國立交通大學

統計學研究所

碩 士 論 文

多維度參考區間

Multivariate Reference Regions

研究生：陳羽偉

指導教授：陳鄰安 博士

中華民國一百年六月

多維度參考區間

Multivariate Reference Regions

Student: Yu-wei Chen　　Advisors: Dr. Lin-an Chen

國立交通大學

統計學研究所

碩士論文

A Thesis

Submitted to Institute of Statistics College of Science

National Chiao Tung University

In partial Fullfillment of the Requirement

For the Degree of Master

In

Statistics

June 2010

Hsinchu, Taiwan, Republic of China

中華民國一百年六月

多維度參考區間

研究生：陳羽偉　　指導教授：陳鄰安 博士

國立交通大學統計學研究所

摘要

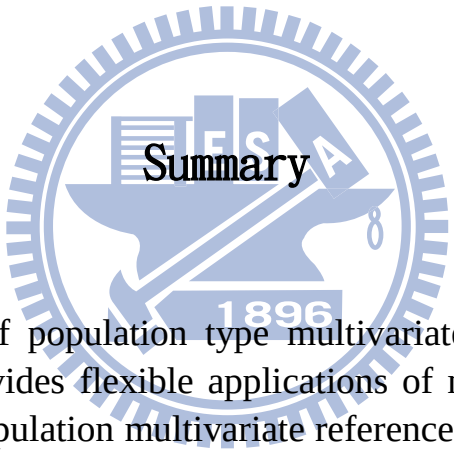
我們介紹了一個非常一般性的多維度參考區間觀念。給定幾個多維機率分配，我們提供不同型態的母體多維度參考區間。因此傳統統計推論中的估計與檢定皆可應用於這一未知區間之預測。在設定為多元常態分配下我們提供了詳細的估計與檢定的方法與模擬分析討論。

Multivariate Reference Regions

Student: Yu-wei Chen Advisors: Dr. Lin-an Chen

Institute of Statistics

National Chiao Tung University



Summary

A general concept of population type multivariate reference region is introduced. This provides flexible applications of multivariate reference region. Given one population multivariate reference region, its estimation and hypothesis testing are important topics in statistical inference for this unknown region. We present several examples of population multivariate reference regions. Given multivariate normal case techniques and criterions for estimation and hypothesis testing are presented and evaluated.

誌　　謝

在碩士這兩年中，由衷的感謝指導教授陳鄰安老師的指導與教誨，讓我不只能順利完成碩士學業，也在其中學到不少做人處事的道理。在做論文的過程中，教授讓我學會如何發現問題，思考問題，和如何解決問題。相信經由這個過程的學習，在未來人生的道路中，碰到任何的挫折與問題時，我都能迎刃而解。也非常謝謝所上的教授們在這兩年中的教導，因為教授們認真仔細的指導，我學會了許多統計分析的技巧。

研究所的生活，非常感謝一起奮戰統研盃的隊友們于憶、士傑、洋德、為翔和瀚宇，一起研究電腦知識的祥福，以及班上的所有人，沒有你們我的碩士生活將不會過得這麼的不平凡和充實。

最後謝謝一路在我徬徨無助或遭遇困難時支持我的爸媽和女友，沒有你們的鼓勵，將不會有今天的我，謝謝你們。

陳羽偉　謹誌于
國立交通大學統計學研究所
中華民國一百年六月

Contents

Multivariate Reference Regions

SUMMARY

A general concept of population type multivariate reference region is introduced. This provides flexible applications of multivariate reference region. Given one population multivariate reference region, its estimation and hypothesis testing are important topics in statistical inference for this unknown region. We present several examples of population multivariate reference regions. Given multivariate normal case, techniques and criterions for estimation and hypothesis testing are presented and evaluated.

Key words: Estimation; hypothesis testing; reference interval; multivariate reference region.

1. Introduction

The determination of intervals to provide reference limits is fundamentally important in clinical chemistry. The reference interval in laboratory chemistry refers to population-based reference values obtained from a well-defined group of reference individuals. This is an interval with two confidence limits which covers the measurement values in the population in some probabilstic sense. The reference interval tells the physician if the patient's value is expected in a healthy or diseased individual or if further testing is warranted. For review of reference intervals, see Horn and Pesce (2003) and Hung, Chen and Welsh (2010).

Most medical decisions require consideration of several co-existing pieces of information, and because these pieces such as blood constituents are often correlated, the multivariate reference regions is more useful than conventional univariate reference

intervals for interpreting clinical laboratory results. There is the uncomfortable statistical fact that when many clinical tests are run on a blood sample from a healthy person, there is a high probability that at least one result will lie outside its reference interval. This indicates that a multidimensional point of correlated observations is likely to lie within the individual's multivariate reference region, even when one or more of the observations lie outside their separate reference intervals for the individual (see Schoen (1970) and Harris, Yasaka et al. (1982)).

Although multivariate reference regions in the practice of clinical chemistry and laboratory medicine is very important, however, it has been received only limited attention in literature and applications. The major reason for this is that there is lack of a natural ordering for multivariate data. This reason also make the existed proposals of multivariate reference regions more or less ad hoc and then most existed ones do not have parametrized versions so that their applications are extremely limited (see Chen and Welsh (2002)). This leads to an unfortunate result. Those laboratories that can not perform their own detailed reference region (interval) studies may need to validate reference regions published elsewhere for their own populations. However, validation of a reference region or interval is generally done through statistical inferences technique such as confidence interval or hypothesis testing that is not allowed to do so if a multivariate reference region is not a sample realization of a population type multivariate reference region. This paper aims to introduce some general but systematic and concise techniques in constructions of probabilistic population multivariate reference regions that allows us to establish statistical inferences such as estimation and hypothesis testing for this unknown region.

In section 2, we introduce general concepts of population multivariate reference region. Examples of this population region for multivariate normal distribution and an beta related multivariate distribution are introduced. In Section 3, a techique of multivariate reference region that may be tranformed from multivariate rectangle is introduced and studied. In Section 4, we introduce two criterions of estimation of unknown multivariate reference region. Simulation results for crietrions of area and mean square error (MSE) are presented. In Section 5, we present a technique that can test a hypothesis of location parameters and scale parameters simultaneously as a tool for validation of a multivariate reference region for a laboratory's population.

2. Multivariate Reference Regions for Independent Transformation Available Distribution

Let Y be random vector of p variables with joint probability density function (pdf) $f(y, \theta)$ where θ parameter vector in Θ. We denote the sample space of random vector Y by Γ_y.

Definition 2.1. A θ dependent subset $C_y(\theta)$ of space Γ_y is called the γ reference region if it satisfies

$$P_\theta(Y \in C_y(\theta)) = \gamma \text{ for } \theta \in \Theta.$$

The interest is how to develop γ reference region $C_y(\theta)$ for a distribution of Y. The difficulty in constructing γ reference region for Y is that elements of Y are generally correlated.

Definition 2.2. We say that the distribution of a random vector Y is independence-transformable if there is invertible function $Z = G(Y, \theta)$ such that elemets, $Z_1, ..., Z_p$, of Z are independent with, respectively, parameter-free pdf's $f_1(z_1), ..., f_p(z_p)$.

Let us denote the sample space of vector Z by Γ_z. In the following example, we present two independence transformable distributions.

Example 1. (a) Suppose that Y has multivariate normal distribution $N_p(\mu, \Sigma)$ where Σ is positive definite matrix. We know that $Z = \Sigma^{-1/2'}(Y - \mu)$ is p vector of i.i.d. random variables with standard normal distribution $N(0, 1)$. Hence Y is independence-transformable where he sample space of Y and Z are, respectively, $\Gamma_y = \Gamma_z = R^n$.

(b) Suppose that bivariate random vector $\binom{Y_1}{Y_2}$ has a joint pdf

$$f_{Y_1Y_2}(y_1, y_2) = \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} y_2^{\alpha-1}(1-y_2)^{\beta-1}(\frac{y_1}{y_2})^{\alpha+\beta-1}(1-\frac{y_1}{y_2})^{\gamma-1}\frac{1}{y_2}, 0 < y_1 < y_2 < 1.$$

By letting $\binom{Z_1}{Z_2} = \binom{Y_2}{Y_1/Y_2}$, we may see that Z_1 and Z_2 are independent random variables, respectively, with distributions $beta(\alpha, \beta)$ and $beta(\alpha+\beta, \gamma)$. Hence Y is independence-transformable where the sample space of Z is $\Gamma_z = (0, 1) \times (0, 1)$ and sample space of Y is $\Gamma_y = \{\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} : 0 < y_1 < y_2 < 1\}$. □

We now consider that for p-vector Y, we have a transformed vector $Z = G(Y, \theta)$ that includes independent and parameter-free elements $Z_1, ..., Z_p$. Then, reference region $C_y(\theta)$ may be constructed based on a Z-based reference region through inversion.

Definition 2.3. Suppose that there is a Z-based γ reference region C_z, a subset of Γ_Z. We define the γ reference region for distribution of Y as

$$C_y(\theta) = G_\theta^{-1}(C_z) = \{y \in \Gamma : G(y, \theta) \in C_z\} \tag{2.1}$$

where G_θ^{-1} is the inversion of function G.

Two approaches are available for construction of Z-based γ reference region C_z. First, in some situations, we can introduce C_z through a univariate mapping on Z. Second, since Z has independent elements, it is allowed to construct C_Z through product of element-wise reference intervals. We first introduce the second approach.

Definition 2.4. If there is distribution of a univariate mapping $Q_z = q(Z)$ so that a γ coverage interval of Q_z, denoted by C_q, is available, then we have

$$C_z = \{z : q(z) \in C_q\}.$$

In the follwoing example, we presents two methods in constructing the univariate mapping when Y has a mutivariate normal distribution.

Example 2. Again, let Y be with the multivariate normal distribution $N_p(\mu, \Sigma)$ and we let the independence-transformation be $Z = \Sigma^{-1/2'}(Y - \mu)$. Let the elements of vector Z be $Z_1, ..., Z_p$.

(a) We then have one univariate mapping $(Y - \mu)'\Sigma^{-1}(Y - \mu) = \sum_{i=1}^{p} Z_i^2$ that has chi-square distribution $\chi^2(p)$. One popularly way in construct a γ reference region for Z is based on this chi-square transform as

$$C_Z = \{z \in R^p : z'z \leq \chi_\gamma^2\}$$

where χ_γ^2 is the γ quantile point of the chi-square distribution $\chi^2(p)$ and we implement γ coverage interval $C_q = (0, \chi_\gamma^2]$ for chi-square variable $Q = \sum_{i=1}^{p} Z_i^2$. Through the inversion, we have

$$\begin{aligned} C_y(\mu, \Sigma) =& \{y = \mu + \Sigma^{1/2} z : z'z \leq \chi_\gamma^2\} \\ =& \{y : (y - \mu)'\Sigma^{-1}(y - \mu) \leq \chi_\gamma^2\}. \end{aligned} \tag{2.2}$$

(b) We can consider the univariate mapping $\frac{1}{\sqrt{p}}1_p'Z = \frac{1}{\sqrt{p}}1_p'\Sigma^{-1/2'}(Y-\mu) \sim N(0,1)$. Since $(-\phi(\frac{1+\gamma}{2}), \phi(\frac{1+\gamma}{2}))$ covers the standard normal random variable with probability γ, an alternative γ reference region for Z is

$$C_Z = \{z \in R^p : \frac{1}{\sqrt{p}}1_p'z \in (-\Phi^{-1}(\frac{1+\gamma}{2}), \Phi^{-1}(\frac{1+\gamma}{2}))\}. \tag{2.3}$$

Through inversion, we have

$$C_y(\mu,\Sigma) = \{y = \mu + \Sigma^{1/2}z : \frac{1}{\sqrt{p}}1_p'z \in (-\Phi^{-1}(\frac{1+\gamma}{2}), \Phi^{-1}(\frac{1+\gamma}{2}))\}. \quad \square$$

3. Reference Region Transformed from Multivariate Rectangles

In this section, we start from constructing coverage intervals for independent transformed variables $Z_1, ..., Z_p$ and then take inversion from the product of these element-wise coverage intervals.

Definition 3.1. Let $C_1, ..., C_p$ be, respectively, the $\gamma^{1/p}$ coverage intervals for independent variables $Z_1, ..., Z_p$. With product $C_z = C_1 \times C_2 \times ... \times C_p$, we may define the γ reference region for distribution of Y as $C_y(\theta) = G_\theta^{-1}(C_z)$.

The following example gives the γ reference region from the multivariate rectangle.

Example 3. (a) We continue the settings for multivariate normal vector Y and transformation Z. By letting $\delta = \gamma^{1/p}$, we choose quantile $z_{\frac{1+\delta}{2}}$. Let C_z be the product of p $\gamma^{1/p}$-coverage intervals, respectively, for $Z_1, ..., Z_p$. Then the reference region for Y then is

$$C_y(\mu,\Sigma) = \{y : y = \mu + \Sigma^{1/2}z, z \in C_z\}.$$

Generally we choose shortest element-wise coverage intervals that leads to the product C_z as

$$C_z = \{ \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{pmatrix} : -\Phi^{-1}(\frac{1+\delta}{2}) \leq z_j \leq \Phi^{-1}(\frac{1+\delta}{2}), j = 1, ..., p\}$$

while a general type product reference region is

$$C_z = \{ \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{pmatrix} : z_{j1} \leq z_j \leq z_{j2}, j = 1, ..., p \text{ with } P(z_{j1} \leq Z_j \leq z_{j2}) = \gamma^{1/p}\}.$$

(b) We next consider the beta distribution case. With $\delta = \gamma^{1/2}$, a product of coverage intervals is

$$C_z = \{ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} : z_1 \in (F_{Z_1}^{-1}(\frac{1-\delta}{2}), F_{Z_1}^{-1}(\frac{1+\delta}{2})), z_2 \in (F_{Z_2}^{-1}(\frac{1-\delta}{2}), F_{Z_2}^{-1}(\frac{1+\delta}{2}))\}.$$

The reference region for Y then is

$$C_y(\alpha, \beta) = \{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} z_1 z_2 \\ z_1 \end{pmatrix} : \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in C_z\}. \quad \square$$

4. Estimators and Statistical Properties for Reference Region

Now, suppose that we have a random sample $Y_1, ..., Y_n$ from the distribution $f(y, \theta)$. It is desired to introduce concepts and methods of statistical inferences when the γ reference region is unknown. In our settings, the γ reference regions are unknown due to that there unknown distribution parameters. Hence, statistical inferences for unknown γ reference region may be reduced to inferences for unknown distribution parameters. We start from point estimation aspect.

Definition 4.1. Let T be a random region in Γ constructed by the random sample $Y_1, ..., Y_n$. We define its expectation as $E(T) = \{E(t_n) : t_n \in T\}$ if all $E(t_n)$ exist

and probability limit $Plim(T) = \{Plim(t_n) : t_n \in T\}$ if all $Plim(t_n)$ exist where $Plim(t_n) = a$ if t_n converges to a in probability. We then say that an estimator $\hat{C}_y(\theta)$ is unbiased estimator of $C_y(\theta)$ if $E(\hat{C}_y(\theta)) = C_y(\theta)$ and it is consistent for $C_y(\theta)$ if $Plim(\hat{C}_y(\theta)) = C_y(\theta)$.

An estimator of γ reference region $C_y(\theta)$ may be obtained by plugging θ by $\hat{\theta}$ when estimator $\hat{\theta}$ is available.

Definition 4.2. Let $\hat{\theta}$ be an estimator of parameter θ. We let estimator of γ reference region be $\hat{C}_y(\theta) = C_y(\hat{\theta})$. Then $\hat{C}_y(\theta)$ is a maximum likelihood estimator (mle) of $C_y(\theta)$ if mle $\hat{\theta}$ exists.

Example 4. Now, suppose that random sample $Y_1, ..., Y_n$ is drawn from normal distribution $N_p(\mu, \Sigma)$. We know that $\bar{Y}$ and $\hat{\Sigma} = S_y = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(Y_i - \bar{Y})'$ are, respectively, mle's of μ and Σ.

In straight forward way, the estimator of γ reference regions for $C_y(\mu, \Sigma)$ of (2.2) and (2.3) are, respectively,

$$\hat{C}_y(\mu, \Sigma) = \{y = \bar{Y} + S_y^{1/2} z : z'z \leq \chi^2_\gamma\}$$

and

$$\hat{C}_y(\mu, \Sigma) = \{y = \bar{Y} + S_y^{1/2} z : \frac{1}{\sqrt{p}} 1_p' z \in (-\Phi^{-1}(\frac{1+\gamma}{2}), \Phi^{-1}(\frac{1+\gamma}{2}))\}.$$

These two estimated γ reference regions are consistent, respectively, for $C_y(\mu, \Sigma)$ of (2.2) and (2.3) since $\bar{Y}$ and S_y are, resctively, consistent for μ and Σ. $\square$

We here consider some other criterions for evaluation of estimator of γ reference region. Let us denote the area of the true γ reference region by A_C. With replications $m = 10,000$, we perform a simulation from normal distribution

$N_2(0_2, \Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{12} \\ \sigma_{21} & \sigma_y^2 \end{pmatrix})$. Let $A^i_{\hat{C}}$ represents the area of the estimate $\hat{C}_y(\mu, \Sigma)$ in ith replication. We define the averaging area as

$$A_{\hat{C}} = \frac{1}{m} \sum_{i=1}^{m} A^i_{\hat{C}}$$

and the square root of the mean square error (MSE) as

$$SMSE_{\hat{A}} = (\frac{1}{m} \sum_{i=1}^{m} (A^i_{\hat{C}} - A_C)^2)^{1/2}.$$

The simulated results of averaging area $A_{\hat{C}}$, square root of the mean square error $SMSE_{\hat{A}}$ associated with the true area A_C are displayed in Table 1.

Table 1. Comparison of areas of estimated and true γ reference region associated with $SMSE_{\hat{A}}$

	A_C	$A_{\hat{C}}$ ($SMSE_{\hat{A}}$) $n = 30$	$n = 50$	$n = 70$	$n = 100$
$\sigma_y^2 = 1$					
$\sigma_{12} = 0.3$	10.32	9.834 (2.22)	10.00 (1.91)	10.07 (1.49)	10.10 (1.27)
$\sigma_{12} = 0.5$	9.372	8.949 (2.07)	9.042 (1.68)	9.137 (1.37)	9.249 (1.10)
$\sigma_{12} = -0.3$	10.32	9.746 (2.23)	10.06 (1.75)	10.16 (1.52)	10.16 (1.30)
$\sigma_{12} = -0.5$	9.372	8.935 (2.40)	9.026 (1.67)	9.189 (1.29)	9.202 (1.20)
$\sigma_y^2 = 0.3$					
$\sigma_{12} = 0.09$	3.09	2.910 (0.69)	2.986 (0.53)	3.013 (0.46)	3.059 (0.38)
$\sigma_{12} = 0.15$	2.811	2.690 (0.61)	2.730 (0.48)	2.750 (0.40)	2.773 (0.36)
$\sigma_{12} = -0.09$	3.097	2.917 (0.68)	2.991 (0.54)	3.057 (0.44)	3.044 (0.37)
$\sigma_{12} = -0.15$	2.811	2.649 (0.62)	2.712 (0.47)	2.755 (0.41)	2.755 (0.34)

Several comments may be drawn from the results in Table 1:

(a) In terms of area for a region, the estimated area and the true area of γ reference region in the designed cases are all under estimated, however, not with too much differences.

(b) As expected, the variation showing in MSE is larger when the variance σ_y^2 is larger.

Comparison of areas and MSE's between the estimated one and the true region is not sufficient to evaluate the efficiency of an estimator of unknown γ reference region. It requires further study to see if the estimated and the true regions are really overlapping closely. For this need, we define the area non-overllaped between these two as

$$ANL = \text{Non-overlapping Area} = A_C + A_{\hat{C}} - 2A_{C_y(\mu,\Sigma)\cap\hat{C}_y(\mu,\Sigma)}$$

and the following MSE

$$\begin{aligned} MSE &= \frac{1}{m}\sum_{j=1}^{m}\left(\frac{\text{Non-overlapping Area}}{2\times(\text{Length }(C(\theta))+\text{Width }(C(\theta))}\right)^2 \\ &= \frac{1}{m}\sum_{j=1}^{m}\left(\frac{A_C + A_{\hat{C}} - 2A_{C_y(\mu,\Sigma)\cap\hat{C}_y(\mu,\Sigma)}}{2\times(\text{Length }(C(\theta))+\text{Width }(C(\theta)))}\right)^2 \end{aligned}$$

where we choose this denominator term to make this MSE dimension-free. The simulated results of ANL and MSE are displayed in Table 2.

Table 2. Efficiencies of estimation of γ reference region through area difference and MSE

	$n = 30$	$n = 50$	$n = 70$	$n = 100$
$\sigma_y^2 = 0.3$				
$\sigma_{12} = 0.09, ANL$	1.074	0.869	0.745	0.641
MSE	0.149	0.120	0.103	0.088
$\sigma_{12} = 0.15, ANL$	0.987	0.794	0.669	0.583
MSE	0.137	0.110	0.092	0.081
$\sigma_{12} = -0.09, ANL$	1.065	0.868	0.746	0.649
MSE	0.147	0.120	0.103	0.090
$\sigma_{12} = -0.15, ANL$	0.981	0.795	0.677	0.591
MSE	0.136	0.110	0.094	0.082
$\sigma_y^2 = 1$				
$\sigma_{12} = 0.3, ANL$	4.000	3.068	2.552	2.099
MSE	0.304	0.233	0.194	0.159
$\sigma_{12} = 0.5, ANL$	3.275	2.625	2.237	1.972
MSE	0.248	0.199	0.170	0.149
$\sigma_{12} = -0.3, ANL$	3.622	2.885	2.480	2.109
MSE	0.275	0.219	0.188	0.160
$\sigma_{12} = -0.5, ANL$	3.289	2.621	2.237	1.930
MSE	0.250	0.199	0.170	0.146

We have two comments for the simulated results:

(a) The non-overlapping area ANL decreases and then the efficiency of point estimator increases when sample size n rises or variance σ_y^2 decreases.

(b) The dimension-free MSE shows that the estimation of true reference region is satisfactory.

5. Testing for Hypothesis of Reference Region

The establishment of reference region requires careful planning, control, and documentation of each aspect of the study. Thus, the resulting reference regions are well-characterized in terms of the variation attributable to pre-analytical and analytical factors. With this consideration, to establish a laboratory's own reference region (interval) is difficult due to costs and forces. Even large laboratories are finding it increasingly difficult to conduct these comprehensive studies cost-effectively.

Therefore, laboratories are becoming more reliant on manufacturers to establish scientifically sound reference regions that can be verified using simpler, less labor-intensive, and lower cost approaches. One important approach requiring less effort for the establishment of reference regions (intervals) is the validation through hypothesis testing to verify if an established reference region can match the use for this laboratory's specific population. This task can be done statistically only when the unknown reference region is function of distributional parameters.

We require the γ reference region $C_y(\theta)$ for the laboratory's population to be dependent on unknown parameter θ that fulfills

$$\gamma = P_\theta(Y \in C_y(\theta)) \text{ for } \theta \in \Theta.$$

When the γ reference region is unknown, we only know that it is one with the space of $C_y(\theta)$ as

$$\{C_y(\theta) : \theta \in \Theta\}. \tag{5.1}$$

We assume that $C_y(\theta_1) \neq C_y(\theta_2)$ if $\theta_1 \neq \theta_2$. Any set $D \subset \Gamma_y$ is a γ reference region if there exists $\theta_0 \in \Theta$ such that $D = C_y(\theta_0)$. Hence, testing hypothesis of γ reference region such as

$$H_0 : C_y(\theta) = D$$

is equivalent to test the hypothesis of unknown parameter as

$$H_0 : \theta = \theta_0. \tag{5.2}$$

Suppose that the random vector Y has the normal distribution $N_p(\mu, \Sigma)$. Then a testing hypothesis of any type of γ reference region is equivalent to test the following hypothesis of distribution parameters

$$H_0 : \mu = \mu_0, \Sigma = \Sigma_0. \tag{5.3}$$

In literature, we can see approaches for testing hypothesis about mean vector as $H_0 : \mu = \mu_0$ and approaches for testing hypothesis about covariance matrix as $H_0 : \Sigma = \Sigma_0$. It is rare to have approaches to test hypothesis for mean vector and covariance matrix simultaneously. The hypothesis about the reference region is reduced to test hypothesis in (5.3) that requires a new test.

With the normality assumption, we have

$$(Y - \mu_0)'\Sigma_0^{-1}(Y - \mu_0) \sim \chi^2(p)$$

when H_0 is true. Suppose that we further have a random sample $Y_1, ..., Y_n$ from $N_p(\mu, \Sigma)$. Then,

$$Q = \sum_{i=1}^{n}(Y_i - \mu_0)'\Sigma_0^{-1}(Y_i - \mu_0) \sim \chi^2(np)$$

when H_0 is true. A rule for testing H_0 is

$$\text{rejecting } H_0 \text{ if } Q \geq \chi_\alpha(np)$$

where $\chi_\alpha(np)$ is the $(1-\alpha)$th quantile of the chi-square distribution $\chi^2(np)$.

We consider the hypothesis (5.3) by choosing data from the following distribution

$$N(\mu_0 + r\begin{pmatrix}1\\1\end{pmatrix}, \Sigma_0 + \lambda I_2)$$

where $(r, \lambda) = (0, 0)$ corresponds to distribution of H_0. With replications $m = 10,000$, we perform a silmulation to verify the power performance of this chi-square test defined as

$$\frac{1}{m}\sum_{j=1}^{m} I(Q_j \geq \chi_\alpha(np))$$

where Q_j is the observation of statistic Q from j-th sample. By settings $\mu_0 = \begin{pmatrix}0\\0\end{pmatrix}$ and $\Sigma_0 = \begin{pmatrix}1 & 0.3\\0.3 & 1\end{pmatrix}$, we display the simulated results in Tables 3 and 4 respectively for significance level $\alpha = 0.05$ and 0.1.

Table 3. Power performance ($\alpha = 0.05$)

(r, λ)	$n = 30$	$n = 50$	$n = 70$	$n = 100$
$(0, 0)$	0.0466	0.0517	0.0500	0.0503
$(0.5, 0)$	0.2707	0.3640	0.4549	0.5645
$(1, 0)$	0.9461	0.9980	1	1
$(0, 0.5)$	0.7888	0.9287	0.9756	0.9971
$(0, 1)$	0.9886	0.9996	1	1
$(0.25, 0.25)$	0.4860	0.6471	0.7684	0.8774
$(0.5, 0.5)$	0.9253	0.9864	0.9982	1

Table 4. Power performance ($\alpha = 0.1$)

(r, λ)	$n = 30$	$n = 50$	$n = 70$	$n = 100$
$(0, 0)$	0.1016	0.0940	0.0962	0.1017
$(0.5, 0)$	0.3875	0.5041	0.5959	0.6890
$(1, 0)$	0.9744	0.9980	1	1
$(0, 0.5)$	0.8706	0.9594	0.9889	0.9986
$(0, 1)$	0.9948	1	1	1
$(0.25, 0.25)$	0.6149	0.7668	0.8581	0.9305
$(0.5, 0.5)$	0.9613	0.9955	0.9999	1

We have several comments drawn from the simulated power results in Tables 3 and 4:

(a) The results for $(r, \lambda) = (0, 0)$ are all close to values α's ensuring that this is a level α test.

(b) Large sample size does improve to raise the power.

(c) Power performance reflecting from the shift in scale is stronger than the shift in location.

References

Chen, L.-A. and Welsh, A. H. (2002). Distribution-function-based bivariate quantiles. *Journal of Multivariate Analysis*, 83, 208-231.

Harris, E. K., Yasaka, T., Horton, M., R. and Shakarji, G. (1982). Comparing multivariate and univariate subject-specific reference regions for blood constituents in healthy persons. *Clinical Chemistry*, 28, 422-426.

Horn, P. S. and Pesce, A. J. (2003). Reference intervals: an update. *Clinica Chimica Acta*, 334, 5-23.

Huang, J.-Y., Chen, L.-A. and Welsh, A.H. (2010). A note on reference limits. *IMS Collections*, Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Hornor of Professoe Jana Jureckova. 7, 84-94.

Schoen, I. and Brooks, S. (1970) Judgement based on 95% confidence limits. *American Journal of Clinical Pathology*. 53, 190-193.