International Journal of Applied Mathematics & Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 8, Issue 2, Feb - Mar 2019; 35-42

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

ESTIMATION IN THE MULTIVARIATE NORMAL DISTRIBUTION

William W.S. Chen

Research Scholar, Department of Statistics, The George Washington University, Washington

ABSTRACT

We present two methods for estimating the population mean vector and variance-covariance matrix in the multivariate normal distribution. We introduce two algorithms, both of which maximize the loglikelihood function. The first method is based on the least square results, and some proven identities to demonstrate the parameter matrix Φ *replaced by F, the solution of normal equation, can maximize the loglikelihood function. This means the least square solution coincides with the maximum likelihood estimates. The second methods will completely depend on matrix differentiation method. We also discuss the problem of how to identify a given data set that fits the multivariate normal distribution better than other distributions.*

KEYWORDS: *Estimation, Maximize The Loglikelihood Function, Mean Vector, Multivariate Normal Distribution, Q-Q Plot, Skewness and Kurtosis Measure, Test Normality, Variance-Covariance Matrix*

Article History

Received: 18 Feb 2019 | Revised: 23 Feb 2019 | Accepted: 19 Mar 2019

INTRODUCTION

The objective of this paper seeks the estimators of the first two moments of the multivariate normal distribution, since almost allof the multivariate inferential procedures are based on this distribution. Before we seek the solution, we ask how we know this given real data set follows the multivariate normal distribution. In this section, we provide some general concepts and leave the technicalities to the concluding remarks. Many tests and graphical procedures have been suggested for evaluating whether a data set likely originated from a multivariate normal distribution. One possibility is to check each variable separately for univariate normality. There are some good reviews for both the univariate and multivariate case studies given by Gnanadesikan (1997) and Seber(1984). A basic graphical approach for checking normality is the Q-Q plot comparing quantiles of a sample against the population quantiles of the univariate normal. If the points are close to a straight line, then there is no indication of a departure from normality.

Deviation from a straight line indicates nonnormality. Checking for multivariate normality is conceptually not as straightforward as assessing univariate normality. We adopt the procedure set forth in Mardia (1970), which we discuss in more detail in the concluding remarks. After we identify the data set belonging to multivariate normal, we can calculate thefirst two moments. In section 2, we give two algorithms and use both the least square method and the maximum likelihoodmethod to find the best estimate of the population parameters.

The Model

The multivariate general linear model is given by

$$E(\underline{y}_1, \underline{y}_2, \ldots \underline{y}_p) = X(\underline{\psi}_1, \underline{\psi}_2, \ldots \underline{\psi}_p)$$
$$\text{or} \quad E(Y) = X\Phi \qquad (2.1)$$

Where $Y_{(nxp)}$ response matrix, $X_{(nxm)}$ design matrix and $\Phi_{(mxp)}$ parameter matrix. We select the kth column from Y, then we can write the linear model

$$E(\underline{y}_k) = X(\underline{\psi}_k) \qquad (2.2)$$

model, we make the following three assumptions:

$Var(\underline{y}_k^{/}) = \Omega, \ \text{Cov}(\underline{y}_i, \underline{y}_j) = 0, \ i \neq j, 1,2,\ldots p$

and $\underline{y}_i \sim N(\underline{x}_i^{/}\Phi, \Omega)$

Based on these three assumptions and model (2.2), and using the least square technique, we can derive the normal equation,

$$X'XF = X'Y \qquad (2.3)$$

Where F is the well-known least square estimator of the unknown parameter vector Φ. Usually, $X'X$ may not full rank, if $X^{/}X$

is singular then $(X'X)^{(-1)}$ conditional inverse exists and is not unique. We need to find a real matrix C such that $C(X^{/}X)^{(-1)}X^{/}X = C$ to promise CF uniqueness. We prove that the following propositions hold.

Proposition 2.1

$$\textit{Show that } (Y-X\Phi)^{/}(Y - X\Phi) = Y^{/}Y - Y^{/}XF + (F-\Phi)^{/}X^{/}X(F-\Phi)$$

Proof: Aware of the fact that $X'X$ may not be full rank, in general, we consider there exists a conditional inverse. Let

$$F = (X'X)^{(-1)}X^{/}Y, \ \textit{Right side} = Y^{/}Y - Y^{/}XF + (F-\Phi)^{/}X^{/}X(F-\Phi)$$

$$= Y^{/}Y - Y^{/}X(X^{/}X)^{(-1)}X^{/}Y + (Y^{/}X(X^{/}X)^{(-1)} - \Phi^{/})X^{/}X((X^{/}X)^{(-1)}X^{/}Y - \Phi)$$

$$= Y^{/}Y - Y^{/}X(X^{/}X)^{(-1)}X^{/}Y + Y^{/}X(X^{/}X)^{(-1)}X^{/}X(X^{/}X)^{(-1)}X^{/}Y$$

$$-\Phi^{/}X^{/}X(X^{/}X)^{(-1)}X^{/}Y - Y^{/}X(X^{/}X)^{(-1)}X^{/}X\Phi + \Phi^{/}X^{/}X\Phi$$

$$= (Y-X\Phi)^{/}(Y - X\Phi)$$

Proposition 2.2

Show that $I-X(X^{/}X)^{(-1)}X^{/}$ is idempotent matrix

$(I-X(X^{/}X)^{(-1)}X^{/})(I-X(X^{/}X)^{(-1)}X^{/})$

$$=I-X(X^{/}X)^{(-1)}X^{/}-X(X^{/}X)^{(-1)}X^{/}+X(X^{/}X)^{(-1)}X^{/}X(X^{/}X)^{(-1)}X^{/}$$

$$=I-X(X^{/}X)^{(-1)}X^{/}$$

Proposition 2.3

Show that E matrix is positive definite

$$E=Y^{/}Y-Y^{/}XF=Y^{/}Y-Y^{/}X(X^{/}X)^{(-1)}X^{/}Y$$

$$=Y^{/}(I-X(X^{/}X)^{(-1)}X^{/})Y$$

$$=Y^{/}(I-X(X^{/}X)^{(-1)}X^{/})(I-X(X^{/}X)^{(-1)}X^{/})Y=\underline{u}^{/}\underline{u}$$

where $\underline{u}=(I-X(X^{/}X)^{(-1)}X^{/})Y$, $\underline{u}$ is order (nxp) and

$I-X(X^{/}X)^{(-1)}X^{/}$ has rank n-r, where r is the rank of X.

r is effective number of paramdeter. IF $n-r\geq p$

then $\text{rank}(\underline{u})=\min(n_e,p)=p$, so $E=\underline{u}^{/}\underline{u}$ is of rank p $(n_e\geq p)$

E is nonsingular and positive definite.

Proposition 2.4

If Ω is positive definite, then so is Ω^{-1}

Proof: there exists a real vector $\underline{a}^{/}$ such that

$\underline{a}^{/}\Omega\underline{a}>0$ implies that $\underline{a}^{/}\Omega\Omega^{-1}\Omega\underline{a}>0$

let $\underline{b}=\Omega\underline{a}$ then $\underline{b}^{/}\Omega^{-1}\underline{b}>0$ hence Ω^{-1} is positive definite.

It is also well known that of Ω^{-1} is positive definite,

then there exists a matrix B such that $\Omega^{-1}=BB^{/}$

Proposition 2.5

$$tr\Omega^{-1}(Y-\Phi)^{/}X^{/}X(Y-\Phi)=trBB^{/}(Y-\Phi)^{/}X^{/}X(Y-\Phi)$$

$$=trB^{/}(Y-\Phi)^{/}X^{/}X(Y-\Phi)B=tr\underline{u}^{/}\underline{u}$$

$$=tr(gramian\ \text{matrix})>0\ \text{ where }\ \underline{u}=X(Y-\Phi)^{/}B$$

Now, we can spell out the joint density function and show that

the least square estimation may coincide with the maximum likelihood estimator.

$$L(\underline{y_1},\underline{y_2},....\underline{y_p})=\prod_{i=1}^{n} f(\underline{y_1},\underline{y_2},....\underline{y_p})$$

$$=\frac{1}{(2\pi)^{\frac{np}{2}}|\Omega|^{\frac{n}{2}}}Exp\{-\frac{1}{2}\sum_{i=1}^{n}(\underline{y_i}-\underline{x_i'}\Phi)'\Omega^{-1}(\underline{y_i}-\underline{x_i'}\Phi)\} \quad (2.4)$$

Let us pay attention to exponential part only,

$$Exp\{-\frac{1}{2}\sum_{i=1}^{n}(\underline{y_i}-\underline{x_i'}\Phi)'\Omega^{-1}(\underline{y_i}-\underline{x_i'}\Phi)\}$$
$$=Exp\{-\frac{1}{2}tr\Omega^{-1}\sum_{i=1}^{n}(\underline{y_i}-\underline{x_i'}\Phi)(\underline{y_i}-\underline{x_i'}\Phi)'\}$$

$$=Exp\{-\frac{1}{2}tr\Omega^{-1}(Y-X\Phi)(Y-X\Phi)'\}$$
$$=Exp\{-\frac{1}{2}tr\Omega^{-1}(Y'Y-Y'XF+(F-\Phi)'X'X(F-\Phi)\}$$

$$=Exp\{-\frac{1}{2}tr\Omega^{-1}E+tr\Omega^{-1}(F-\Phi)'X'X(F-\Phi)\}$$
$$=Exp\{-\frac{1}{2}tr\Omega^{-1}E+trB'(F-\Phi)'X'X(F-\Phi)B\} \quad (2.5)$$

After taking logarithm on equation (2.4), we finally get,

$$\ln L(\underline{y_1},\underline{y_2},.....\underline{y_p})=-\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\Omega|-\frac{1}{2}tr\Omega^{-1}E-h^2 \quad \text{where } h=X(F-\Phi)B$$

where $h^2 \geq 0$ if $h^2=0$ then $\tilde{\Phi}=F$. Hence LnL is maximized if $\tilde{\Phi}$ is replaced by F.

This demonstrates that F is that maximum likelihood estimator of Φ and its maximum value is:

$$\ln L(\widehat{\Phi}=F)=-\frac{np}{2}\ln(2\pi)-\frac{n}{2}\ln|\Omega|-\frac{1}{2}tr\Omega^{-1}E$$

Where E is matrix of the sum of the square and product of error. If $n_e > p$, then E^{-1} exists. where n_e is degree of freedom of error. In the next section, we will discuss the matrix differentiation method

Matrix Differentiation Method

A random mx1 column vector $\underset{\sim}{X}$ is said to be the multivariate normal distribution if its probability density function is given by

$$f(\underset{\sim}{x}) = (2\pi)^{-\frac{m}{2}} |\Omega|^{-\frac{1}{2}} Exp\{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu})^{/} \Omega^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu})\} \qquad (3.1)$$

for $\underset{\sim}{x} \in R^m$, where $\underset{\sim}{\mu}$ *is* an mx1 *vector* and Ω ,a nonsingular symmetric mxm matrix, In short, $\underset{\sim}{x} \sim N_n(\underset{\sim}{\mu}, \Omega)$, , the parameter $\underset{\sim}{\mu}$ and Ω are just the expectation and variance matrix of $\underset{\sim}{x}$ with $E(\underset{\sim}{x}) = \underset{\sim}{\mu}$ and $\mathrm{Var}(\underset{\sim}{x}) = \Omega$.

We wish to show that the maximum likelihood estimator of $\underset{\sim}{\mu}$ and Ω

are $\widehat{\underset{\sim}{\mu}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{\underset{\sim}{x}}$, $\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}(\underset{\sim}{x_i} - \overline{\underset{\sim}{x}})(\underset{\sim}{x_i} - \overline{\underset{\sim}{x}})^{/}$. From the density function (3.1), we can derive the loglikelihood as follows:

$$\ln L(\underset{\sim}{x}) = -\frac{mn}{2}\ln(2\pi) - \frac{n}{2}\ln|\Omega| - \frac{1}{2} tr\,\Omega^{-1} z \qquad (3.2)$$

where $z = \sum(\underset{\sim}{x_i} - \underset{\sim}{\mu})(\underset{\sim}{x_i} - \underset{\sim}{\mu})^{/}$. In equation (3.2), we can take derivative with respective to $\underset{\sim}{\mu}$ and Ω as follows:

$$\frac{\partial \ln L(\underset{\sim}{x})}{\partial \underset{\sim}{\mu}} = -\frac{1}{2} tr\Omega^{-1}(\frac{\partial z}{\partial \underset{\sim}{\mu}}) = 0$$

$$\frac{\partial z}{\partial \underset{\sim}{\mu}} = \sum\{-\frac{\partial \underset{\sim}{\mu}}{\partial \underset{\sim}{\mu}}(\underset{\sim}{x_i} - \underset{\sim}{\mu})^{/} + (\underset{\sim}{x_i} - \underset{\sim}{\mu})(-\frac{\partial \underset{\sim}{\mu}}{\partial \underset{\sim}{\mu}})^{/}\}$$

$$= \sum\{-2((\underset{\sim}{x_i} - \underset{\sim}{\mu})\} = 0 \qquad \sum \underset{\sim}{x}_i = n\widehat{\underset{\sim}{\mu}}$$

$$\text{so} \quad \widehat{\underset{\sim}{\mu}} = \underset{\sim}{x} \qquad (3.3)$$

$$\frac{\partial \ln L(\underset{\sim}{x})}{\partial \Omega} = -\frac{n}{2}\Omega^{-1} + \frac{1}{2}\Omega^{-1} z \Omega^{-1} = 0$$

$$\Omega^{-1} z \Omega^{-1} = n\Omega^{-1}$$

$$\widehat{\Omega} = \frac{1}{n} z = \frac{1}{n}\sum(\underline{x_i} - \underline{\overline{x}})(\underline{x_i} - \underline{\overline{x}})^{/} \qquad (3.4)$$

The equation (3.3) and (3.4) are the required maximum likelihood estimator of μ and Ω. However, we still need to prove that theyare really make the equation (3.2) maximized. To reach this purpose, we take first derivative with respect to the loglikelihood function (3.2) as follow.

$$d\Lambda(\underline{\mu},\Omega) = -\frac{n}{2} d\log|\Omega| - \frac{1}{2} tr(d\Omega^{-1})z - \frac{1}{2} tr\Omega^{-1}dz$$

$$= -\frac{n}{2} tr\Omega^{-1} d\Omega + \frac{1}{2} tr\Omega^{-1}(d\Omega)\Omega^{-1}z$$

$$+\frac{1}{2} tr\Omega^{-1}\{\sum(\underline{x}_i - \underline{\mu})(d\underline{\mu})' + (d\underline{\mu})\sum(\underline{x}_i - \underline{\mu})'\}$$

$$= \frac{1}{2} tr(d\Omega)\Omega^{-1}(z\text{-}n\Omega)\Omega^{-1} + n(d\underline{\mu})'\Omega^{-1}(\bar{\underline{x}} - \underline{\mu}) \qquad (3.5)$$

If we ignore the symmetry constraint on Ω, we obtain the first order conditions $\Omega^{-1}(z\text{-}n\Omega)\Omega^{-1} = 0$ and $\Omega^{-1}(\bar{\underline{x}} - \underline{\mu}) = 0$ from which

equation (3.3) and (3.4) follow immediately. To prove that we have in fact found the maximum value of equation (3.5), we differentiate (3.5) again. This can yield:

$$d^2\Lambda(\mu,\Omega) = \frac{1}{2} tr(d\Omega)\{d\Omega^{-1}(z\text{-}n\Omega)\Omega^{-1} + \Omega^{-1}[(z - n\Omega)d\Omega^{-1} + (dz\text{-}nd\Omega)\Omega^{-1}]\}$$

$$+n(d\underline{\mu})'[d\Omega^{-1}(\bar{\underline{x}} - \underline{\mu}) + \Omega^{-1}(-d\underline{\mu})']$$

at the point ($(\hat{\underline{\mu}},\hat{\Omega})$, we have $\bar{\underline{x}} = \underline{\mu}$, $\hat{z} - n\hat{\Omega} = 0$ and $d\hat{z} = 0$. Hence

$d^2\Lambda(\mu,\Omega) = -\frac{n}{2} tr(d\Omega)\Omega^{-1}(d\Omega)\Omega^{-1} - n(d\underline{\mu})'\Omega^{-1}d\mu < 0$ unless $d\underline{\mu} = 0$ and $d\Omega = 0$. It follows that Λ has a strict local maximum at

$$(\hat{\underline{\mu}},\hat{\Omega})$$

Concluding Remarks

Mardia(1970) method for assessing multivariate normality is a generalization of the univariate test based on the skewness and kurtosis measures. Let $\underline{y}$ and $\underline{x}$ vectors are independent, identically distributed with mean vector $\underline{\mu}$ and covariance matrix

$\hat{\Omega}$. The skewness and kurtosis for multivariate population are defined by Mardia as :

$$\beta_{1,p} = E((\underline{y} - \underline{\mu})'\Omega^{-1}(\underline{x} - \underline{\mu}))^3, \quad \beta_{2,p} = E((\underline{y} - \underline{\mu})'\Omega^{-1}(\underline{y} - \underline{\mu}))^2$$

Since third-order central moments for the multivariate normal distribution are zero, $\beta_{1,p} = 0$ when $\underline{y} \sim N_p(\underline{\mu},\Omega)$. Then the estimates of $\beta_{1,p}$ *and* $\beta_{2,p}$ are given by,

$$b_{1,p} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}^3, \quad b_{2,p} = \frac{1}{n}\sum_{i=1}^{n} h_{ii}^2, \quad where\ h_{ij} = (\underline{y}_i - \bar{\underline{y}})'\widehat{\Omega^{-1}}(\underline{y}_i - \bar{\underline{y}})$$

Mardia(1970,1974) gives percentage points $b_{1,p}$ and $b_{2,p}$ for p=2,3,4, which can be used in testing for multivariate normality. Forother values of p when n>50, the approximation tests are available. For $b_{1,p}$, the statistic

$$z_1 = \frac{(p+1)(n+1)(n+3)}{6((n+1)(P+1)-6)} b_{1,p} \sim \chi_{\frac{1}{6}p(p+1)(p+2)}$$

We reject the hypothesis of multivariate normality if $z_1 > \chi^2_{0.05}$.

With $b_{2,p}$, we wish to reject the extreme values (too peaked or too flat). For the upper 0.025 points of $b_{2,p}$ use

$$z_2 = \frac{b_{2,p} - p(p+2)}{\sqrt{8p(p+2)/n}} \sim N(0,1)$$

For the lower 0.025 points we break into two possible cases :

When $50 \le n \le 400,$ we use $z_3 = \frac{b_{2,p} - p(p+2)(n+p+1)/n}{\sqrt{8p(p+2)/(n-1)}} \sim N(0,1)$ (2)When $n \ge 400,$ we use z_2.

REFERENCES

1. *Anderson, T.W. (1984) An Introduction to Multivariate Statistical Analysis, second edition, New York, Wiley.*
2. *Mardia, K.V.(1970) Measures of Multivariate Skewness and Kurtosis with application. Biometrika, 57, 519-530.*
3. *Mardia, K.V.(1974) Applications of Some Measures of Multivariate Skewness and Kurtosis for Testing Normality and Robustness Studies. Sankhya, Series B, 36, 115-128.*
4. *. Gnanadesikan, R.(1997)Methods for Statistical Data Analysi sof Multivariate Observations 2nd edition, NewYork, Wiley.*
5. *Seber, G.A.F.(1984) Multivariate Observations, New York, Wiley*