ESTIMATION FROM CENSORED BIVARIATE SAMPLES

By M. N. Bhattacharyya

Defence Science Laboratory, New Delhi

1. Introduction

The problem of estimation from censored samples has been treated by many authors. A. K. Gupta[1] studied the case of estimating the mean and standard deviation of a normal population from a sample of size n out of which $(n-k)$ greatest or smallest observations are censored. Cohen[2] considered the estimation of the mean and standard deviation of a univariate normal population from singly and doubly censored samples. In singly censored sample, the observations below or above a given point are censored while for a doubly censored sample the observations above the upper and below the lower truncation points are not available. He[3] has also considered the estimation of the parameters of a bivariate population from a sample, either singly or doubly censored, and has recently extended his studies to the multivariate case, considering truncation on one of the variates.

But in practice occasionally we come across with bivariate samples which are doubly censored on both the variates. For example in a firing trial on a finite target screen, the rounds may be missing on any side of the screen. In this case the estimation of the mean point of impact and dispersions, both horizontal and vertical, is equivalent to estimating the means and standard deviations of a bivariate normal population from a sample doubly censored on both the variates.

In this paper we shall derive the maximum likelihood equations for estimating the means and standard deviations of a bivariate population from a sample doubly censored on both the variates, which are assumed to be uncorrelated, *i.e.*, $\rho = 0$. The information matrix has also been given. The solutions of the equations have been obtained by iterative process using the normal curve tables for the ordinates and areas.

2. Maximum Likelihood Estimates

Let a, b and c, d be the left and right truncation points on the X Y-axis respectively and R_1 and R_2 be the respective distances between runcation points. Let N be the sample size of which n_0 pairs nd Y $(a \leqslant X \leqslant b;\ c \leqslant Y \leqslant d)$ are measured, *i.e.*, $(N - n_0)$ observa- e outside the above limits. We change the origin to the left terminal points for both the variates by changing the variables to $x = X - a$ and $y = Y - c$. Let the truncation points on the x and y-axes be designated ξ', ξ'' and η', η'' respectively in standardized units, *i.e.*,

$$\xi' = \frac{a - m_1}{\sigma_1}, \quad \xi'' = \frac{b - m_1}{\sigma_1} \tag{1}$$

$$\eta' = \frac{c - m_2}{\sigma_2}, \quad \eta'' = \frac{d - m_2}{\sigma_2}$$

where m_1, m_2 and σ_1, σ_2 are the mean and standard deviation for the variates X and Y respectively.

The probability density function for such a normal bivariate sample is

$$L = \frac{N!}{n_0!\,(N - n_0)!}\,(1 - A)^{N-n_0}\left(\frac{1}{2\pi\sigma_1\sigma_2}\right)^{n_0}$$

$$\exp\left[-\tfrac{1}{2}\sum^{n_0}\left\{\left(\xi' + \frac{x}{\sigma_1}\right)^2 + \left(\eta' + \frac{y}{\sigma_2}\right)^2\right\}\right] \tag{2}$$

where

$$A = \text{Prob.}\ (a \leqslant X \leqslant b;\ \ c \leqslant Y \leqslant d)$$

$$= \int_{\xi'}^{\xi''} \phi(t)\,dt \times \int_{\eta'}^{\eta''} \phi(t)\,dt\ ;$$

$$\phi(t) \text{ being equal to } \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}$$

Putting

$$I' = \int_{\xi'}^{\infty} \phi(t)\,dt;\quad I'' = \int_{\xi''}^{\infty} \phi(t)\,dt$$

$$J' = \int_{\eta'}^{\infty} \phi(t)\,dt;\quad J'' = \int_{\eta''}^{\infty} \phi(t)\,dt$$

so,

$$A = (I' - I'') \times (J' - J'') \tag{3}$$

Since R_1 and R_2 are the censored ranges of the two variates, we have from (1)

$$\xi'' = \xi' + \frac{R_1}{\sigma_1}$$

and

$$\eta'' = \eta' + \frac{R_2}{\sigma_2}$$

We shall first obtain the following partial derivatives which will be utilised in subsequent derivations.

$$\frac{\partial I'}{\partial \xi'} = -\phi' \qquad \frac{\partial \phi'}{\partial \xi'} = -\phi'\xi'$$

$$\frac{\partial I''}{\partial \xi'} = -\phi'' \qquad \frac{\partial \phi''}{\partial \xi'} = -\phi''\xi''$$

$$\frac{\partial I'}{\partial \sigma_1} = 0 \qquad \frac{\partial \phi'}{\partial \sigma_1} = 0 \tag{5}$$

$$\frac{\partial I''}{\partial \sigma_1} = \phi'' \frac{R_1}{\sigma_1^2} \qquad \frac{\partial \phi''}{\partial \sigma_1} = \phi''\xi'' \frac{R_1}{\sigma_1^2}$$

$$\frac{\partial J'}{\partial \eta'} = -\psi' \qquad \frac{\partial \psi'}{\partial \eta'} = -\psi'\eta'$$

$$\frac{\partial J''}{\partial \eta'} = -\psi'' \qquad \frac{\partial \psi''}{\partial \eta'} = -\psi''\eta'' \tag{6}$$

$$\frac{\partial J'}{\partial \sigma_2} = 0 \qquad \frac{\partial \psi'}{\partial \sigma_2} = 0$$

$$\frac{\partial J''}{\partial \sigma_2} = \psi'' \frac{R_2}{\sigma_2^2} \qquad \frac{\partial \psi''}{\partial \sigma_2} = \psi''\eta'' \frac{R_2}{\sigma_2^2}$$

where

$$\phi' = \frac{1}{\sqrt{2\pi}} e^{-\xi'^2/2} \quad ; \quad \phi'' = \frac{1}{\sqrt{2\pi}} e^{-\xi''^2/2}$$

$$\psi' = \frac{1}{\sqrt{2\pi}} e^{-\eta'^2/2} \quad ; \quad \psi'' = \frac{1}{\sqrt{2\pi}} e^{-\eta''^2/2}$$

Logarithm of the likelihood function is

$$\log L = \log \frac{N!}{n_0!\,(N-n_0)!} + (N-n_0)\log(1-A)$$

$$- n_0 \log(2\pi\sigma_1\sigma_2) - \tfrac{1}{2}\sum^{n0} \left\{\left(\xi' + \frac{x}{\sigma_1}\right)^2 + \left(\eta' + \frac{y}{\sigma_2}\right)^2\right\} \tag{7}$$

Differentiating (7) with the assistance of (5) and (6) and equating to zero we obtain the maximum likelihood estimating equations

$$\frac{\partial \log L}{\partial \xi'} = (N-n_0)\frac{(\phi'-\phi'')(J'-J'')}{(1-A)} - \Sigma\left(\xi' + \frac{x}{\sigma_1}\right) = 0$$

$$\frac{\partial \log L}{\partial \sigma_1} = (N - n_0) \frac{\phi''(J' - J'')}{(1 - A)} \frac{R_1}{\sigma_1^2} - \frac{n_0}{\sigma_1}$$

$$+ \frac{1}{\sigma_1^2} \sum \left\{ \left(\xi' + \frac{x}{\sigma_1} \right) x \right\} = 0 \tag{8}$$

$$\frac{\partial \log L}{\partial \eta'} = (N - n_0) \frac{(\psi' - \psi'')(I' - I'')}{(1 - A)} - \Sigma \left(\eta' + \frac{y}{\sigma_2} \right) = 0$$

$$\frac{\partial \log L}{\partial \sigma_2} = (N - n_0) \frac{\psi''(I' - I'')}{(1 - A)} \frac{R_2}{\sigma_2^2} - \frac{n_0}{\sigma_2}$$

$$+ \frac{1}{\sigma_2^2} \Sigma \left\{ \left(\eta' + \frac{y}{\sigma_2} \right) y \right\} = 0$$

Let

$$Q_1 = \left(\frac{N - n_0}{n_0} \right) \frac{\phi'(J' - J'')}{(1 - A)},$$

$$Q_2 = \left(\frac{N - n_0}{n_0} \right) \frac{\phi''(J' - J'')}{(1 - A)}, \tag{9}$$

$$Q_3 = \left(\frac{N - n_0}{n_0} \right) \frac{\psi'(I' - I'')}{(1 - A)},$$

and

$$Q_4 = \left(\frac{N - n_0}{n_0} \right) \frac{\psi''(I' - I'')}{(1 - A)}.$$

then equation's (8) can be written as

$$\sigma_1 [Q_1 - Q_2 - \xi'] - v_1 = 0 \tag{10.1}$$

$$\sigma_1^2 \left[1 - \xi'(Q_1 - Q_2 - \xi') - \frac{R_1 Q_2}{\sigma_1} \right] - v_2 = 0 \tag{10.2}$$

$$\sigma_2 [Q_3 - Q_4 - \eta'] - \gamma_1 = 0 \tag{10.3}$$

$$\sigma_2^2 \left[1 - \eta'(Q_3 - Q_4 - \eta') - \frac{R_2 Q_4}{\sigma_2} \right] - \gamma_2 = 0 \tag{10.4}$$

where

$$v_i = \sum^{n_0} \frac{x^i}{n_0}$$

and

$$\gamma_i = \sum^{n_0} \frac{y^i}{n_0}$$

Since Q's are functions of ξ', η', σ_1 and σ_2 their values for any given values of these arguments can be noted from standard tables for ordinates and areas of a univariate normal curve. The equations (10) are to be solved for getting the estimates of ξ', σ_1, η', and σ_2. From these estimates m_1 and m_2 can be estimated from the relations given in (I).

3. Numerical Example

The equations (10) can be solved numerically either by the modified Newton-Raphson method or by the iterative process. In the present case the modified Newton-Raphson method is quite cumbersome because for each approximation, after calculating the values of partial derivatives of Q's with respect to ξ', ξ'', η' and η'', a 4×4 determinant will have to be evaluated whereas for the iterative process the values of Q's only are required. Hence for the simplicity of computational work the iterative process was followed, though the convergence of this process may be slow.

The iterative equations derived from (10) are as follows:—

$$\xi' = F_1(\xi', \xi'', \eta', \eta'') = \frac{Q_1 - Q_2 - \frac{\nu_1}{R_1}\xi''}{\left(1 - \frac{\nu_1}{R_1}\right)}$$

$$\xi'' = F_2(\xi', \xi'', \eta', \eta'') = \frac{\left(\frac{R_1}{\xi'' - \xi'}\right) - \xi'\left(\nu_1 - \frac{\nu_2}{R_1}\right) - Q_2 R_1}{\frac{\nu_2}{R_1}}$$

$$\eta' = F_3(\xi', \xi'', \eta', \eta'') = \frac{Q_3 - Q_4 - \frac{\gamma_1}{R_2}\eta''}{\left(1 - \frac{\gamma_1}{R_2}\right)} \tag{11}$$

$$\eta'' = F_4(\xi', \xi'', \eta', \eta'') = \frac{\left(\frac{R_2}{\eta'' - \eta'}\right) - \eta'\left(\gamma_1 - \frac{\gamma_2}{R_2}\right) - Q_4 R_2}{\frac{\gamma_2}{R_2}}$$

The above equations will be convergent provided the sum of the absolute values of the partial derivatives of F_1, F_2, F_3 and F_4 with respect to any of the variables ξ', ξ'', η' and η'' is less than unity.

A random sample of size fifty from an uncorrelated bivariate population with $m_1 = 10$, $\sigma_1 = 1$, $m_2 = 12$ and $\sigma_2 = 1\cdot 5$ was taken with the help of Prof. Mahalanobis's tables.[4] Subsequently taking $a = 8$, $b = 12 \cdot 25$, $c = 9$, and $d = 16$, forty-two pairs of x and y were retained. The first and second moments for x and y, referred to the left truncation points, were calculated.

Thus for the sample selected

$$N = 50, \quad \nu_1 = 2 \cdot 00810, \quad \gamma_1 = 3 \cdot 18888$$

$$n_0 = 42, \quad \nu_2 = 4 \cdot 89544, \quad \gamma_2 = 11 \cdot 67827$$

$$R_1 = 4 \cdot 25 \quad R_2 = 7$$

The initial set of approximations obtained from the relation

$$(1 - 2p)^2 = \frac{n_0}{N}, \tag{12}$$

where

$$p = \int_{-\infty}^{\xi'} \phi(t)\, dt = \int_{\xi''}^{\infty} \phi(t)\, dt = \int_{-\infty}^{\eta'} \phi(t)\, dt$$

$$= \int_{\eta''}^{\infty} \phi(t)\, dt,$$

is as follows:—

$$\xi' = -1 \cdot 73079, \quad \xi'' = 1 \cdot 73079$$

$$\eta' = -1 \cdot 73079, \quad \eta'' = 1 \cdot 73079 \tag{13}$$

The convergence of the iterative equations (11) can be seen from the following relations:—

$$\left|\frac{\partial F_1}{\partial \xi'}\right| + \left|\frac{\partial F_2}{\partial \xi'}\right| + \left|\frac{\partial F_3}{\partial \xi'}\right| + \left|\frac{\partial F_4}{\partial \xi'}\right| = \cdot 81966$$

$$\left|\frac{\partial F_1}{\partial \xi''}\right| + \left|\frac{\partial F_2}{\partial \xi''}\right| + \left|\frac{\partial F_3}{\partial \xi''}\right| + \left|\frac{\partial F_4}{\partial \xi''}\right| = \cdot 95490 \tag{14}$$

$$\left|\frac{\partial F_1}{\partial \eta'}\right| + \left|\frac{\partial F_2}{\partial \eta'}\right| + \left|\frac{\partial F_3}{\partial \eta'}\right| + \left|\frac{\partial F_4}{\partial \eta'}\right| = \cdot 88039$$

$$\left|\frac{\partial F_1}{\partial \eta''}\right| + \left|\frac{\partial F_2}{\partial \eta''}\right| + \left|\frac{\partial F_3}{\partial \eta''}\right| + \left|\frac{\partial F_4}{\partial \eta''}\right| = \cdot 89366$$

Starting with the initial set of approximations (13) and using the iterative equations (11) ten successive approximations have been calculated and given in Table I. To examine the accuracy of the estimates the left-hand side of equations (10) have been evaluated for each of these approximations and have also been shown in the table.

TABLE I

Approximations					Value of left-hand side of the estimating equations			
No.	ξ'	ξ''	η'	η''	(10·1)	(10·2)	(10·3)	(10·4)
1	−1·55026	1·99344	−1·44818	2·38132	−·063	−·147	−·309	−1·041
2	−1·64962	1·97195	−1·75812	2·30774	·008	−·045	−·017	·393
3	−1·63697	1·92735	−1·77632	2·46269	·021	·004	−·002	−·224
4	−1·60403	1·90084	−1·88943	2·48977	·024	−·015	−·052	·043
5	−1·56619	1·85639	−1·94947	2·56504	·025	·015	−·064	−·066
6	−1·52733	1·83236	−2·02577	2·61488	·021	·008	−·051	−·011
7	−1·49540	1·80938	−2·08743	2·67234	·018	·010	−·049	−·039
8	−1·46852	1·79260	−2·14907	2·71832	·015	·009	−·040	−·014
9	−1·44720	1·78017	−2·19998	2·76395	·011	·008	−·036	−·025
10	−1·43095	1·77148	−2·24693	2·80114	·008	·007	−·029	−·011

The estimates of m_1, m_2, σ_1 and σ_2 obtained from relations (1) are

$$\hat{\sigma}_1 = 1\cdot32712; \quad \hat{m}_1 = 9\cdot89900$$

$$\hat{\sigma}_2 = 1\cdot38667; \quad \hat{m}_2 = 12\cdot11575$$

4. ASYMPTOTIC VARIANCES AND COVARIANCES OF THE ESTIMATES

The second partial derivatives of the logarithm of the likelihood function are given below:—

$$\frac{\partial^2 \log L}{\partial \xi'^2} = n_0 \, h_{11}(\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial \xi' \, \partial \sigma_1} = \frac{n_0}{\sigma_1} h_{12}(\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial \xi' \, \partial \eta'} = n_0 \, h_{13}(\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial \xi' \partial \sigma_2} = \frac{n_0}{\sigma_2} h_{14} (\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial {\sigma_1}^2} = \frac{n_0}{{\sigma_1}^2} h_{22} (\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial \sigma_1 \partial \eta'} = \frac{n_0}{\sigma_1} h_{23} (\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial \sigma_1 \partial \sigma_2} = n_0 h_{24} (\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial \eta'^2} = n_0 h_{33} (\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial \eta' \partial \sigma_2} = \frac{n_0}{\sigma_2} h_{34} (\xi', \xi'', \eta', \eta'')$$

$$\frac{\partial^2 \log L}{\partial {\sigma_2}^2} = \frac{n_0}{{\sigma_2}^2} h_{44} (\xi', \xi'', \eta', \eta'')$$

where

$$h_{11} = -\left[1 + Q_1\xi' - Q_2\xi'' + \frac{n_0}{N - n_0} (Q_1 - Q_2)^2\right]$$

$$h_{12} = \left[(Q_1 - Q_2 - \xi') + \frac{Q_2 R_1}{\sigma_1}\left\{\left(\frac{n_0}{N - n_0}\right) (Q_2 - Q_1) - \xi''\right\}\right]$$

$$h_{13} = \left[\frac{(Q_1 - Q_2)\ (Q_4 - Q_3)}{A} \left(\frac{n_0}{N - n_0}\right)\right].$$

$$h_{14} = -\left[\frac{R_2}{{\sigma_2}^2} \left(\frac{n_0}{N - n_0}\right) \frac{(Q_1 - Q_2)}{A}\ Q_4\right]$$

$$h_{22} = \left\{\left(\frac{R_1}{\sigma_1}\right)^2 Q_2 \left(\xi'' - \frac{n_0}{N - n_0}\ Q_2\right)\right\}$$

$$- \left\{2 - \xi' (Q_1 - Q_2 - \xi') - \frac{Q_2 R_1}{\sigma_1}\right\}$$

$$h_{23} = \left[\frac{R_1}{{\sigma_1}^2} \left\{\frac{Q_2 (Q_4 - Q_3)}{A} \left(\frac{n_0}{N - n_0}\right)\right\}\right]$$

$$h_{24} = -\left[\frac{Q_2\ Q_4}{A} \left(\frac{n_0}{N - n_0}\right) \left(\frac{R_1\ R_2}{{\sigma_1}^2\ {\sigma_2}^2}\right)\right]$$

$$h_{33} = -\left[1 + \eta' Q_3 - \eta'' Q_4 + \left(\frac{n_0}{N - n_0}\right)(Q_3 - Q_4)^2\right]$$

$$h_{34} = \left[\frac{R_2}{\sigma_2} Q_4 \left\{\left(\frac{n_0}{N - n_0}\right)(Q_4 - Q_3) - \eta''\right\}\right.$$

$$\left. + (Q_3 - Q_4 - \eta')\right]$$

$$h_{44} = \left[\left(\frac{R_2}{\sigma_2}\right)^2 Q_4 \left(\eta'' - \frac{n_0}{N - n_0} Q_4\right)\right.$$

$$\left. - \left\{2 - \eta'(Q_3 - Q_4 - \eta') - \frac{Q_4 R_2}{\sigma_2}\right\}\right]$$

The information matrix can be written as

$$\begin{pmatrix} -n_0 h_{11} & -\frac{n_0}{\sigma_1} h_{12} & -n_0 h_{13} & -\frac{n_0}{\sigma_2} h_{14} \\ -\frac{n_0}{\sigma_1} h_{12} & -\frac{n_0}{\sigma_1{}^2} h_{22} & -\frac{n_0}{\sigma_1} h_{23} & -n_0 h_{24} \\ -n_0 h_{13} & -\frac{n_0}{\sigma_1} h_{23} & -n_0 h_{33} & -\frac{n_0}{\sigma_2} h_{34} \\ -\frac{n_0}{\sigma_2} h_{14} & -n_0 h_{24} & -\frac{n_0}{\sigma_2} h_{34} & -\frac{n_0}{\sigma_2{}^2} h_{44} \end{pmatrix}$$

The inverse of this matrix will give the variance-covariance matrix of the estimates $\hat{\xi}'$, $\hat{\sigma}_1$, $\hat{\eta}'$ and $\hat{\sigma}_2$.

5. Summary

Maximum likelihood equations for estimating the parameters of a bivariate normal population ($\rho = 0$) from a sample having known truncation points on both the variates, have been derived. It has been shown that these equations can be solved by iterative process using the tables of ordinates and areas of normal curve. Information matrix has also been given from which the asymptotic variances and covariances can be obtained. Practical application of the results has been illustrated by a numerical example.

6. Acknowledgement

The author's thanks are due to Dr. P. V. K. Iyer, Senior Scientist, Defence Science Laboratory, New Delhi, for his valuable guidance in preparing this paper.

References

1. Gupta, A. K. .. "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, 1952, **39**, 260.
2. Cohen, A. C. .. "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Ann. Math. Stat.*, 1950, **21**, 557.
3. ——— .. "Estimation in truncated bivariate normal distributions" (Preliminary report), *ibid.*, 1953, **24**, 497.
4. Mahalanobis, P. C. .. "Tables of random samples from a normal population," *Sankhya*, 1934, **1**, 289.
5. Scarborough, J. B. .. *Numerical Mathematical Analysis*, The Johns Hopkins Press, Baltimore, 1950, 206.