Statistica Sinica 3(1993), 103-115

ON THE LOCAL OPTIMALITY OF OPTIMAL LINEAR TESTS FOR RESTRICTED ALTERNATIVES

Ming-Tan M. Tsai and Pranab Kumar Sen

Academia Sinica and University of North Carolina at Chapel Hill

Abstract: For the multivariate normal mean vector testing problem, it is shown that in the light of local power, the most stringent somewhere most powerful test (MSSMPT) performs better than the likelihood ratio test (LRT) for the entire positive orthant space.

Key words and phrases: Likelihood ratio test, most stringent and somewhere most powerful test, positive orthant alternative, restricted alternatives, uniformly most powerful test.

1. Introduction

Let $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ik})'$, $i = 1, \ldots, n$, be n independent and identically distributed random (k-) vectors (i.i.d.r.v.) having a k-variate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ and non-negative definite (n.n.d.) dispersion matrix $\boldsymbol{\Delta}$. Consider the problem of testing the null hypothesis $H_0 : \boldsymbol{\mu} = \mathbf{0}$ against the restricted alternative $H_1 : \boldsymbol{\mu} \geq \mathbf{0}$, $\|\boldsymbol{\mu}\|^2 > 0$ (which constitutes the positive orthant space R^{+k}). A popular test for this restricted alternative problem is based on the usual likelihood ratio test (LRT) criterion (see Bartholomew (1959 a,b), Kudô (1963), Nüesch (1966), Perlman (1969) and extensive literature cited in Robertson et al. (1988)). However, the (asymptotic) optimality properties of the LRT are not that well known (compared to the case of the global alternative $\boldsymbol{\mu} \neq \mathbf{0}$ where the LRT has the asymptotic most stringent and best average power properties). Because of this, Abelson and Tukey (1963) considered an optimal linear test (OLT) statistic which was later extended by Schaafsma and Smid (1966), Snijders (1979), Shi (1987) and Shi and Kudô (1987), among others; these are refered to as the most stringent and somewhere most powerful (MSSMP) tests. The advantage of an OLT is that the test statistic has a normal distribution so that the critical level can be computed very easily, and the power can be expressed in terms of a normal distribution. Also, such an OLT is most powerful for alternatives in a certain direction. On the other hand, for such restricted alternatives, the LRT may not be a Bayes test, and hence, it is not generally a most stringent test (even asymptotically), and may not be most powerful even for a specific part of the restricted parameter space R^{+k}. However, from the consistency point of view, Perlman (1969) has shown that the region of consistency is larger for the LRT than the corresponding OLT. An analytic comparison of the (exact or asymptotic) power functions for the LRT and OLT is difficult (due to the fact that their null as well as non-null distributions are not of comparable forms). For the specific case, $\boldsymbol{\Delta} = \boldsymbol{I}$, Oosterhoff (1969) has shown that the LRT is asymptotically optimal in the sense that the maximum shortcoming of the LRT converges uniformly to 0 (at an exponential rate) when the level of significance converges to 0; this result for a general $\boldsymbol{\Delta}$ is due to Kallenberg (1978). Oosterhoff (1969) also claimed that the OLT inherits the same asymptotic optimality property of the LRT when the level of significance converges to 0. In the current study, we also consider this situation when the level of significance is small, and in the light of some local power considerations, a comparative picture of the OLT and LRT is drawn. In Section 2, the case of known covariance matrix is considered, where it is shown that in terms of the slopes of the power functions at the origin, the OLT has a better picture than the LRT (uniformly in the entire parameter space R^{+k}). In Section 3, we incorporate the curvature of the power function in the case of known $\boldsymbol{\Delta}$. In the last section, the case when $\boldsymbol{\Delta} = \sigma^2\boldsymbol{\Delta}_0$, where $\boldsymbol{\Delta}_0$ is known and σ^2 is unknown, is considered. For this model, parallel results are studied.

2. First Order Local Picture

Without loss of generality, we let $n = 1$. We also write $\boldsymbol{\Delta} = ((\delta^{jj'}))$ and let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_k)'$, where $\delta_j = \sqrt{\delta^{jj}}$, $j = 1, \ldots, k$. Then, following the arguments of Shi (1987), Shi and Kudô (1987), we claim that there exists a particular subset J $(\emptyset \subseteq J \subseteq K)$ for which $\boldsymbol{\Delta}_{JJ:J'}\boldsymbol{\delta}_J > \boldsymbol{0}$ and $\boldsymbol{\delta}_{J'} + \boldsymbol{\Delta}^{-1}_{J'J'}\boldsymbol{\Delta}_{J'J}\boldsymbol{\delta}_J \leq \boldsymbol{0}$ with $\boldsymbol{B}^{-1}$ denoting the generalized inverse of $\boldsymbol{B}$, so that the OLT statistic is of the form

$$T = \boldsymbol{\ell}'_{J:J'}\boldsymbol{X} \tag{2.1}$$

where

$$\boldsymbol{\ell}_{J:J'} = \boldsymbol{\delta}'_J\boldsymbol{B}_J\{\boldsymbol{\delta}'_J\boldsymbol{\Delta}_{JJ:J'}\boldsymbol{\delta}_J\}^{-1/2} \text{ with } \boldsymbol{B}_J = \boldsymbol{I} - \boldsymbol{\Delta}_{JJ'}\boldsymbol{\Delta}^{-1}_{J'J'}. \tag{2.2}$$

Let a be any subset of $K = \{1, \ldots, k\}$ and a' be its complementary subset $(\emptyset \subseteq a \subseteq K)$. For each a in K, partition (following possible rearrangement) $\boldsymbol{X}$ and $\boldsymbol{\Delta}$ as

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_a \\ \boldsymbol{X}_{a'} \end{pmatrix} \qquad \boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{aa} & \boldsymbol{\Delta}_{aa'} \\ \boldsymbol{\Delta}_{a'a} & \boldsymbol{\Delta}_{a'a'} \end{pmatrix}. \tag{2.3}$$

Also for each a $(\emptyset \subseteq a \subseteq K)$, let

$$\boldsymbol{X}_{a:a'} = \boldsymbol{X}_a - \boldsymbol{\Delta}_{aa'}\boldsymbol{\Delta}^{-1}_{a'a'}\boldsymbol{X}_{a'}; \tag{2.4}$$

$$\boldsymbol{\Delta}_{aa:a'} = \boldsymbol{\Delta}_{aa} - \boldsymbol{\Delta}_{aa'}\boldsymbol{\Delta}^{-1}_{a'a'}\boldsymbol{\Delta}_{a'a}. \tag{2.5}$$

Then, the LRT statistic is of the form

$$Q^2 = \sum_{\emptyset \subseteq a \subseteq K} \{\boldsymbol{X}'_{a:a'} \boldsymbol{\Delta}^{-1}_{aa:a'} \boldsymbol{X}_{a:a'}\} 1(\boldsymbol{X}_{a:a'} > \boldsymbol{0}, \boldsymbol{\Delta}^{-1}_{a'a'} \boldsymbol{X}_{a'} \leq \boldsymbol{0}), \tag{2.6}$$

where $1(B)$ stands for the indicator function of the set B.

Note that the OLT statistic has a normal distribution function (d.f.) while the LRT statistic has a chi-bar d.f. (under H_0), and under H_1, their non-null distributions are not of comparable forms to allow the usual Pitman R.E. (relative efficiency) measure to study their relative performances. For this reason, we use some local measures to compare the performances of the OLT and LRT.

Theorem 1. *Let $\boldsymbol{X}_i \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Delta})$, $i = 1, \ldots, n$, where $\boldsymbol{\Delta}$ is known and non-negative definite. For testing the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ against $H_1 : \boldsymbol{\mu} \geq \boldsymbol{0}$, $\|\boldsymbol{\mu}\| > 0$, the* OLT *is uniformly locally more powerful than the* LRT *as the level of significance α is made to converge to* 0.

Proof. Let $\beta_\alpha(Q^2; \boldsymbol{\mu})$ ($\beta_\alpha(T; \boldsymbol{\mu})$) and y^2_α (τ_α) be the corresponding power function and critical point of the LRT (OLT) when the level of significance is α and $\boldsymbol{\mu} \in R^{+k}$. Thus if we write $\boldsymbol{\mu} = \mu \boldsymbol{t}$, $\boldsymbol{t} = (t_1, \ldots, t_k)'$, $\boldsymbol{t} \geq \boldsymbol{0}$, $\boldsymbol{t}'\boldsymbol{\Delta}^{-1}\boldsymbol{t} = 1$, then, following some standard steps, it follows that

$$\beta'_\alpha(T; \boldsymbol{t}) = (\partial/\partial\mu)\beta_\alpha(T; \mu\boldsymbol{t})|_{\mu=0} = (\boldsymbol{t}'\boldsymbol{\ell}_{J:J'})\phi(\tau_\alpha); \tag{2.7}$$

where $\phi(\cdot)$ stands for the standard normal density function. Similarly,

$$\begin{aligned} &\beta'_\alpha(Q^2; \boldsymbol{t}) \\ = \; &(\partial/\partial\mu)\beta_\alpha(Q^2; \mu\boldsymbol{t})\Big|_{\mu=0} = \sum_{\{\emptyset \subseteq a \subseteq K\}} E_0\{\boldsymbol{X}'\boldsymbol{\Delta}^{-1}\boldsymbol{t} 1(S_a)\} \\ = \; &(\boldsymbol{\ell}'_{J:J'}\boldsymbol{t}) \sum_{\{\emptyset \subseteq a \subseteq K\}} E_0\{\boldsymbol{X}'\boldsymbol{\Delta}^{-1}(\boldsymbol{\ell}_{J:J'}\boldsymbol{\ell}'_{J:J'})^{-1}\boldsymbol{\ell}_{J:J'} 1(S_a)\} \end{aligned} \tag{2.8}$$

where E_0 denotes the expectation under the null hypothesis and

$$S_a = \{\boldsymbol{X} : \boldsymbol{X}'_{a:a'}\boldsymbol{\Delta}^{-1}_{aa:a'}\boldsymbol{X}_{a:a'} \geq y^2_\alpha; \boldsymbol{X}_{a:a'} > \boldsymbol{0}, \boldsymbol{\Delta}^{-1}_{a'a'}\boldsymbol{X}_{a'} \leq \boldsymbol{0}\}. \tag{2.9}$$

Note that both (2.7) and (2.8) depend on the direction vector $\boldsymbol{t}$ through the multiplicative factor $\boldsymbol{t}'\boldsymbol{\ell}_{J:J'}$, so that if we let

$$\delta_\alpha(\boldsymbol{t}) = \beta'_\alpha(Q^2; \boldsymbol{t}) \Big/ \beta'_\alpha(T; \boldsymbol{t}), \tag{2.10}$$

$$\delta_\alpha = \sum_{\{\emptyset \subseteq a \subseteq K\}} E_0\{\boldsymbol{X}'\boldsymbol{\Delta}^{-1}(\boldsymbol{\ell}_{J:J'}\boldsymbol{\ell}'_{J:J'})^{-1}\boldsymbol{\ell}_{J:J'} 1(S_a)\} \Big/ \phi(\tau_\alpha) \tag{2.11}$$

then we have $\delta_\alpha(\boldsymbol{t}) = \delta_\alpha$, uniformly in $\boldsymbol{t} : \boldsymbol{t} \geq \boldsymbol{0}$, $\boldsymbol{t}'\boldsymbol{\Delta}^{-1}\boldsymbol{t} = 1$.

Let

$$\boldsymbol{m}_{J:J'} = \boldsymbol{\Delta}^{-1/2}(\boldsymbol{\ell}_{J:J'}\boldsymbol{\ell}'_{J:J'})^{-1}\boldsymbol{\ell}_{J:J'}, \text{ and } \boldsymbol{Y} = \boldsymbol{\Delta}^{-1/2}\boldsymbol{X}. \tag{2.12}$$

Then (2.10) reduces to

$$\delta_\alpha = \Big[1/\phi(\tau_\alpha)\Big]\left[\sum_{\{\emptyset\subseteq a\subseteq K\}} E_0\{\boldsymbol{Y}'\boldsymbol{m}_{J:J'}1(S_a)\}\right]. \tag{2.13}$$

Consider the usual polar transformation from $\boldsymbol{Y}$ to $(W, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{k-1})'$,

$$Y_j = W\left(\sin\theta_j \prod_{r=1}^{j-1}\cos\theta_r\right),\ j = 1, \ldots, k-1,\ Y_k = W(\cos\theta_1 \cdots \cos\theta_{k-1}), \tag{2.14}$$

$-\pi/2 \leq \theta_j \leq \pi/2$, for $j = 1, \ldots, k-2$ and $0 \leq \theta_{k-1} \leq 2\pi$; the Jacobian of this transformation is given by

$$J(\boldsymbol{Y}/W, \boldsymbol{\theta}) = J(W\boldsymbol{\tau}/W, \boldsymbol{\theta}) = W^{k-1}\prod_{r=1}^{k-2}(\cos\theta_r)^{k-r-1}, \tag{2.15}$$

where

$$\boldsymbol{\tau} = \left(\sin\theta_1, \cos\theta_1\sin\theta_2, \ldots, \prod_{r=1}^{k-1}\cos\theta_r\right)'. \tag{2.16}$$

Note that $\|\boldsymbol{\tau}\|^2 = 1$, and

$$\begin{aligned}
&\|\boldsymbol{m}_{J:J'}\|^2 \\
&= (\boldsymbol{\delta}'_J\boldsymbol{\Delta}_{JJ:J'}\boldsymbol{\delta}_J)\boldsymbol{\delta}'_J\boldsymbol{B}_J\Big[(\boldsymbol{B}'_J\boldsymbol{\delta}_J\boldsymbol{\delta}'_J\boldsymbol{B}_J)^{-1}\Big]'\boldsymbol{\Delta}^{-1}(\boldsymbol{B}'_J\boldsymbol{\delta}\boldsymbol{\delta}'_J\boldsymbol{B}_J)^{-1}\boldsymbol{B}'_J\boldsymbol{\delta}_J \\
&= \mathrm{tr}\Big[(\boldsymbol{\Delta}_{JJ:J'}\boldsymbol{\delta}_J\boldsymbol{\delta}'_J)^{-1}\boldsymbol{\Delta}_{JJ:J'}\boldsymbol{\delta}_J\boldsymbol{\delta}'_J\Big] \\
&= \mathrm{rank}(\boldsymbol{\Delta}_{JJ:J'}\boldsymbol{\delta}_J\boldsymbol{\delta}'_J) = \mathrm{rank}(\boldsymbol{\delta}_J\boldsymbol{\delta}'_J) = 1,
\end{aligned}$$

where $\mathrm{tr}(\boldsymbol{B})$ denotes the trace of $\boldsymbol{B}$. By (2.13) and Cauchy-Schwarz inequality, it follows that for $\alpha \to 0$,

$$\delta_\alpha \leq \Big[1/\phi(\tau_\alpha)\Big]\Big\{c\{\exp(-1/2y_\alpha^2)\}y_\alpha^k\Big[1 + O(y_\alpha^{-2})\Big]\Big\}, \tag{2.17}$$

where

$$c = \left\{\int_{\{\boldsymbol{\Delta}^{1/2}\boldsymbol{\tau}>\boldsymbol{0}\}} d\boldsymbol{\tau}\right\}\Big/(2\pi)^{k/2} > 0. \tag{2.18}$$

Using Mill's ratio, for τ_α, we have

$$\tau_\alpha = \Big\{ -2\log\Big(\alpha(2\pi)^{1/2}\Big)\Big\}^{1/2}, \text{ as } \alpha \downarrow 0. \tag{2.19}$$

Also note that

$$\alpha = \sum_{\{\emptyset \subseteq a \subseteq K\}} E_0\Big[1(S_a)\Big] = c y_\alpha^{k-1} e^{-\frac{1}{2}y_\alpha^2}\Big[1 + O(y_\alpha^{-1})\Big], \text{ as } \alpha \downarrow 0, \tag{2.20}$$

and hence, as $\alpha \downarrow 0$,

$$c y_\alpha^{k-1} e^{-\frac{1}{2}y_\alpha^2}\Big[1 + O(y_\alpha^{-1})\Big] = \tau_\alpha^{-1}\phi(\tau_\alpha)\Big[1 + O(\tau_\alpha^{-2})\Big] \tag{2.21}$$

so that

$$\begin{aligned} y_\alpha - \tau_\alpha &= -2(y_\alpha + \tau_\alpha)^{-1}\Big[\log(c\tau_\alpha(2\pi)^{1/2}) + (k-1)\log y_\alpha\Big]\Big[1 + o(1)\Big] \\ &\to 0, \text{ as } \alpha \downarrow 0. \end{aligned} \tag{2.22}$$

Therefore, by (2.17), (2.21) and (2.22),

$$\lim_{\alpha\downarrow 0} \delta_\alpha \le \lim_{\alpha \downarrow 0}\Big\{[1/\phi(\tau_\alpha)] c y_\alpha^k e^{-\frac{1}{2}y_\alpha^2}\Big\} = 1. \tag{2.23}$$

This completes the proof of the theorem.

In passing, we may remark that Theorem 1 relates to the case where $\alpha \downarrow 0$. This relative result that δ_α (defined in (2.11)) is not greater than one still holds even for moderate values of α. Towards this, we present the case $\boldsymbol{\Delta} = \boldsymbol{I}$ (for simplicity of calculation) in the following:

Table 1. Values of the slope-ratio δ_α for typical combinations of (α, k)

$\alpha\backslash k$	2	3	4	5	6	7	8
0.2000	0.8735	0.8230	0.7971	0.7813	0.7707	0.7632	0.7577
0.1000	0.8749	0.8255	0.8000	0.7844	0.7740	0.7665	0.7607
0.0500	0.8712	0.8233	0.7979	0.7827	0.7724	0.7650	0.7592
0.0250	0.8717	0.8240	0.7991	0.7835	0.7733	0.7660	0.7604
0.0100	0.8796	0.8322	0.8070	0.7919	0.7813	0.7738	0.7687
0.0050	0.8814	0.8341	0.8098	0.7938	0.7843	0.7763	0.7708
0.0010	0.8740	0.8256	0.8036	0.7881	0.7754	0.7717	0.7647
0.0001	0.8751	0.8261	0.8000	0.7778	0.8125	0.8000	0.7857

3. Second Order Local Power Comparison

In Theorem 1, we have observed that the slopes of the power functions of the OLT and LRT (at the origin) are proportional to $\boldsymbol{t}'\boldsymbol{\ell}_{J:J'}$, and hence, the same local efficiency (relative to each other) prevails over the entire positive orthant. The picture may not be different when we take into account the curvature of the local power functions at the origin (which behaves differently in the different directions within the positive orthant). For simplicity of presentation, consider here the setup of $\boldsymbol{\Delta} = \boldsymbol{I}$ (a similar picture will hold for an arbitrary $\boldsymbol{\Delta}$), and also a second order local expansion of the two power functions. Note that $\beta_\alpha(T;\boldsymbol{\mu})$ is increasing in each μ_i when the others are held fixed $(1 \leq i \leq k)$. Also, $\beta_\alpha(Q^2,\boldsymbol{\mu}) \geq \beta_\alpha(Q^2;\boldsymbol{\mu}^*)$ if $\boldsymbol{\mu}-\boldsymbol{\mu}^*\varepsilon-\Gamma^0$, where Γ^0 is the dual cone of Γ. Hence, by the same arguments as in Theorem 2.6.2 of Robertson et al. (1988), we conclude that the power function $\beta_\alpha(Q^2;\boldsymbol{\mu})$ is also increasing in each μ_i when the others are held fixed $(1 \leq i \leq k)$. Invoking the symmetry structure (in $\boldsymbol{\mu}$) in these power functions, we claim that the minimum of each power function is attained at the k edges: $\mu_i > 0$, $\mu_j = 0$, $\forall j \neq i$; $i = 1,\ldots,k$, and the maximum of each power function is attained at the half-line $\mu_1 = \cdots = \mu_k > 0$. By the somewhere most powerful character of the OLT, it also follows that maximum power of the OLT (on this half-line) is never below that of the LRT. Hence, we compare the two local powers (up to the second order) over one of the edges, for example, the edge $\mu\boldsymbol{e}_1$, where $\boldsymbol{e}_1 = (1,\boldsymbol{0}')'$, $\mu > 0$. Then note that

$$\beta''_\alpha(T;\boldsymbol{t}) = (\partial^2/\partial\mu^2)\beta_\alpha(T;\mu\boldsymbol{t})|_{\mu=0} = (\boldsymbol{c}'\boldsymbol{t})^2\tau_\alpha\phi(\tau_\alpha),\ \forall \boldsymbol{t} \geq \boldsymbol{0}, \tag{3.1}$$

where $\boldsymbol{c}' = k^{-1/2}(1,\ldots,1) = k^{-1/2}\boldsymbol{1}'_k$. Also, after some routine steps, it follows that for every $\boldsymbol{t} \geq \boldsymbol{0}$; $\|\boldsymbol{t}\|^2 = 1$,

$$\begin{aligned}
&\beta''_\alpha(Q^2;\boldsymbol{t})\\
=\ &(\partial^2/\partial\mu^2)\beta_\alpha(Q^2;\mu\boldsymbol{t})\Big|_{\mu=0}\\
=\ &2^{-k}\sum_{r=0}^{k}\sum_{\{i_1,\cdots,i_r\}}\Bigg\{\sum_{j=1}^{r} t_{i_j}^2\Big[P\{\chi^2_{r+2}\geq y^2_\alpha\}-P\{\chi^2_r\geq y^2_\alpha\}\Big]\\
&+(2/\pi)\Big[\sum_{j=1}^{r}\sum_{m=1_{j\neq m}}^{r} t_{i_j}t_{i_m}P\{\chi^2_{r+2}\geq y^2_\alpha\}-\sum_{j=r+1}^{k}\sum_{m=r+1_{j\neq m}}^{k} t_{i_j}t_{i_m}P\{\chi^2_r\geq y^2_\alpha\}\Big]\Bigg\}\\
=\ &2^{-k}\left(\sum_{r=0}^{k}\binom{k-1}{r}\Big[P\{\chi^2_{r+3}\geq y^2_\alpha\}-P\{\chi^2_{r+1}\geq y^2_\alpha\}\Big]\right.\\
&\left.-\frac{2}{\pi}\sum_{r=0}^{k-2}\binom{k-2}{r}\Big[P\{\chi^2_{r+4}\geq y^2_\alpha\}-P\{\chi^2_r\geq y^2_\alpha\}\Big]\right)\\
&+(\boldsymbol{c}'\boldsymbol{t})^2(2^{k-1}\pi)^{-1}k\sum_{r=0}^{k-2}\binom{k-2}{r}\Big[P\{\chi^2_{r+4}\geq y^2_\alpha\}-P\{\chi^2_r\geq y^2_\alpha\}\Big],
\end{aligned} \tag{3.2}$$

where χ_p^2 denotes a central chi-squared random variable with p degrees of freedom (DF) and $\{i_1, \ldots, i_k\}$ presents a permutation of $1, \ldots, k$. The first term (sum) on the right hand side of (3.2) does not depend on $\boldsymbol{t}$ while the last term is a function of $(\boldsymbol{c}'\boldsymbol{t})^2$. Thus, the two curvatures in (3.1) and (3.2) are not proportional to each other (for every $\boldsymbol{t}$ on the unit sphere (restricted to the positive orthant)). Hence, a Taylor's expansion of the two power functions (up to the second order) reveals that the local power efficiency does not remain constant in all directions over the positive orthant space. Define

$$\gamma_\alpha(\boldsymbol{t}) = \beta_\alpha''(Q^2; \boldsymbol{t})/\beta_\alpha''(T; \boldsymbol{t}), \text{ for every } \boldsymbol{t} \geq \boldsymbol{0} \text{ (where } \|\boldsymbol{t}\|^2 = 1). \tag{3.3}$$

Then, for $\boldsymbol{\mu} = \mu \boldsymbol{e}_1$, i.e., $\boldsymbol{t} = \boldsymbol{e}_1$, and for small $\mu (> 0)$,

$$\begin{aligned} & \beta_\alpha(Q^2; \mu\boldsymbol{e}_1) - \beta_\alpha(T; \mu\boldsymbol{e}_1) \\ = \; & k^{-1/2}\mu\phi(\tau_\alpha)(\delta_\alpha - 1) + \frac{1}{2k}\tau_\alpha\phi(\tau_\alpha)\mu^2(\gamma_\alpha(\boldsymbol{e}_1) - 1) + O(\mu^3) \\ = \; & k^{-1/2}\mu\phi(\tau_\alpha)\Big[(\delta_\alpha - 1) + \{(\mu\tau_\alpha)/(2k^{1/2})\}\{(\gamma_\alpha(\boldsymbol{e}_1) - 1)\} + O(\mu^2)\Big], \end{aligned} \tag{3.4}$$

where

$$\delta_\alpha = \exp\{(\tau_\alpha^2 - y_\alpha^2)/2\}k^{1/2}2^{-k+1}\sum_{r=0}^{k-1}\binom{k-1}{r}\Big\{(y_\alpha^2/2)^{r/2}\Big/\Gamma(r/2+1)\Big\}, \tag{3.5}$$

and

$$\gamma_\alpha(\boldsymbol{e}_1) = 2^{-k}k\sum_{r=0}^{k}\binom{k-1}{r}\Big[P\{\chi_{r+3}^2 \geq y_\alpha^2\} - P\{\chi_{r+1}^2 \geq y_\alpha^2\}\Big]\Big/\{\tau_\alpha\phi(\tau_\alpha)\}. \tag{3.6}$$

Recall that if $g_p(x)$ stands for the density function of the central chi square d.f. with p DF, the $P\{\chi_{r+2}^2 \geq x\} - P\{\chi_r^2 \geq x\} = 2g_{r+2}(x) = (2/r)g_r(x)x$, $x \geq 0$, so that using (2.21), (2.22) and the above identity, it follows from (3.5) and (3.6) that

$$\begin{aligned} \lim_{\alpha\downarrow 0}\delta_\alpha &= \lim_{\alpha\downarrow 0}\left\{\sqrt{\left(\frac{k}{\pi}\right)}\frac{\Gamma(k/2)[-2\log\{2^{(3k-1)/2}\Gamma(k/2)\alpha\}]^{1/2}}{\Gamma((1+k)/2)[-2\log((2\pi)^{1/2}\alpha)]^{1/2}}\right\} \\ &= (k/\pi)^{1/2}\Big[\Gamma(k/2)/\Gamma((k+1)/2)\Big], \end{aligned} \tag{3.7}$$

and

$$\lim_{\alpha\downarrow 0}\gamma_\alpha(\boldsymbol{e}_1) = 1, \text{ for every } k \geq 1. \tag{3.8}$$

Consequently, for every fixed $\mu > 0$ (3.4) is non-positive for very small values of α. For moderate values of α, numerical study indicates that the OLT may not have a very distinct advantage over the LRT for not very small $\mu > 0$. Towards this, we present the following numerical results:

Table 2. Table for the values of $\gamma_\alpha(e_1)$ for some typical

$\alpha \backslash k$	2	3	4	5	6	7	8
0.200	1.7093	2.1682	2.5238	2.8218	3.0828	3.3174	3.5327
0.100	1.4490	1.7416	1.9704	2.1637	2.3328	2.4855	2.6256
0.050	1.3353	1.5576	1.7320	1.8797	2.0103	2.1279	2.2371
0.025	1.2727	1.4551	1.5994	1.7217	1.8302	1.9283	2.0183
0.010	1.2306	1.3815	1.5014	1.6040	1.6950	1.7774	1.8527
0.005	1.2069	1.3405	1.4489	1.5407	1.6223	1.6960	1.7640

Table 2 shows that for moderate values of α, the $\gamma_\alpha(e_1)$ are much larger than their asymptotes, and this is reflected in the following two tables as well.

Table 3. Table for the values of $\beta_\alpha(T;\mu e_1)$ for some specific values of (μ,α,k)

α	k	2	3	4	5	6	7	8
0.200	$\mu=0.2$	.2408	.2328	.2281	.2250	.2226	.2208	.2194
0.100		.1264	.1211	.1180	.1159	.1144	.1133	.1123
0.050		.0663	.0631	.0612	.0599	.0590	.0583	.0577
0.025		.0345	.0326	.0314	.0307	.0302	.0298	.0294
0.010		.0143	.0134	.0129	.0125	.0123	.0121	.0119
0.005		.0074	.0069	.0066	.0064	.0062	.0061	.0060
0.200	$\mu=0.3$	.2634	.2509	.2435	.2386	.2350	.2322	.2300
0.100		.1417	.1331	.1282	.1249	.1225	.1207	.1192
0.050		.0759	.0705	.0675	.0654	.0639	.0628	.0619
0.025		.0402	.0370	.0351	.0339	.0331	.0324	.0319
0.010		.0171	.0155	.0146	.0141	.0136	.0133	.0131
0.005		.0089	.0080	.0076	.0073	.0070	.0069	.0067
0.200	$\mu=0.4$	.2870	.2696	.2595	.2527	.2477	.2439	.2408
0.100		.1581	.1459	.1390	.1343	.1310	.1284	.1264
0.050		.0866	.0787	.0742	.0713	.0692	.0676	.0663
0.025		.0468	.0419	.0392	.0374	.0362	.0352	.0345
0.010		.0203	.0179	.0166	.0157	.0151	.0147	.0143
0.005		.0108	.0094	.0087	.0082	.0078	.0076	.0074

Table 4. Table for the values of $\beta_\alpha(Q^2; \mu e_1)$ for some specific values of (μ, α, k)

α	k	2	3	4	5	6	7	8
0.200	$\mu = 0.2$	.2385	.2299	.2352	.2222	.2200	.2183	.2170
0.100		.1251	.1193	.1162	.1142	.1128	.1117	.1109
0.050		.0650	.0616	.0597	.0585	.0576	.0570	.0565
0.025		.0337	.0317	.0306	.0299	.0294	.0290	.0287
0.010		.0140	.0131	.0126	.0123	.0120	.0119	.0117
0.005		.0072	.0067	.0064	.0062	.0061	.0060	.0059
0.200	$\mu = 0.3$	.2608	.2475	.2401	.2353	.2318	.2292	.2271
0.100		.1401	.1308	.1259	.1228	.1205	.1188	.1174
0.050		.0742	.0687	.0656	.0637	.0623	.0623	.0604
0.025		.0391	.0358	.0341	.0329	.0321	.0315	.0310
0.010		.0166	.0151	.0142	.0137	.0133	.0130	.0128
0.005		.0087	.0078	.0073	.0070	.0068	.0067	.0065
0.200	$\mu = 0.4$	.2851	.2667	.2564	.2497	.2448	.2412	.2382
0.100		.1568	.1437	.1368	.1323	.1291	.1266	.1247
0.050		.0845	.0766	.0723	.0695	.0676	.0660	.0649
0.025		.0452	.0406	.0380	.0363	.0352	.0343	.0336
0.010		.0196	.0173	.0161	.0153	.0148	.0143	.0140
0.005		.0104	.0091	.0084	.0079	.0076	.0074	.0072

In passing, we make a few comments on Tables 3 and 4. First, the power of the OLT, in Table 3 for a given (μ, α, k) is the exact power while the power of the LRT in Table 4, is approximate (up to the second order). Since the third derivative of the power function of the LRT at the null point is still positive, the LRT compares very favorably with the OLT over the edges for not very small μ.

4. The Covariance Matrix has an Unknown Scalar Factor

In this section, we consider the case of $\boldsymbol{\Delta} = \sigma^2 \boldsymbol{\Delta}_0$, where $\boldsymbol{\Delta}_0$ is known (n.n.d.) and σ^2 unknown. Recall that we have n i.i.d. r.v.'s $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and $\bar{\boldsymbol{X}}_n = n^{-1} \sum_{i=1}^n \boldsymbol{X}_i$. In this setup, we may estimate σ^2 by

$$(nk-1)S_n^2 = \sum_{i=1}^n \boldsymbol{X}_i' \boldsymbol{\Delta}_0^{-1} \boldsymbol{X}_i - n \bar{\boldsymbol{X}}_n' \boldsymbol{\ell}_{J:J'} \boldsymbol{\ell}_{J:J'}' \boldsymbol{X}_n, \tag{4.1}$$

where $\boldsymbol{\ell}_{J:J'}$ is defined in the same way as in (2.2) with $\boldsymbol{\Delta}_0$ equivalent up to a scalar factor. Note that the OLT statistic corresponding to the MSSMP similar region for testing H_0 vs. H_1 is given by

$$T_n^0 = \sqrt{n} \boldsymbol{\ell}_{J:J'}' \bar{\boldsymbol{X}}_n / S_n. \tag{4.2}$$

In the particular case of $n = 1$ and $\boldsymbol{\Delta}_0 = \boldsymbol{I}$, Shi and Kudô (1987) have developed an alternative MSSMP form of T_n^0, and the current one is a natural generalization of theirs. As such, under H_0, T_0^0 has the student t-distribution with $nk-1$ DF; in the Shi and Kudô (1987) form, the corresponding DF is equal to $k-1$. Similarly the LRT statistic is of the form

$$Q_n^{0^2} = \sum_{\emptyset \subseteq a \subseteq K} \left\{ \frac{n\bar{\boldsymbol{X}}'_{n(a:a')}\boldsymbol{\Delta}^{-1}_{0(aa:a')}\bar{\boldsymbol{X}}_{n(a:a')}}{\sum_{i=1}^{n} \boldsymbol{X}_i'\boldsymbol{\Delta}_0^{-1}\boldsymbol{X}_i} \right\} 1\left\{\bar{\boldsymbol{X}}_{n(a:a')} > \boldsymbol{0},\ \boldsymbol{\Delta}^{-1}_{0(a'a')}\bar{\boldsymbol{X}}_{n(a')} \le \boldsymbol{0}\right\}, \tag{4.3}$$

where the corresponding partitions of vectors and matrices are defined the same way as in (2.3)–(2.5) with $\boldsymbol{\Delta}$ replaced by $\boldsymbol{\Delta}_0$. Then, under H_0, the distribution of $Q_n^{0^2}$ is given by

$$P_0\left\{Q_n^{0^2} \le c\right\} = \sum_{\ell=0}^{k} w_\ell B_{\frac{\ell}{2},\frac{nk-\ell}{2}}(c) \tag{4.4}$$

where $w_\ell = \sum_{\{a:|a|=\ell\}} P_0\{\boldsymbol{Z}_{a:a'} > \boldsymbol{0}, \boldsymbol{\Delta}^{-1}_{0(a'a')}\boldsymbol{Z}_{a'} \le \boldsymbol{0}\}$, $\ell = 0,1,\ldots,k$, with $\boldsymbol{Z} \sim N(\boldsymbol{0},\boldsymbol{\Delta}_0)$ as well as $|a|$ being the cardinality of the set a, and $B_{p,q}$ represents a random variable having a beta distribution with parameters p and q ($B_{0,q} \equiv 0$). Without loss of generality, take $\boldsymbol{\Delta}_0 = \boldsymbol{I}$, also let ρ_α and $t_{nk-1;\alpha}$ be the α-level critical value of $Q_n^{0^2}$ and the upper $100\alpha\%$ point of the student t-distribution with $nk-1$ DF respectively. Then after some routine calculations, for every $\boldsymbol{t} \ge \boldsymbol{0}$, $\|\boldsymbol{t}\|^2 = 1$,

$$\begin{aligned} \beta'_\alpha(T_n^0;\boldsymbol{t}) &= (\partial/\partial\mu)\beta_\alpha(T_n^0;\mu\boldsymbol{t})\Big|_{\mu=0} \\ &= \frac{1}{2}(\boldsymbol{1}'\boldsymbol{t})\sqrt{\frac{n}{k}}\sqrt{\frac{2}{\pi}}P\left\{B_{1,\frac{nk-1}{2}} \ge \frac{t^2_{nk-1;\alpha}}{(nk-1)+t^2_{nk-1;\alpha}}\right\} \end{aligned} \tag{4.5}$$

and

$$\begin{aligned} \beta'_\alpha(Q_n^{0^2};\boldsymbol{t}) &= (\partial/\partial\mu)\beta_\alpha(Q_n^{0^2};\mu\boldsymbol{t})\Big|_{\mu=0} \\ &= (\boldsymbol{1}'\boldsymbol{t})\sqrt{\frac{2}{\pi}}\sqrt{n}2^{-k}\sum_{\ell=0}^{k-1}\binom{k-1}{\ell} \\ &\quad \times \left[P\left\{B_{\frac{\ell+2}{2},\frac{nk-\ell-1}{2}} \ge \rho_\alpha\right\} - P\left\{B_{\frac{\ell}{2},\frac{nk-\ell+1}{2}} \ge \rho_\alpha\right\}\right]. \end{aligned} \tag{4.6}$$

Note that both (4.5) and (4.6) depend on the direction vector $\boldsymbol{t}$ through the multiplicative factor $\boldsymbol{1}'\boldsymbol{t}$, so that if we define

$$\delta^0_\alpha(\boldsymbol{t}) = \beta'_\alpha(Q_n^{0^2};\boldsymbol{t})\Big/\beta'_\alpha(T_n^0;\boldsymbol{t}) \tag{4.7}$$

then $\delta_\alpha^0(\boldsymbol{t}) = \delta_\alpha^0$ uniformly in $\boldsymbol{t}$; $\boldsymbol{t} \geq \boldsymbol{0}$ and $\|\boldsymbol{t}\|^2 = 1$, where

$$\begin{aligned}\delta_\alpha^0 &= \sqrt{k}2^{-k+1}\sum_{\ell=0}^{k-1}\binom{k-1}{\ell}\Big[P\Big\{B_{\frac{\ell+2}{2},\frac{nk-\ell-1}{2}} \geq \rho_\alpha\Big\} - P\Big\{B_{\frac{\ell}{2},\frac{nk-\ell+1}{2}} \geq \rho_\alpha\Big\}\Big]\Big/ \\ &\quad P\Big\{B_{1,\frac{nk-1}{2}} \geq \frac{t_{nk-1;\alpha}^2}{(nk-1)+t_{nk-1;\alpha}^2}\Big\}. \end{aligned} \tag{4.8}$$

Write $z_\alpha = t_{nk-1;\alpha}^2 \Big/ \Big[(nk-1)+t_{nk-1;\alpha}^2\Big]$, and note that for every positive ρ_α and z_α

$$\frac{1}{2}P\Big\{B_{\frac{1}{2},\frac{nk-1}{2}} \geq z_\alpha\Big\} = \alpha \text{ and } 2^{-k}\sum_{\ell=0}^{k}\binom{k}{\ell}P\Big\{B_{\frac{\ell}{2},\frac{nk-\ell}{2}} \geq \rho_\alpha\Big\} = \alpha. \tag{4.9}$$

For $n = 1$, the minimal level of significance of $Q_n^{0^2}$ is 2^{-k} (as $B_{\frac{k}{2},0} = 1$). Hence take $n \geq 2$. Then $n \geq 2$, as $\alpha \downarrow 0$, and we have

$$(1-z_\alpha)^{\frac{nk-1}{2}} \sim \alpha(nk-1)B\Big(\frac{1}{2},\frac{nk-1}{2}\Big) \tag{4.10}$$

and

$$(1-\rho_\alpha)^{\frac{(n-1)k}{2}} \sim \alpha 2^{-k+1}(n-1)kB\Big(\frac{k}{2},\frac{(n-1)k}{2}\Big), \tag{4.11}$$

where $a \sim b$ means that $a/b \to 1$.

From (4.8), (4.10) and (4.11), it follows that

$$\begin{aligned}\lim_{\alpha\downarrow 0}\delta_\alpha^0 &= \Big(\frac{k}{2}\Big)^{1/2}\Gamma\Big(\frac{k}{2}\Big)\Big/\Gamma\Big(\frac{k+1}{2}\Big) &(4.12)\\ &= e_k^0, \text{ say.} &(4.13)\end{aligned}$$

We may remark that e_k^0 is equivalent to $\lim_{\alpha\downarrow 0}\delta_\alpha(\boldsymbol{t})$ (defined in (3.7)) when $\boldsymbol{\Delta} = \boldsymbol{I}$ and it is easy to see that e_k^0 is less than one for every $k \geq 2$. In passing, we also note that by arguments parallel to those in the proof of Theorem 1, and the results of (4.10)–(4.11), the conclusion in Theorem 1 remains true for the case $\boldsymbol{\Delta} = \sigma^2\boldsymbol{\Delta}_0$. Also note that

$$\begin{aligned}\beta_\alpha''(T_n^0;\boldsymbol{t}) &= (\partial^2/\partial\mu^2)\beta_\alpha(T_n^0;\mu\boldsymbol{t})\Big|_{\mu=0} \\ &= \frac{n}{2}\Big[(k-2)P\Big\{B_{\frac{1}{2},\frac{nk-1}{2}} \geq z_\alpha\Big\} + P\Big\{B_{\frac{3}{2},\frac{nk-1}{2}} \geq z_\alpha\Big\}\Big] \\ &\quad + n\sum_{j=1}^{k}\sum_{j'=1,j\neq j'}^{k} t_j t_{j'}\mathcal{E}\{n\bar{X}_{nj}\bar{X}_{nj'}1(T_n^0 \geq t_{nk-1;\alpha})\} \end{aligned} \tag{4.14}$$

and

$$\begin{aligned}\beta''_\alpha(Q_n^{0^2};t) &= (\partial^2/\partial\mu^2)\beta_\alpha(Q_n^{0^2};\mu t)\Big|_{\mu=0}\\ &= n2^{-k}\sum_{\ell=0}^{k}\binom{k}{\ell}\Big[kP\Big\{B_{\frac{\ell+2}{2},\frac{nk-\ell}{2}}\ge\rho_\alpha\Big\}-P\Big\{B_{\frac{\ell}{2},\frac{nk-\ell}{2}}\ge\rho_\alpha\Big\}\Big]\\ &\quad+n\sum_{j=1}^{k}\sum_{j'=1,j'\neq j}^{k}t_jt_{j'}\mathcal{E}\Big\{n\bar{X}_{n_j}\bar{X}_{n'_j}1(Q_n^{0^2}\ge\rho_\alpha)\Big\}. \qquad (4.15)\end{aligned}$$

Thus, for small μ,

$$\begin{aligned}&\beta_\alpha(Q_n^{0^2};\mu e_1)-\beta_\alpha(T_n^0;\mu e_1)\\ &= \beta'_\alpha(T_n^0;e_1)(\delta_\alpha^0-1)\mu+\beta''_\alpha(T_n^0;e_1)[\gamma_\alpha^0(e_1)-1]\frac{\mu^2}{2}+O(\mu^3) \qquad (4.16)\end{aligned}$$

where

$$\begin{aligned}\gamma_\alpha^0(e_1) &= 2^{-k+1}\sum_{\ell=0}^{k}\binom{k}{\ell}\Big[kP\Big\{B_{\frac{\ell+2}{2},\frac{nk-\ell}{2}}\ge\rho_\alpha\Big\}-P\Big\{B_{\frac{\ell}{2},\frac{nk-\ell}{2}}\ge\rho_\alpha\Big\}\Big]\Big/\\ &\quad\Big[(k-2)P\Big\{B_{\frac{1}{2},\frac{nk-1}{2}}\ge z_\alpha\Big\}+P\Big\{B_{\frac{3}{2},\frac{nk-1}{2}}\ge z_\alpha\Big\}\Big]. \qquad (4.17)\end{aligned}$$

By using (4.9)–(4.11), we then have

$$\lim_{\alpha\downarrow 0}\gamma_\alpha^0(e_1)=1 \quad \forall\, k\ge 1. \qquad (4.18)$$

Therefore similar conclusions to those in Section 3 can be made; we omit the details.

Acknowlegements

The authors are grateful to the referees for their most useful comments on the manuscript.

References

Abelson, R. P. and Tukey, J. W. (1963). Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. *Ann. Math. Statist.* **34**, 1347–1369.

Bartholomew, D. J. (1959a,b). A test of homogeneity for ordered alternatives. *Biometrika* **46**, 36–48; 328–335.

Kallenberg, W. C. M. (1978). Asymptotic optimality of likelihood ratio tests in exponential families. *Math. Centre Tract.* **77**, Amsterdam, Netherlands.

Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–418.

Nüesch, P. E. (1966). On the problem of testing location in multivariate populations for restricted alternatives. *Ann. Math. Statist.* **37**, 113–119.

Oosterhoff, J. (1969). Combination of one-sided tests. *Math. Centre Tracts* **28**, Amsterdam, Netherlands.

Perlman, M. D. (1969). One sided testing problems in multivariate analysis. *Ann. Math. Statist.* **40**, 549–567.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* John Wiley, New York.

Schaafsma, W. and Smid, L. J. (1966). Most stringent somewhere most powerful tests against alternatives restricted by a number of linear inequalities. *Ann. Math. Statist.* **37**, 1161–1172.

Shi, N. Z. (1987). Testing a normal mean vector against the alternative determined by a convex cone. *Mem. Fac. Sci. Kyushu Univ.* **41**, 133–145.

Shi, N. Z. and Kudô, A. (1987). The most stringent somewhere most powerful one-sided test of the multivariate normal mean. *Mem. Fac. Sci. Kyushu Univ.* **29**, 303–328.

Snijders, T. A. B. (1979). Asymptotic optimality theory for testing problems with restricted alternatives. *Math. Centre Tracts* **113**, Amsterdam, Netherlands.

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.
Department of Biostatistics and Statistics, University of North Carolina, Chapel Hill, NC 27599-7400, U.S.A.

(Received November 1990; accepted May 1992)