International Journal of Statistics and Applied Mathematics

**ISSN:** 2456-1452
Maths 2021; 6(4): 83-85

www.mathsjournal.com
Received: 15-04-2021
Accepted: 03-06-2021

**Jianqi Yu**
College of Science, Guilin University of Technology, Guilin Guangxi, 541004, China

**Shaoling Ding**
College of Science, Guilin University of Technology, Guilin Guangxi, 541004, China

**Xiang Wang**
College of Science, Guilin University of Technology, Guilin Guangxi, 541004, China

**Corresponding Author:**
**Jianqi Yu**
College of Science, Guilin University of Technology, Guilin Guangxi, 541004, China

# Maximum likelihood estimation for multivariate normal with auxiliary information

**Jianqi Yu, Shaoling Ding and Xiang Wang**

**DOI:** https://doi.org/10.22271/maths.2021.v6.i4b.708

**Abstract**
Closed forms are obtained for the maximum likelihood estimators (MLE) of the mean vectors and the covariance matrix of a multivariate normal samples using known means of auxiliary variables. The likelihood function is decomposed as product of several independent normal and conditional normal likelihood functions. The parameters are transformed into a new set of parameters of which the MLEs are easy to derive. Since the MLE are invariant, the MLE of the original parameters are derived using the inverse transformation.
MSC2000 subject classi cations: 62H12; 62H15.

**Keywords:** Auxiliary variables, maximum likelihood estimators, multivariate normal

## Introduction

Auxiliary information is very common and easily obtained in practice. Making full use of auxiliary information can improve the accuracy of inference effectively. For instance, we usually use the sample mean to estimate the population mean, but when there is auxiliary information, there are other better estimates. Cochran (1940) [1] proposed the ratio estimation of the population mean in simple random sample survey which reached the best when the research variables and auxiliary variables were highly positively correlated and the regression line passed through the origin. The product estimation was firrst proposed by Robson (1957) [5] and rediscovered by Murthy (1964) [4], which is suitable for the situation where the research variables and auxiliary variables are highly negatively correlated. The regression estimation proposed by Watson (1937) [8] is suitable for the case that the regression line of the research variable and auxiliary variable does not pass through the origin.

In this paper, we consider the MLEs for multivariate normal population with auxiliary variables.

Let the research variable be y and p dimensional, the auxiliary variable be x and q dimensional, and the expectation of x is known. And together they satisfy.

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N_{q+p}\left[\begin{pmatrix} c \\ \mu \end{pmatrix}, \begin{pmatrix} \Sigma_{11}, \Sigma_{12} \\ \Sigma_{21}, \Sigma = \Sigma_{22} \end{pmatrix}\right]$$

where c is a known constant. Suppose that there are n independent samples on both y and x, and in addition, there are m extra observations on x solely. In other words, we have a random sample as follows:

$$\begin{gathered} x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_{n+m} \\ y_1, y_2, \ldots, y_n \end{gathered} \tag{1.1}$$

We consider the MLEs of the mean vector $\mu$ and covariance matrix $\Sigma$ of y.

**Maximum Likelihood Estimation**
Partition the data in (1.1) as follows:

$$D_1 = \begin{pmatrix} x_1, x_2, ..., x_n \\ y_1, y_2, ..., y_n \end{pmatrix},$$
$$D_2 = (x_1, x_2, ..., x_n, x_{n+1}, ..., x_{n+m}) \quad (2.1)$$

Let $\overline{D}_1 = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$ and $S = \begin{pmatrix} S_{11}, S_{12} \\ S_{21}, S_{22} \end{pmatrix}$ denote the sample mean vector and the sums of squares and products matrix respectively based on $D_1$. Similarly, let $\overline{D}_2$ and V denote respectively the sample mean vector and the sums of squares and products matrix based on $D_2$.
Consider the density function of data in (1.1). We note that the density of x and y can be written as the marginal density of x times the conditional density of y given x (we indicate the density of normal distribution by f( ) here), that is

$$f(x, y \mid c, \mu, \Sigma_{\bullet\bullet}) = f(x \mid c, \Sigma_{11}) f(y \mid \mu_{2.1} + B_{2.1}x, \Sigma_{2.1})$$

Where

$$B_{2.1} = \Sigma_{21}\Sigma_{11}^{-1}, \mu_{2.1} = \mu - B_{2.1}c, \Sigma_{2.1} = \Sigma_{22} - B_{2.1}\Sigma_{12} \quad (2.2)$$

The likelihood function can be written as

$$L(\mu, \Sigma_{\bullet\bullet}) = \prod_{i=1}^{n+m} f(x_i \mid c, \Sigma_{11}) \prod_{i=1}^{n} f(y_i \mid \mu_{2.1} + B_{2.1}x_i, \Sigma_{2.1}) \quad (2.3)$$

The maximum likelihood estimates of $\Sigma_{11}, \mu_{2.1}, B_{2.1}, \Sigma_{2.1}$ are those values that maximize

To maximize (2.3) with respect to $\Sigma_{11}$, we maximize $\prod_{i=1}^{n+m} f(x_i \mid c, \Sigma_{11})$. This gives us the usual estimates of the parameters of a normal distribution based on n+m observations, namely,

$$\hat{\Sigma}_{11} = \frac{1}{n+m}\sum_{i=1}^{n+m}(x_i - c)'(x_i - c) \quad (2.4)$$

To maximize (2.3) with respect to $\mu_{2.1}, B_{2.1}, \Sigma_{2.1}$, we maximize the second term of the right hand side of (2.3). This gives the usual estimates of regression parameters, namely,

$$\hat{B}_{2.1} = S_{12}S_{11}^{-1}, \hat{\mu}_{21} = \hat{y} - \hat{B}_{2.1}\hat{x}, \hat{\Sigma}_{2.1} = (S_{22} - \hat{B}_{2.1}S_{12})/n \quad (2.5)$$

It is easy to see that the maximum likelihood estimates of the original parameters $\mu, \Sigma_{12}, \Sigma_{22}$ are obtained by solving (2.2), where

$$B_{2.1} = \hat{B}_{2.1}, \mu_{21} = \hat{\mu}_{21}, \Sigma_{2.1} = \hat{\Sigma}_{2.1}$$

Hence, we have

$$\hat{\mu} = \bar{y} - \hat{B}_{2.1}(\bar{x} - c), \hat{\Sigma} = \hat{\Sigma}_{22} = \hat{\Sigma}_{2.1} + \hat{B}_{2.1}\hat{\Sigma}_{12} \quad (2.6)$$

with

$$\hat{B}_{2.1} = S_{12}S_{11}^{-1}, \hat{\Sigma}_{12} = \hat{\Sigma}_{11}\,\hat{B}_{2.1}$$

It is obvious that $\hat{\mu}$ is determined completely by $D_1$ and it is the same as regression estimator in sample survey. However, $\hat{\Sigma}$ is different. It is not only related to $D_1$, but also related to the extra observations on x in $D_2$.

**References**


1. Cochran W. "The estimation of the yields of cereal experiments by sampling for the ratio gain to total produce," Journal of Agriculture Science 1940;30:262-275.
2. Kadilar G. "A new exponential type estimator for the population mean in simple random sampling," Journal of Modern Applied Statistical Methods 2016;15(2):207-214.
3. Khan S, Ali H, Manzoor S, Alamgir. "A class of transformed e cient ratio estimators of nite population mean," Pakistan Journal of Statistics 2015;31(4):353-362.
4. Murthy M. "Product method of estimation," Sankhya. 1964;26:294-307.
5. Robson D. " Application of multivariate polykays to the theory of unbiased ratio type estimation," Journal of American Stattical Association 1957;52:411-422.
6. Singh H, Tailor R. " Use of known correlation coe cient in estimating the nite population means," Statistical. Transition 2003;6(4):555-560.
7. Singh R, Chauhan P, Sawan N, Smarandache F. " Improvement in es-timating the population mean using exponential estimator in simple random sam-pling," Bulletin of Statistics and Economy 2009;3:13-18.
8. Watson D. "The estimation of leaf area in eld crops," Journal of Agriculture Science 1937;27:474-483.
9. Yan Z, Tian B. "Ratio method to the mean estimation using coe cient of skewness of auxiliary variable," ICICA. Part II, 2010;106:103-110.