THE EGYPTINA STATISTICA JOURNAL
ISSR, CAIRO UNIV., VOL., 45, NO.2, 2001

On EstimatingThe Parameters Of The Bivariate Normal Distribution

Ahmed M. M Sultan
Egyptian Air Force
Cairo, Egypt

Albert H. Moore
Air Force Institute of Technology
Patteson AFB, Ohio 45433

Hala Mahmoud Khaleel
Zagazig University
Zagazig, Egypt

Abstract

A technique is applied to estimate the parameters of the bivariate normal distribution with unknown mean vector and unknown covariance matrix by minimizing the Cramer von Mises distance from a non-parametric density estimate and the parametric estimate at the order statistics. The maximum likelihood estimators were found and a comparison was made with the proposed estimator. For different parameters of the true density the proposed estimators were tested using a Monte Carlo experiment . The results show an improvement in mean integrated square error which is taken as a measure of the closeness of the estimated density and the true density.

1. INTRODUCTION

Among the different criteria for the choice of the parameter estimators for a probability density are the unbiasedness, the consistency, the minimum variance and the sufficiency. A number of authors considered the estimation of the parameters of the bivariate normal density using the method of moments, the maximum likelihood, besides other methods. Section 2 discusses the maximum likelihood estimators for the parameters of the bivariate normal density. The log likelihood equations are solved to give the estimators for the mean and covariance matrix. In section 3 the application of a non-parametric density estimator to obtain estimates of the parameters of the bivariate normal distribution is discussed. A Monte Carlo comparison of the maximum likelihood estimators and the minimum distance estimators is given using the integrated squared error between the true density and the estimated true model for a sample size of 5(5)20 and different Monte Carlo repetitions in tables. A comparison is made between the estimators for a given mean vector and a given covariance matrix for sample sizes 5(5)20 in tables and figures.

2. MAXIMUM LIKELIHOOD ESTIMATORS FOR THE PARAMETERS OF THE BIVARIATE NORMAL DENSITY

If $X = [x_{ir}]$; i=1,2 and r=1,2,...,n be a $2 \times n$ matrix which represents a sample of size n from the bivariate nonsingular normal distribution with mean vector $\mu = [\mu_1, \mu_2]'$ and variance covariance matrix $\Sigma = [\sigma_{ij}]$, i,j=1,2, then the joint p.d.f. will be given as:

$$L(\mu,\Sigma) = (2\pi)^{n}|\Sigma|^{n/2}\exp\left\{-\frac{1}{2}tr\Sigma^{-1}(X-\mu E_{1n})(X-\mu E_{1n})'\right\} \qquad \ldots(2\text{-}1)$$

where E_{1n} is an $1 \times n$ matrix with all elements unity.

The loglikelihood function of the sample observations is given by:

$$\begin{aligned}
&\log_e L(\mu,\Sigma)\\
&= -\frac{n}{2}\log_e(2\pi) - \frac{n}{2}\log_e|\Sigma| - \frac{1}{2}tr\Sigma^{-1}(X-\mu E_{1n})(X-\mu E_{1n})'\\
&= -\frac{n}{2}\log_e(2\pi) - \frac{n}{2}\log_e|\Sigma| - \frac{1}{2}tr\Sigma^{-1}(X-\overline{X}E_{1n})(X-\overline{X}E_{1n})'\\
&\qquad -\frac{n}{2}tr\Sigma^{-1}(\overline{X}E_{1n}-\mu E_{1n})(\overline{X}E_{1n}-\mu E_{1n})'\\
&= -\frac{n}{2}\log_e(2\pi) - \frac{n}{2}\log_e|\Sigma| - \frac{1}{2}tr\Sigma^{-1}S - \frac{n}{2}tr\Sigma^{-1}(\overline{X}-\mu)(\overline{X}-\mu)' \qquad \ldots(2\text{-}2)
\end{aligned}$$

where

$\overline{X} = \frac{1}{n}XE_{n1}$ is the vector of sample means,

$E_{n1} = n \times 1$ matrix with all elements unity ($E_{1n} = E'_{n1}$) and

$S = X\left[I_n - \frac{1}{n}E_{n1}E_{1n}\right]X'$ is the matrix of corrected sum of squares and sum of products.

To find the maximum likelihood estimator for μ, we differentiate $\log_e L(\mu,\Sigma)$ in (2-2) with respect to μ, using $\left[\frac{\partial(X'AX)}{\partial X} = 2AX\right]$, and set the resulting expression equal to 0:

$$\frac{\partial \log_e L(\mu,\Sigma)}{\partial \mu} = \Sigma^{-1}(\overline{X}-\mu) = 0 \qquad \ldots(2\text{-}3)$$

which gives

$$\hat{\mu} = \bar{X} \qquad \text{... (2-4)}$$

Before differentiating $\log_e L(\mu, \Sigma)$ to find $\hat{\Sigma}$, we substitute $\hat{\mu} = \bar{X}$ in (2-2) and rewrite $\log_e|\Sigma|$ in terms of Σ^{-1} to obtain

$$\log_e L(\hat{\mu}, \Sigma) = \frac{-n}{2}\log_e(2\pi) + \frac{n}{2}\log_e\left|\Sigma^{-1}\right| - \frac{1}{2}tr\Sigma^{-1}S \qquad \text{...(2-5)}$$

Using

$$\frac{\partial tr(AB)}{\partial A} = B + B' - diag(B),$$

$$\frac{\partial \log_e|A|}{\partial A} = 2A^{-1} - diag(A^{-1})$$

and differentiate (2-5) with respect to Σ^{-1} we get

$$\frac{\partial \log_e L(\hat{\mu}, \Sigma)}{\partial \Sigma^{-1}} = n\Sigma - \frac{n}{2}diag(\Sigma) - S + \frac{1}{2}diag(S) = 0 \qquad \text{...(2-6)}$$

from which we have

$$\hat{\Sigma} - \frac{1}{2}diag(\hat{\Sigma}) = \frac{1}{n}[S - \frac{1}{2}diag(S)] \qquad \text{...(2-7)}$$

note that for the off-diagonal elements we have

$$\hat{\sigma}_{ij} = \frac{1}{n}s_{ij}, \quad i \neq j$$

and for the diagonal elements we have

$$\hat{\sigma}_{ii} - \frac{1}{2}\hat{\sigma}_{ii} = \frac{1}{n}(s_{ii} - \frac{1}{2}s_{ii})$$

or

$$\frac{1}{2}\hat{\sigma}_{ii} = \frac{1}{2n}s_{ii}$$

i.e

$$\hat{\sigma}_{ii} = \frac{1}{n}s_{ii}$$

thus, we have

$$\hat{\Sigma} = \frac{1}{n}S \qquad \text{...(2-8)}$$

3. MINIMUM DISTANCE ESTIMATION

Minimum distance estimation (MDE) has characterization and properties that could be found in Parr and Schucany (1980), Wolfwitz (1957). Hobbs, Moore, and James (1984) used MDE to find the location of the gamma distribution. Similarly, Hobbs, Moore, and Miller (1985) used MDE to estimate location of the Weibull, Sultan and Moore (1995) used MDE for the parameters of the two parameter logistic distribution.

MDE selects as estimates those p.d..f parameters which minimize the discrepancy between the sample data and the estimated distribution. The distance measures, which are minimized are 'Goodness of fit statistics'.

The MDE has the following characterization and properties:-

1. Not susceptible to outliers (Parr and Schucany, 1980).
2. Statistically consistent(Wolfwitz,1957).
3. Easily applied to all the parameters(Parr and Schucany 1980).

A series of logical candidates for the distance estimation task is studied by Fuchs(1984).

In this section it is required to estimate μ,Σ for the bivariate normal distribution function such that a goodness of fit statistic is minimized using a nonparametric estimator $\hat{f}(X)$. In this case $X = (x_1, x_2)'$ for the bivariate density f(X) given by:

$$f(X) = \frac{1}{(2\pi)|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\right\}$$

where

$$\mu = (\mu_1, \mu_2)'$$

and

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

as the mean vector and the covariance matrix respectively.

The kernel estimator is to be used with a Gaussian kernel which is defined as :

$$\hat{f}(X) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left\{\frac{1}{h}(X - X_i)\right\}$$

with K(X) given by

$$K(X) = (2\pi)^{-1}\exp\left(-\frac{1}{2}X'X\right)$$
$$= (2\pi)^{-1}\exp\left[-\frac{1}{2}\left(x_1^2 + x_2^2\right)\right]$$

which results the following expression for $\hat{f}(X)$:

$$\hat{f}(x_1,x_2) = \frac{1}{nh^2}\sum_{i=1}^{n}\exp\left\{\frac{1}{2h^2}\left[(x_1 - x_{1i})^2 + (x_2 - x_{2i})^2\right]\right\}$$

This $\hat{f}(x_1,x_2)$ defines the kernel density estimator with a Gaussian kernel. The c.d.f of this kernel density $\hat{F}(X)$ is given as:

$$\hat{F}(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{1}{nh^2}\sum_{i=1}^{n}\exp\left\{\frac{1}{2h^2}\left[(x_1 - x_{1i})^2 + (x_2 - x_{2i})^2\right]\right\}dx_1 dx_2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{1}{h^2}\exp\left\{\frac{1}{2h^2}\left[(x_1 - x_{1i})^2 + (x_2 - x_{2i})^2\right]\right\}dx_1 dx_2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{X - X_i}{h}\right)$$

where $\Phi(X)$ denotes the c.d.f for the standard bivariate normal random variable.

The Cramer von Mises statistic W^2 defined as:

$$CvM = n\int_{-\infty}^{\infty}\left[\hat{F}(X) - F(X)\right]^2 dF(X)$$

or the computational formula:

$$W^2 = \sum_{j=1}^{n}\left[F(X_j) - \frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{X - X_i}{h}\right)\right]^2 + \frac{1}{12n}$$

is to be used. The optimal value of the window width h (in the MISE sense) depends on the choice of the kernel K, the underlying unknown density f(X) and the sample size i.e

$$h_{opt} = f_1(K).f_2(f(X)).f_2(n)$$

An approximation for the optimal window width for a normal sample in case of univariate distribution was suggested to be ; where k is a real constant (Silverman 1986). Although this approximation simplifies the optimal expression for the window width and works fine with the normal distribution it is not good for other distributions (Sultan, 1995). This leads to the idea of introducing an approximate expression for h_{opt}.

An efficient alternative for computing the window width that gives an improvement in the sense of applying nonparametric density estimation in parameter estimation is the empirical choice of h equals $kSn^{\cdot}$ where S represents the sample standard deviation. In fact, the choice of the constant k that guarantees a close enough MISE to the theoretical optimal h needed a relatively extensive calculations. The choice of the h parameter is data dependent which is a function of both the sample standard deviation and the sample size.

The choice of the h parameter for the univariate case, can frequently be chosen visually in a satisfactory manner. However this is not exactly the same in the bivariate case. The behavior of different distributions under a proposed choice of the h parameter had been studied (Moore and Sultan in 1990).

4. THE MONTE CARLO EXPERIMENT

To evaluate the performance of the method a Monte Carlo experiment is designed. Deviates from a bivariate normal distribution with a given mean vector and a given variance covariance matrix are generated. The data based choice of the smoothing parameter is calculated for each sample of the Monte Carlo experiment. The integrated square error ISE given as:

$$ISE = \int \left[\hat{f}(X) - f(X) \right]^2 dX$$

is computed for each sample.

An estimate of the mean integrated square error MISE is obtained by averaging the ISE from the Monte Carlo repetitions. Likewise, an estimate of the standard deviation of MISE is computed.

The Monte Carlo procedure used here could be described in the following steps:

1. Different samples from the bivariate normal distribution with a given mean vector and a given variance covariance matrix for different sample sizes were generated. The bivariate normal deviates were generated using the RNMVN routine from IMSL.
2. The MLE estimators for μ and Σ were computed as discussed earlier.
3. The CvM statistics was computed for the estimated density with MLE estimators for the parameters.
4. The CvM statistics was minimized over the parameter space with μ_1, μ_2 and σ_{11}, σ_{12}, σ_{21}, σ_{22} as decision variables.
5. The new parameter estimates were compared with those of MLE, using the ISE as a measure for the comparison.

The following table (table 1) shows the MISE for both estimators (MLE and minimum distance estimator with CvM)

le size	5	10	15	20
No. of Monte Carlo	529	125	341	385
E (MLE)	1.04061E-3 *(2.23041E-3)*	3.46079E-4 *(4.50843E-4)*	2.04458E-4 *(2.12746E-4)*	1.89423E-4 *(8.200738E-5)*
MISE (CvM)	9.61347E-4 *(2.45467E-3)*	2.65217E-4 *(3.2176E-4)*	1.88698E-4 *(9.848806E-5)*	1.50644E-4 *(1.39581E-4)*

table 1. Results from M.C experiment for sample size 5(5)20

To show the effect of the variation of the covariance matrix on the MISE, a different experiment was run for a sample size of 20 and 4 values of $\Sigma(\Sigma_1,\cdots,\Sigma_4)$ The results of this experiment is shown in table 2 for the following values of the Σ matrix:

$$\Sigma_1=\begin{bmatrix}10 & -.375\\ -.375 & 20\end{bmatrix},\quad \Sigma_2=\begin{bmatrix}10 & 3.0\\ 3.0 & 5\end{bmatrix},\quad \Sigma_3=\begin{bmatrix}10 & -17\\ -17 & 200\end{bmatrix},\quad \Sigma_4=\begin{bmatrix}100 & 3.75\\ 3.75 & 18\end{bmatrix}$$

Bivar. Nor.	MISE *(CvM)*	MISE *(MLE)*
$N(\mu_0,\Sigma_1)$	3.91250E-3 *(5.98327E-3)*	2.41766E-2 *(1.81924E-3)*
$N(\mu_0,\Sigma_2)$	3.508693E-3 *(4.80272E-3)*	6.68145E-3 *(4.76252E-3)*
$N(\mu_0,\Sigma_3)$	3.14077E-3 *(3.76742E-3)*	4.342011E-3 *(3.043978E-3)*
$N(\mu_0,\Sigma_4)$	3.18853E-3 *(1.93254E-3)*	3.14943E-3 *(1.23398E-3)*

table 2. Results from M.C size 1000 for sample size 20

The table shows that both the MLE method and the new technique are statistically the same for covariance matrix Σ_4. However the new technique shows a significant improvement over the MLE method for $\Sigma_1,\Sigma_2,\Sigma_3$.

Together with the results from the previous table, the new technique shows an improvement over the MLE method. Different cases are shown with the integrated square error (ISE) for the nonparametric density estimation approach as well as for the

MLE. The parameters estimators for both cases are given. The variations in h together with the variation in MISE indicate that the method is an adaptive one in the sense that the choice of the parameter h which is data dependent varies with the variation of the distribution parameters and the sample size. Graphs for these cases are given in figures (fig1 - fig4), while table2 shows the resulting MISE together with its standard deviation for sample size 20 for the different parameter values for both the new proposed estimation technique concurrently with the MISE for the MLE.

The final conclusion is that the minimum distance estimation method using the CvM statistic as a measure of the difference between a nonparametric estimator based on a suggested window width and a parametric density with unknown parameters gives in general a smaller MISE value than the maximum likelihood method.

Bivariate Normal Sample (*Sample Size = 5*)

Bivariate Normal Data Points

x_1	x_2
-.150383	.494668
3.438728	-.381145
-.827254	-1.614013
-2.758851	2.203328
.141007	-.835404

TRUE PARAMETERS ARE

MEAN VECTOR = (5.000, 5.000)

COVARIANCE MATRIX :	10.000	-.375
	-.375	20.000

MLE PARAMETERS ARE

MEAN VECTOR = (.000, .000)

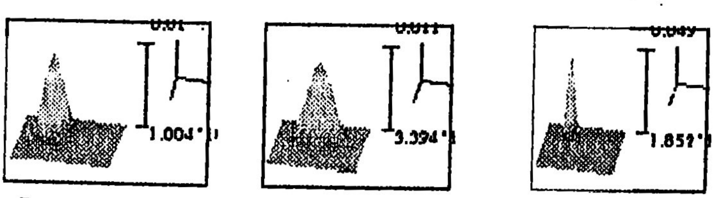
COVARIANCE MATRIX :	5.000	-1.563
	-1.563	2.100

MDCVM PARAMETERS ARE

MEAN VECTOR = (2.654, -.026)

COVARIANCE MATRIX :	30.020	-.990
	-.990	7.818

ISE MLE	.0065
ISE MDCVM	.0002



C O M

O....True density M....MLE C....MDE CvM

Fig 1

Bivariate Normal Sample (*Sample Size = 10*)

Bivariate Normal Data Points

x_1	x_2	x_1	x_2
4.85	10.65	5.35	5.42
8.44	8.27	4.82	6.65
4.17	8.16	3.84	7.45
2.24	6.46	6.49	6.71
5.14	13.48	4.41	2.00

TRUE PARAMETERS ARE

MEAN VECTOR = (5.000, 5.000)

COVARIANCE MATRIX :	10.000	-.375
	-.375	20.000

MLE PARAMETERS ARE

MEAN VECTOR = (.000, 2.500)

COVARIANCE MATRIX :	2.700	0.768
	0.768	9.300

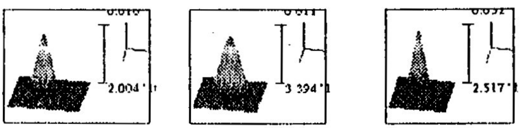
MDCVM PARAMETERS ARE

MEAN VECTOR = (3.071, 2.526)

COVARIANCE MATRIX :	10.494	0.990
	0.990	9.242

ISE MLE .0023

ISE MDCVM .0002



C O M

O....True density M....MLE C....MDE CvM

Fig 2

Bivariate Normal Sample (***Sample Size* = 15**)

Bivariate Normal Data Points

x_1	x_2	x_1	x_2	x_1	x_2
4.85	5.44	5.35	5.79	8.99	1.10
8.44	6.51	4.82	6.12	7.41	4.67
4.17	7.44	3.84	4.33	7.21	5.31
2.24	6.87	6.49	-.01	5.96	-.63
5.14	1.97	4.41	7.47	11.01	1.93

TRUE PARAMETERS ARE

MEAN VECTOR = (5.000, 5.000)

COVARIANCE MATRIX :	10.000	-.375
	-.375	20.000

MLE PARAMETERS ARE

MEAN VECTOR = (1.000, -.700)

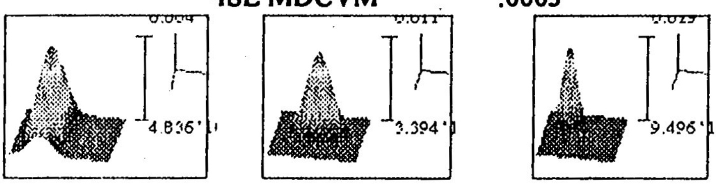
COVARIANCE MATRIX :	5.200	-2.795
	-2.795	7.400

MDCVM PARAMETERS ARE

MEAN VECTOR = (2.014, -.712)

COVARIANCE MATRIX :	117.482	-.990
	-.990	11.316

ISE MLE	.0013
ISE MDCVM	.0003



C O M

O....True density M....MLE C....MDE CvM

Fig 3

Bivariate Normal Sample (*Sample Size = 20*)

Bivariate Normal Data Points

x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
4.85	5.81	5.35	1.24	8.99	2.89	5.31	1.76
8.44	5.99	4.82	4.76	7.41	3.65	6.16	3.92
4.17	4.32	3.84	5.43	7.21	.26	6.71	7.27
2.24	.15	6.49	-.65	5.96	10.97	6.25	-4.42
5.14	7.45	4.41	2.18	11.01	4.38	2.86	1.97

TRUE PARAMETERS ARE

MEAN VECTOR = (5.000, 5.000)

COVARIANCE MATRIX :	10.000	-.375
	-.375	20.000

MLE PARAMETERS ARE

MEAN VECTOR = (0.9, -1.500)

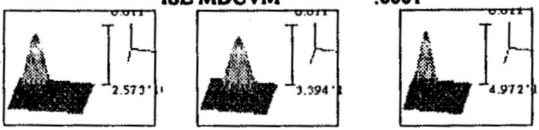
COVARIANCE MATRIX :	4.400	0.642
	0.642	11.500

MDCVM PARAMETERS ARE

MEAN VECTOR = (3.484, -1.533)

COVARIANCE MATRIX :	18.353	0.990
	0.990	11.375

ISE MLE	.0009
ISE MDCVM	.0001



C O M

O....True density M....MLE C....MDE CvM

Fig 4

Bibliography

[1] Anderson. T.W. *An introduction to multivariate statistical analysis*. New York: Wiley (1984).

[2] Bowman, A. W. A comparative study of some kernel based nonparametric density estimators. *Journal of Stat. Simul. 21*: 313-327 (1985).

[3] Devroye. Luc *A course in density estimation*. Birkhauser (1987).

[4] Devroye. Luc *Nonparametric density estimation: the L1 view*. New York: Wiley (1992).

[5] Es. A. J. Van Aspects of non parametric density estimation *Centrum voor wiskund en Informatica*. (1991).

[6] Hearne. Leonard B *Probability density estimation on a high dimensional space using random tessellations*. Ph.D. dissertation (1993).

[7] Hobbs. J. R. ; Moore. A. H. ; and James. W. Minimum distance estimation of the three parameter gamma distribution. *IEEE Transactions on Reliability 33 NO.3*: 237-240(1984).

[8] Hobbs. J. R. : Moore. A. H. ; and Miller. R. M. Minimum distance estimation of the three parameter Weibull distribution. *IEEE Transactions on Reliability 34 NO.5*: 495-496(1985).

[9] Parr. W.C. and Schucany. W.R. Minimum distance and robust estimation. *JASA 75 NO 3* :616-624 (1980).

[10] Revesz. P. Density estimation. *Handbook of Statistics vol. 4* :531-549 (1984).

[11] Scott. David W. *Multivariate density estimation: Theory, practice, and visualization*. New York. Wiley (1992).

[12] Silverman. B. W. *Density estimation for statistics and data analysis*. Chapman and Hall ..... (1986)

[13] Sultan. A. M. *Applications of non-parametric density estimation* Ph.D. dissertation. Ohio A.U. 1990.

[14] Sultan. A. M .; Moore. A. H. A comparative study of parameter estimation for the logistic distribution. *INFORMS LA Spring National Meeting* (April 23-26. 1995).

[15] Wahba, Grace Optimal smoothing of density estimates. *Annals of Math. Stat.*:423-452 (1983).

[16] Wegman, E. J. Density estimation. *Encyclopedia of Stat Science vol 2*:309-315 (1982).

[17] Wetz, Wolfgang *Statistical density estimation: A survey*. Van denhoeck and ruprecht (1978)

[18] Wolfwitz, J. The minimum distance method. *Annals of Math. Stat. 28*:75-88 (1957).