in L_{N_0}$, it is easy to check the following

$$(36) \quad \varphi(\mu) = \varphi(P_{N_0}\mu), \quad \text{and}$$

$$(37) \quad \text{the acceptance region for the testing problem; } H: \mu = 0 \text{ and } AH: \mu \in L_{N_0} \cap C_N - \{0\} \text{ is the same as } K_\alpha.$$

Therefore, since we may set $\mu_0 = 0$ and $\mu_1 \in L_{N_0} \cap C_N$ from (36), it suffices, from (37), to treat this problem when $\Sigma = I$ and $L_{N_0} = \{0\}$, then $K_\alpha = \{x: |x|^2 \leq t_\alpha\}$. To prove this, we need the following elementary lemmas. Let $l_\mu = \{t\mu: t \in R\}$ and $C'_N = \{x \in C_N: |x| = 1\}$.

LEMMA 6. *$x \in K_\alpha$ if and only if $\mu'x \leq t_\alpha^{1/2}$ for any $\mu \in C'_N$.*

PROOF.

Sufficiency: When $\hat{x} = 0$, $x \in K_\alpha$ is obvious. When $\hat{x} \neq 0$, $|x|^{-1}x \in C'_N$ as $\hat{x} \in C_N$. Since $\hat{x}$ is, from Lemma 1, an orthogonal projection to some linear subspace, we have $|\hat{x}|^{-1}\hat{x}'\hat{x} = |\hat{x}|^{-1}\hat{x}'x \leq t_\alpha^{1/2}$. Hence $x \in K_\alpha$. Necessity: Suppose $\mu'x > t_\alpha^{1/2}$ for some $\mu \in C'_N$. $(\mu'x)\mu \in C_N$ is the orthogonal projection of x to l_μ . Therefore, from the definition of $\hat{x}$, we have

$$\hat{x}'\hat{x} \geq \{(\mu'x)\mu\}'\{(\mu'x)\mu\} = (\mu'x)^2 > t_\alpha.$$

This is a contradiction. Hence $x \notin K_\alpha$.

LEMMA 7. Suppose $\mu \in C'_N$. If $x \in K_\alpha$ and $\mu'x \geq 0$ then $(I - 2\mu\mu')x \in K_\alpha$.

PROOF. It suffices, from Lemma 6, to show for any $\theta \in C'_N$

$$\theta'(I - 2\mu\mu')x = \theta'x - 2(\theta'\mu)(\mu'x) \leq t_\alpha^{1/2}.$$

When $\theta'\mu \geq 0$, it is obvious, since $\theta'x \leq t_\alpha^{1/2}$ and $\mu'x \geq 0$. On the other hand, since C'_N is a convex cone and $(I - 2\mu\mu')\theta = \theta + 2(-\theta'\mu)\mu$ we have $(I - 2\mu\mu')\theta \in C'_N$ for $\theta'\mu < 0$. Therefore, from the definition of x , the conclusion follows obviously.

LEMMA 8. Let $\mu \in C'_N$ and $d \in l_\mu^\perp$. If $(d + l_\mu) \cap K_\alpha \neq \emptyset$ then either of the following holds

- i) $(d + l_\mu) \cap K_\alpha = \{e - t\mu : t \geq 0\}$ for some e , or
- ii) $(d + l_\mu) \cap K_\alpha = \{(1 - t)e + tf : 0 \leq t \leq 1\}$ for some e and f . In this case there exists $h \leq 0$ s.t. $\frac{1}{2}(e + f) = d + h\mu$.

PROOF. The intersection of the straight line $d + l_\mu$ and the closed convex set K_α should be a line connecting two points, a segment starting from a point, or $d + l_\mu$. Therefore it remains to show that the intersection cannot be of the form $\{e + t\mu : t \geq 0\}$ or $d + l_\mu$. In fact if $K_\alpha \ni d + t_\mu$ for any large positive t then $\mu'(d + t\mu) = t \leq t_\alpha^{1/2}$ by Lemma 6. This is a contradiction.

Further assume $h > 0$ in case of (ii). Without loss of generality we may set $\mu'e \geq 0$. From $\frac{1}{2}((I - 2\mu\mu')e + e) = d$ and $\frac{1}{2}(f + e) = d + h\mu$ we have $f - (I - 2\mu\mu')e = 2h\mu$. It follows, from the definition of f , that $(I - 2\mu\mu')e \notin K_\alpha$. This contradicts Lemma 7, and hence $h \leq 0$.

Using the above lemmas, we show that $Pr\{X \in K_\alpha | X \sim N_p(t\mu, I)\}$, $\mu \in C_N - \{0\}$, is strictly decreasing in $t \geq 0$, i.e., the power is strictly increasing. Without loss of generality we may set $\mu \in C'_N$. Let $d \in l_\mu^\perp$ and $X \sim N_p(\mu t, I)$. The conditional distribution of X given $X \in d + l_\mu$ is a normal distribution with the mean vector $d + \mu t$, which, by the obvious correspondence between $d + l_\mu$ and the one dimensional space, corresponds to a one dimensional normal distribution with the mean t . By lemma 8, $(d + l_\mu) \cap K_\alpha \neq \emptyset$ corresponds to either $(-\infty, a)$ or (b, a) with $b + a \leq 0$. This correspondence and the property of a normal distribution enable us to see that the conditional probability of $(d + l_\mu) \cap K_\alpha$ is strictly decreasing in $t \geq 0$. Therefore, its integration with respect to the marginal distribution of d on $l_\mu^\perp$, implies that $Pr\{X \in K_\alpha | X \sim N_p(t\mu, I)\}$ is strictly decreasing in $t \geq 0$.

For a application of this Theorem, see Shiraishi [11].

§ 6. Some examples

In case $\Sigma = I$, the MLE of x is nothing but the orthogonal projection of x to the convex cone. For the case of general Σ , the MLE is the solution of the minimization problem involving the seminorm $\|\cdot\|_{\Sigma}$. Recall that it is called the projection with respect to Σ .

The MLE of the mean of normal distribution is a projection with respect to the variance matrix. There are situations where the MLE of a parameter is a projection with respect to some norm, or some matrix A , while the underlying distribution is not normal. In such a case Theorem 1 and 4 are applicable. Moreover, if the asymptotic distribution is normal and its variance matrix is A , then Theorem 5 is applicable, too.

Let $N_0 = \{(i, j): 1 \leq i, j \leq p, i \neq j\}$ and $a_{(i,j)}$ be a p -vector with 1 in the i -th component, -1 in the j -th component, and 0 in others. And let N be a subset of N_0 such that $C_N \neq L_{N_0}$. Thus H is reduced to the equalities of $\mu'_i s$, and AH is reduced to some order restrictions among $\mu'_i s$.

EXAMPLE 1. Consider a exponential type distribution with the density $f(x: \theta) = C(\theta) \exp [q(\theta)K(x)]$ with respect to a measure ν , where $q(\theta)$ has a continuous derivative with $q'(\theta) > 0$ and $C(\theta)$ has a derivative $C'(\theta) = -\theta q'(\theta)C(\theta)$. The mean and the variance of $K(x)$ are derived by twice differentiating $\int f(x: \theta) d\nu(x) = 1$, and they are θ and $q'(\theta)^{-1}$ respectively.

Let $(X_{i1}, \dots, X_{in_i})$ $i = 1, \dots, p$ be the independent random samples from this distribution with $\theta = \mu_i$, $i = 1, \dots, p$ respectively. Let $X = (n_1^{-1} \sum_{j=1}^{n_1} K(X_{1j}), n_2^{-1} \sum_{j=1}^{n_2} K(X_{2j}), \dots, n_p^{-1} \sum_{j=1}^{n_p} K(X_{pj}))'$, then $E(X) = (\mu_1, \mu_2, \dots, \mu_p)'$ and $V(X) = \text{Diag} [n_1^{-1} q'(\mu_1)^{-1}, n_2^{-1} q'(\mu_2)^{-1}, \dots, n_p^{-1} q'(\mu_p)^{-1}]$. The MLE's $\hat{x}$ and $\tilde{x}$ are discussed by Robertson and Wegman [9], which turn out to be the projection of X on L_{N_0} and C_N with respect to the matrix $\text{Diag} [n_1^{-1}, n_2^{-1}, \dots, n_p^{-1}]$. (Also see Nomakuchi [6]) Therefore theorem 4 is applicable.

Let λ be the likelihood ratio, and let $\Sigma = q'(\mu_1)^{-1} \text{Diag} [c_1^{-1}, c_2^{-1}, \dots, c_p^{-1}]$. When $\mu = (\mu_1, \mu_1, \dots, \mu_1)$, we have $-2 \log \lambda = n(\hat{x} - \tilde{x})' \Sigma^{-1} (\hat{x} - \tilde{x}) + o(n)$ as $n \rightarrow \infty$ under the condition $n = nc_i + o(n)$ with $\Sigma c_i = 1$ and $c_i > 0$ ($i = 1, \dots, p$), and the asymptotic distribution of $\sqrt{n}(x - \mu)$ is $N_p(0, \Sigma)$. Therefore Theorem 5 can be applied to obtain the upper tail probabilities.

EXAMPLE 2. Consider a multinomial distribution based on n trials with

parameters $\mu_1, \dots, \mu_p$ ($\sum \mu_i = 1$), and let $nX = (X_1, X_2, \dots, X_p)'$, where X_i is the number of occurrence of the i -th event. Similarly as in Example 1, $\tilde{X}$ and $\hat{X}$ can be seen to be the projection of X on L_{N_0} and C_N with respect to I . (See [2], [6], [10]) Under $H: \mu = (p^{-1}, \dots, p^{-1})'$, the asymptotic relation is $-2 \log \lambda = np(\hat{X} - \tilde{X})'(\hat{X} - \tilde{X}) + o(n)$, and the asymptotic distribution of $\sqrt{np}(x - \mu)$ is $N_p(0, I - p^{-1}E)$ where E is a $p \times p$ matrix whose elements are all 1, and all of $a(i, j)$ are easily seen to be eigen vectors corresponding to a non zero eigen value 1 of $I - p^{-1}E$, and hence we can apply Theorem 6. The above two conclusions are important. In testing homogeneity of multinomial parameters $(\mu_1, \dots, \mu_p)$ against any order alternative, we can treat the data $X = n^{-1}(X_1, X_2, \dots, X_p)'$ as if $X \sim N_p(\mu, (np)^{-1}I)$ we can compute the statistics $\hat{X}$ and $\tilde{X}$ in this way as they are projections with respect to I . The asymptotic distribution of $-2 \log \lambda$ is identical to the one given by (26) of Theorem 2 for testing $H: \mu = (p^{-1}, \dots, p^{-1})'$ against an ordered alternative. The problem of order restricted inference for the multinomial parameters was first treated by Chacko [2] and also Robertson [10].

References

- [1] BARLOW, R. E., BARTHOROMEW, D. J., BREMNER, J. M. AND BRUNK, H. D.; *Statistical Inference under Order Restrictions*, New York, John Wiley and Sons Inc, 1972.
- [2] CHACKO, V. J.; *Modified chi-square test for ordered alternatives*, Sankhyā, **A23** 25–40.
- [3] KHATRI, C. G.; Some results for the singular normal multivariate regression models, Sankhyā, **A30** (1961) 267–280.
- [4] KUDÔ, A.; *A multivariate analogue of the one sided test*, Biometrika, **50** (1961) 403–418.
- [5] KUDÔ, A. AND CHOI, J. R.; *A generalized multivariate analogue of the one sided test*, Mem. Fac. Sci. Kyushu Univ. **29** (1975) 303–328.
- [6] NOMAKUCHI, K.; *Projection on a convex polyhedral cone and the restricted maximal likelihood estimate*, to appear.
- [7] NUESCH, P. E.; *On the problem of testing location in multivariate populations for restricted alternatives*, Ann. Math. Statist., **37** (1966) 113–119.
- [8] RAO, C. R.; *Linear Statistical Inference and its Applications* (second edition), New York, John Wiley and Sons Inc, 1973.
- [9] ROBERTSON, T. and WEGMAN, E. J.; *Likelihood ratio tests for order restrictions*, Ann. Statist., **6** (1978) 485–505.
- [10] ROBERTSON, T.; *Testing for and against on order restriction on multinomial parameters*, J.A.S.A., **73** (1978) 197–202.
- [11] SHIRAIISHI, T.; *Testing homogeneity against trend based on rank in one-way layout*, Commun. Statist. Theor. Meth., **11** (1982) 1255–1268.
- [12] STOER, J. and WITZGALL, C.; *Convexity and Optimization in Finite Dimensions I*, Berlin, Springer-Verlag, 1970.
- [13] WYNN, H. P.; *Integrals for one-sided confidence bounds: a general result*, Biometrika, **62** (1975) 393–396.