Journal of Statistical Research *ISSN 0256 - 422 X*
2008, Vol. 42, No. 1, pp. 37-44
Bangladesh

A NOTE ON THE MEAN VOLUME OF CONFIDENCE ELLIPSOID FOR THE MEAN OF MULTIVARIATE NORMAL DISTRIBUTION

Jie Mi
Department of Statistics, Florida International University
Miami, FL 33199, USA
Email: mi@fiu.edu

SUMMARY

This paper shows that more data or information is better in estimating the mean of a multivariate normal distribution. Precisely, the mean volume of confidence ellipsoid for the mean decrease with addition of independent observation. This result is obvious when the variance-covariance matrix of the multivariate normal distribution is known, but it not so straight to see the result when the variance-covariance matrix is unknown.

Keywords and phrases: Hotelling's T^2; confidence ellipsoid; projection; mean volume; change of sign.

AMS Classification: 62F10, 62F11

1 Introduction

It is widely believed that more data or information is always better in statistical inference. Some are so extreme that they think this is an obvious principle and there is no need to verify it mathematically at all. However, this intuitive or naive principle is, in fact, not always true. Hengartner (1999) gave an example in which two independent samples of binary random variables were drawn; in the first one, the probability of success is p (unknown); in the second one the probability of success is q (unknown) and it is known that $q < p$. It is found in Hengartner (1999) that depending on the value of p, the maximum likelihood estimator using both samples has a larger variance and mean squared error than the maximum likelihood estimator of p that uses only the first sample. Kim and Verducci (1999) gave another example in which it is possible to construct a uniformly most powerful (UMP)test for a sample of size one, but there is no UMP test for larger sample size. Both examples are in contradiction with the above mentioned naive principle. Therefore, for each individual statistical inference, if the effect of more sampling is a concern, then the foregoing naive principle should be carefully verified unless it is trivial. In the present paper

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

we will show that the naive principle holds in estimating the mean of multivariate normal population.

In multivariate data analysis often we want to obtain confidence region for an unknown parameter vector. In particular, in the case of multivariate, say p-dimensional, normal population $N_p(\mu, \Sigma)$, we want to obtain the confidence region for the population mean μ. If the population variance-covariance matrix Σ is unknown, then based on Hotelling's T^2 test statistic, the confidence region is actually a p-dimensional ellipsoid. If Σ is known, then in addition to Hotelling's T^2 method, we can use chisquare distribution to obtain confidence region for μ. The resulting confidence region is also a p-dimensional ellipsoid. These methods are discussed, for instance, in Anderson (1984), Johnson and Wichern (1998), and Rencher (1998). From the confidence ellipsoids obtained from applying either Hotelling's T^2-method or chisquare method we can further find their projections (shadow) into any q-dimensional subspace with $1 \leq q \leq p$ to obtain the confidence regions for the projections of μ onto these subspaces. These projections are again ellipsoids but of course, of q-dimension.

The above discussion indicates that if Σ is known in the case of $N_p(\mu, \Sigma)$ population, there are two approaches to constructing confidence ellipsoid for μ, and consequently to obtaining confidence ellipsoids for the projections of μ onto all q-dimensional subspaces. In this paper we will study the effect of the sample size on the mean volume of such ellipsoids. Mean volumes of these ellipsoids correspond to the mean width in the univariate case. In Section 2 some preliminary results are stated. It is proved in Section 3 that as sample size n increases, the mean volume of those confidence ellipsoids strictly decreases no matter which method is used for constructing confidence region. That is, more data is better in estimating the mean of a multivariate normal population. From this aspect this article validates the naive principle.

2 Mean Volume of Confidence Ellipsoid

Let $X_1, X_2, \ldots, ; X_n$ be a random sample from p-dimensional normal population $N_p(\mu, \Sigma)$ where Σ is positive definite. Suppose we are interested in the confidence regions for the projection of μ onto a q-dimensional subspace space. That is, if $\{u_1, \ldots, u_q\}$ is a set of orthogonal basis of a given q-dimensional subspace and matrix U has $u_1, \ldots, u_q$ as its columns, then we want to obtain the confidence regions of vectors $U'\mu$. Suppose Σ is known, then there are two ways to obtain the confidence regions for $U'\mu$. Let $0 < 1 - \alpha < 1$ be any given confidence coefficient. Without using Σ, from the Hotelling's T^2-method, a $100(1-\alpha)\%$ confidence region for μ is the p-dimensional ellipsoid determined by

$$\left\{\mu : \quad n(\bar{X} - \mu)' S^{-1} (\bar{X} - \mu) \leq c_T^2\right\} \tag{2.1}$$

where $\bar{X}$ and S, respectively, are the sample mean and sample variance-covariance matrix based on $X_1, \ldots, X_n$, and

$$c_T^2 = \frac{p(n-1)F_\alpha(p, n-p)}{n-p},$$

where $F_\alpha(p, n-p)$ is the upper α-percentile of F-distribution with numerator and denominator degrees of freedom p and $n-p$, respectively. Then the confidence regions for $U'\mu$ are the projections of the ellipsoid (2.1) onto the q-dimensional space spanned by $u_1, \ldots, u_q$. On the other hand, if Σ is applied, then we can used chisquare method and then a $100(1-\alpha)\%$ confidence region for μ is the p-dimensional ellipsoid given by

$$\left\{\mu : \quad n(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu) \leq c_Z^2\right\} \tag{2.2}$$

where $c_Z^2 = \chi_\alpha^2(p)$ and $\chi_\alpha^2(p)$ is the upper α-percentile of chisquare distribution with p degrees of freedom since $n(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu)$ follows chisquare distribution with p degrees of freedom. Then the confidence regions for $U'\mu$ are the projections of the ellipsoid (2.2) onto the q-dimensional subspaces spanned by $u_1, \ldots, u_q$. In order to investigate the effect of increasing the sample size on the mean volume of the confidence region from the above two approaches, we have to obtain mathematical expressions for the mean volumes of the simultaneous confidence ellipsoids. In the following we use V_T and V_Z to denote the volumes of the q-dimensional ellipsoids obtained as projections of the p-dimensional ellipsoids (2.1) and (2.2), respectively, onto the subspace spanned by $u_1, u_2, \ldots, u_q$.

Applying Hotelling's T^2-method, the projection of the p-dimensional ellipsoid (2.1) onto the subspace spanned by $u_1, \ldots, u_q$ is given by

$$\left\{v : \ n(U\bar{X}-v)'(U'SU)^{-1}(U\bar{X}-v) \leq c_T^2\right\} \tag{2.3}$$

where $v \equiv U'\mu$ belongs to the subspace spanned by $u_1, \ldots, u_q$. This fact regarding projection can be found, for instance, in Johnson and Wichern (1998). The volume of the ellipsoid (2.3) is given by

$$V_T = k_q\sqrt{\det(U'SU)}\ c_T^q \ / \ n^{q/2} \tag{2.4}$$

where the constant k_q is defined by

$$k_q = \frac{2\pi^{q/2}}{q\Gamma(q/2)}.$$

The calculation of volume of ellipsoid is given, for example, in Cramér (1946). Similarly we can obtain the expression of V_Z as

$$V_Z = k_q\sqrt{\det(U'\Sigma U)}\ c_Z^q \ / \ n^{q/2}. \tag{2.5}$$

To compute the mean volume $E(V_T)$ we have to obtain $E(\sqrt{\det(U'SU)})$ as can be seen from (2.4). It is known that $(n-1)S$ follows Wishart distribution with $(n-1)$ degrees of freedom, $(n-1)S \sim W_{n-1}(\cdot|\Sigma)$. It yields $U'(n-1)SU \sim W_{n-1}(\cdot|U'\Sigma U)$. Further, since the distribution of the determinant of the generalized sample variance det(S) is the same as the distribution of $\det(\Sigma)/(n-1)^p$ times the product of q independent factors, the

distribution of the ith factor being the chisquare distribution with $n-i$ degrees of freedom, the distribution of $\sqrt{\det(U'SU)}$ is the same as that of

$$\frac{\sqrt{\det(U'\Sigma U)}}{(n-1)^{q/2}} \prod_{i=1}^{q} \xi_{n-i},$$

where $\xi_{n-i},\ 1 \le i \le q$ are independent and ξ_{n-i} follows chi-distribution with degrees of freedom $n-i$. Then we can obtain

$$\begin{aligned} E\left(\sqrt{\det(U'SU)}\right) &= \frac{\sqrt{\det(U'\Sigma U)}}{(n-1)^{q/2}} \prod_{i=1}^{q} \frac{\sqrt{2}\Gamma((n-i+1)/2)}{\Gamma((n-i)/2)} \\ &= \frac{\sqrt{\det(U'\Sigma U)}}{(n-1)^{q/2}} \cdot 2^{q/2} \frac{\Gamma(n/2)}{\Gamma((n-q)/2)} \end{aligned} \tag{2.6}$$

3 Change of Mean Volume of Confidence Ellipsoid

It is easy to believe intuitively that statistical inference will be more accurate in certain sense as the sample size increases since that would include more information. In this section we consider the following problem. Suppose that we want to obtain confidence ellipsoid for $U'\mu$ where U and μ are the same as defined in the previous section. Will the mean volumes of these ellipsoids decrease as sample size increases? The answer to this question is easy in the case of known variance-covariance matrix Σ. Actually from (2.5) we immediately see that $E(V_Z) = V_Z$ strictly decreases in $n > p$ since the constant $k_q\sqrt{\det(U'\Sigma U)}\ c_Z^q$ does not depend on n at all. However, it is not easy to answer the same question when Σ is not known. To this end we first prove the following auxiliary result.

Lemma. For any $1 \le q \le p < n$, the following inequality holds

$$\left(\frac{\Gamma((n+1)/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((n+1-q)/2)}\right)^{p/q} \left(\frac{\Gamma((n+1)/2)\Gamma((n-p)/2)}{\Gamma(n/2)\Gamma((n+1-p)/2)}\right)^{-1} \le 1 \tag{3.1}$$

and the equality holds if and only if $p = q$.

(Proof) The inequality (3.1) is clearly true when $p = q$. Actually the left hand side of (3.1) equals to 1 exactly when $p = q$. Hence in the following we assume $q < p$. For $z \ge n/2$ define

$$\varphi(z) = \left(\frac{\Gamma(z+1/2)\Gamma(z-q/2)}{\Gamma(z)\Gamma(z-(q-1)/2)}\right)^{p/q} \left(\frac{\Gamma(z+1/2)\Gamma(z-p/2)}{\Gamma(z)\Gamma(z-(p-1)/2)}\right)^{-1}.$$

The desired inequality is equivalent to $\varphi(n/2) < 1$. We will show that $\varphi(z) < 1$ for all $z \ge n/2$. Taking the natural logarithm, we see that we need to show

$$\ln\varphi(z) = \left(\frac{p}{q} - 1\right)\left[\ln\Gamma(z+\frac{1}{2}) - \ln\Gamma(z)\right] - \frac{p}{q}\left[\ln\Gamma(z-\frac{q-1}{2}) - \ln\Gamma(z-\frac{q}{2})\right]$$

$$+\left[\ln\Gamma(z-\frac{p-1}{2})-\ln\Gamma(z-\frac{p}{2})\right]<0,\quad \forall\ z\geq\frac{n}{2}.$$

Notice that

$$\frac{\varphi'(z)}{\varphi(z)}=\left(\frac{p}{q}-1\right)\left[\psi(z+\frac{1}{2})-\psi(z)\right]-\frac{p}{q}\left[\psi(z-\frac{q}{2}+\frac{1}{2})-\psi(z-\frac{q}{2})\right]$$
$$+\left[\psi(z-\frac{p}{2}+\frac{1}{2})-\psi(z-\frac{p}{2})\right],$$

where $\psi(z)\equiv\Gamma'(z)/\Gamma(z)$ is the digamma function. If we further define

$$\eta(u)\equiv\psi(u+\frac{1}{2})-\psi(u),\quad \forall\ u>0,$$

then

$$\begin{aligned}\frac{\varphi'(z)}{\varphi(z)} &= (\frac{p}{q}-1)\eta(z)-\frac{p}{q}\eta\left(z-\frac{q}{2}\right)+\eta\left(z-\frac{p}{2}\right)\\ &= \frac{p}{q}\left\{(1-\frac{q}{p})\eta(z)+\frac{q}{p}\eta(z-\frac{p}{2})-\eta(z-\frac{q}{2})\right\},\quad z\geq\frac{n}{2}. \qquad (3.2)\end{aligned}$$

Now we claim that $\eta(u)$ is a strict convex function on $(0,\infty)$. As a matter of fact we have

$$\eta''(u)=\psi''(u+\frac{1}{2})-\psi''(u)=\psi'''(u+\theta)$$

with $0<\theta<1/2$. But $\psi'''(u)>0,\ \forall u>0$, thus the claim is true. Since

$$\left(1-\frac{q}{p}\right)z+\frac{q}{p}\left(z-\frac{p}{2}\right)=z-\frac{q}{2},$$

by the strict convexity of $\eta(u)$ we conclude that $\varphi'(z)>0,\ \forall z\geq n/2$ from (3.2). It implies that $\varphi(z)$ is strictly increasing in $z\geq n/2$. Note that $\lim_{z\to\infty}\varphi(z)=1$, therefore $\varphi(z)<1,\ \forall\ z\geq n/2$. Particularly, we obtain $\varphi(n/2)<1$. This ends the proof.

The following result shows that the mean volume of the confidence ellipsoid (2.3) strictly decreases in $n\geq p+1$.

Theorem. For any $0<\alpha<1,\ 1\leq q\leq p<n$, the mean volume $E(V_T)$ strictly decreases in $n\geq p+1$.

(Proof) We denote V_T as $V_T(n)$ to emphasize the dependence of V_T on the sample size n. What we need to show is $E(V_T(n))>E(V_T(n+1))$, for any $n\geq p+1$.

From (2.4) and (2.6) we see that

$$E\Big(V_T(n)\Big)=\frac{L}{(n(n-p))^{q/2}}\ \frac{\Gamma(n/2)}{\Gamma((n-q)/2)}\ \left(F_\alpha(p,n-p)\right)^{q/2},$$

where the constant $L = k_q\sqrt{(2p)^q \det(U'\Sigma U)}$ is independent of n. Hence $E(V_T(n)) > E(V_T(n+1))$ is equivalent to

$$F_\alpha(p, n-p) > b(n,p)\ F_\alpha(p, n+1-p), \tag{3.3}$$

where

$$b \equiv b(n,p) \equiv \frac{n(n-p)}{(n+1)(n+1-p)} \left(\frac{\Gamma((n+1)/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((n+1-q)/2)} \right)^{2/q}.$$

In order to show (3.3) it suffices to show that

$$\int_{bt}^{\infty} f_F(x; p, n-p)dx > \int_{t}^{\infty} f_F(x; p, n+1-p)dx, \quad \forall\ t > 0 \tag{3.4}$$

where $f_F(x; p, n-p)$ is the density function of F-distribution with numerator degrees of freedom p and denominator degrees of freedom $n-p$. Because if (3.4) is true, then particularly choosing $t = F_\alpha(p, n+1-p)$ gives

$$\int_{bF_\alpha(p,n+1-p)}^{\infty} f_F(x, p, n-p)dx > \int_{F_\alpha(p,n+1-p)}^{\infty} f_F(x; p, n+1-p)dx = \alpha$$

and consequently

$$F_\alpha(p, n-p) > bF_\alpha(p, n+1-p).$$

The inequality (3.4) is equivalent to

$$\int_{t}^{\infty} bf_F(bx; p, n-p)dx > \int_{t}^{\infty} f_F(x; p, n+1-p)dx, \quad \forall\ t > 0. \tag{3.5}$$

The function $bf_F(bx; p, n-p)$ obviously is a probability density function on $(0, \infty)$. Hence in order to prove (3.5) it is sufficient to show that $bf_F(bx; p, n-p) - f_F(x; p, n+1-p)$ has exactly one change of sign on $(0, \infty)$ and the change of sign occurs from $-$ to $+$ (See Boland et al.(1989), and Shaked and Shanthikumar (1994)).

We define

$$r(x) \equiv \frac{bf_F(bx; p, n-p)}{f_F(x; p, n+1-p)} = \left(\frac{n}{n+1} \right)^{p/2} \left(\frac{\Gamma((n+1)/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((n+1-q)/2)} \right)^{p/q}$$

$$\cdot \left(\frac{\Gamma((n+1)/2)\Gamma((n-p)/2)}{\Gamma(n/2)\Gamma((n+1-p)/2)} \right)^{-1} \sqrt{g(x)},$$

where

$$g(x) = \frac{\left(1 + px \ /\ (n+1-p)\right)^{n+1}}{\left(1 + pbx \ /\ (n-p)\right)^{n}}.$$

It can be verified that

$$\begin{aligned} g'(x)[1+\frac{pbx}{(n-p)}]^{2n} &= (n+1)[1+\frac{px}{(n+1-p)}]^{n}\frac{p}{n+1-p}\cdot[1+\frac{pbx}{(n-p)}]^{n} \\ &- [1+\frac{px}{(n+1-p)}]^{n+1}\cdot n[1+\frac{pbx}{(n-p)}]^{n-1}\frac{pb}{n-p} \\ & p[1+\frac{px}{(n+1-p)}]^{n}[1+\frac{pbx}{(n-p)}]^{n-1}\{\frac{n+1}{n+1-p}[1+\frac{pbx}{(n-p)}] \\ &- \frac{nb}{n-p}[1+\frac{px}{(n+1-p)}]\} \\ &= p[1+\frac{px}{(n+1-p)}]^{n}[1+\frac{pbx}{(n-p)}]^{n-1} \\ &\times \frac{pbx-[nb(n+1-p)-(n+1)(n-p)]}{(n-p)(n+1-p)} \end{aligned} \tag{3.6}$$

Thus it is easy to see that

$$g(x)\begin{cases} \text{strictly decreases, if} & x<x_0; \\ \text{achieves its minimum, if} & x=x_0; \\ \text{strictly increases, if} & x>x_0, \end{cases}$$

where $x_0 \equiv [nb(n+1-p)-(n+1)(n-p)]/pb$. No matter whether $x_0 \le 0$ or $x_0 \ge 0$, we can see that the equation $r(x)=1$ has exactly one solution on $(0,\infty)$ since

$$\begin{aligned} r(0) &= \left(\frac{n}{n+1}\right)^{p/2}\left(\frac{\Gamma((n+1)/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((n+1-q)/2)}\right)^{p/q}\left(\frac{\Gamma((n+1)/2)\Gamma((n-p)/2)}{\Gamma(n/2)\Gamma((n+1-p)/2)}\right)^{-1} \\ &< \left(\frac{\Gamma((n+1)/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((n+1-q)/2)}\right)^{p/q}\left(\frac{\Gamma((n+1)/2)\Gamma((n-p)/2)}{\Gamma(n/2)\Gamma((n+1-p)/2)}\right)^{-1} \\ &\le 1, \end{aligned}$$

where the second inequality follows from the Lemma, and $\lim_{x\to\infty} r(x)=\infty$. This means that the desired property of change of sign is true, and therefore the result of the theorem follows.

References

[1] Ahlfors, L. V. (1979).*Complex Analysis, 3rd Edition.* McGraw-Hill.

[2] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd Edition.* Wiley, New York.

[3] Boland, P. J., Proschan, F., and Tong, Y. L. (1989). Crossing Properties of Mixture Distributions. *Prob. Eng. Inform. Sci.*, **3**, 355–366.

[4] Cramér, H. (1946).*Mathematical Methods of Statistics.* Princeton Univ. Press.

[5] Hengartner, N. W. (1999). A Note on Maximum Likelihood Estimation, *The American Statistician*. **53**, 123–125.

[6] Johnson, R. A. and Wichern, D. W. (1998) *Applied Multivariate Statistical Analysis, 4th Edition*. Prentice Hall.

[7] Kim, Y. and Verducci, J. S. (1999). Too much sampling kills the UMP test, *Statistics & Probability Letters*, **41**, 101–105.

[8] Mi, J. (1996). Crossing Properties of F Distributions, *Statistics & Probability Letters*, **27**, 289–294.

[9] Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*, Wiley.

[10] Shaked, M. and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications*, Academic Press, Boston.