Moments of a Truncated Bivariate Normal Distribution

By S. ROSENBAUM

Army Medical Department, Stanmore, Middlesex

[Received November 1960]

SUMMARY

The moments are obtained for a bivariate normal distribution which is singly truncated with respect to both variables; the variables may be correlated. From these moments the parameters of the distribution can be estimated.

1. INTRODUCTION

THE object is to obtain estimates of the parameters of a distribution from truncated observations. The method of moments was used in the univariate case by K. Pearson, and tables for the solution of particular problems given in *Tables for Statisticians and Biometricians*, Part I (1930). The method of maximum likelihood has been used to the same end, and Cohen (1959), in the latest of a series of papers devoted to this problem, has produced a particularly simple form of computation for the estimates. The same author (1955) found a maximum likelihood solution for the bivariate case where the truncation is with respect to one only of the variables. Singh (1960) has recently published a maximum likelihood solution of the bivariate case where the truncation is with respect to both variables, but the variables are uncorrelated.

In this paper we revert to the method of moments for the bivariate case, singly truncated with respect to both variables, which may be correlated. Cohen (1957) has stated that the method of moments gives the same solution as the method of maximum likelihood.

2. BIVARIATE MOMENTS

2.1. *First Moments*

In the standardized form, the first moment about the origin of the x variable is

$$m_{10} = \frac{1}{2\pi\sqrt{(1-\rho^2)}\,L(h,k;\,\rho)} \int_k^\infty \int_h^\infty x \exp\{-\tfrac{1}{2}(x^2-2\rho xy+y^2)/(1-\rho^2)\}\,dx\,dy$$

where $x = h, y = k$ are the truncation points, and L is the total probability in the truncated distribution.

After the transformation $x-\rho y = z\sqrt{(1-\rho^2)}$, we have, with $u = (h-\rho y)/\sqrt{(1-\rho^2)}$,

$$\begin{aligned} L(h,k;\,\rho)\,m_{10} &= \frac{1}{2\pi}\int_k^\infty \int_u^\infty \{z\sqrt{(1-\rho^2)}+\rho y\} \exp\{-\tfrac{1}{2}(y^2+z^2)\}\,dz\,dy \\ &= \frac{\sqrt{(1-\rho^2)}}{2\pi}\int_k^\infty \exp\{-\tfrac{1}{2}(h^2-2\rho hy+y^2)/(1-\rho^2)\}\,dy \\ &\qquad + \frac{\rho}{2\pi}\int_k^\infty y\,e^{-\frac{1}{2}y^2}\int_u^\infty e^{-\frac{1}{2}z^2}\,dz\,dy. \end{aligned}$$

Integration of the second expression by parts yields a term like the first expression times $\rho^2/(1-\rho^2)$, and another term

$$\frac{\rho}{2\pi}e^{-\frac{1}{2}k^2}\int_{(h-\rho k)/\sqrt{(1-\rho^2)}}^{\infty} e^{-\frac{1}{2}z^2}\,dz = \rho Z(k)\,Q\left\{\frac{h-\rho k}{\sqrt{(1-\rho^2)}}\right\},$$

where Z and Q are the standard frequency and distribution functions of the univariate normal distribution. The first expression can also be put into this form by the transformation

$$z=(y-h\rho)/\sqrt{(1-\rho^2)}$$

and we finally obtain

$$L(h,k;\,\rho)\,m_{10} = Z(h)\,Q\left\{\frac{k-\rho h}{\sqrt{(1-\rho^2)}}\right\}+\rho Z(k)\,Q\left\{\frac{h-\rho k}{\sqrt{(1-\rho^2)}}\right\}. \tag{1}$$

By symmetry, the other first moment is given by

$$L(h,k;\,\rho)\,m_{01} = \rho Z(h)\,Q\left\{\frac{k-\rho h}{\sqrt{(1-\rho^2)}}\right\}+Z(k)\,Q\left\{\frac{h-\rho k}{\sqrt{(1-\rho^2)}}\right\}. \tag{2}$$

2.2. *Second Moments*

The second moment about the origin of the x variable is obtained by using the same transformation as for the first moment, giving

$$L(h,k;\,\rho)\,m_{20} = \frac{1}{2\pi}\int_k^{\infty}\int_u^{\infty}\{z\sqrt{(1-\rho^2)}+\rho y\}^2\exp\{-\tfrac{1}{2}(y^2+z^2)\}\,dz\,dy.$$

Again, after integration by parts, this becomes

$$\frac{1}{2\pi}\int_k^{\infty}e^{-\frac{1}{2}y^2}\left[(1-\rho^2+\rho^2y^2)\int_u^{\infty}e^{-\frac{1}{2}z^2}\,dz+\sqrt{(1-\rho^2)}\,(h+\rho y)\exp\{-\tfrac{1}{2}(h-\rho y)^2/(1-\rho^2)\}\right]dy$$

and, eventually, we have that

$$L(h,k;\,\rho)\,m_{20} = L(h,k;\,\rho)+hZ(h)\,Q\left\{\frac{k-\rho h}{\sqrt{(1-\rho^2)}}\right\}+\rho^2kZ(k)\,Q\left\{\frac{h-\rho k}{\sqrt{(1-\rho^2)}}\right\}$$
$$+\frac{\rho\sqrt{(1-\rho^2)}}{\sqrt{(2\pi)}}Z\left\{\frac{\sqrt{(h^2-2\rho hk+k^2)}}{\sqrt{(1-\rho^2)}}\right\}. \tag{3}$$

By symmetry, the other second moment is given by

$$L(h,k;\,\rho)\,m_{02} = L(h,k;\,\rho)+\rho^2hZ(h)\,Q\left\{\frac{k-\rho h}{\sqrt{(1-\rho^2)}}\right\}+kZ(k)\,Q\left\{\frac{h-\rho k}{\sqrt{(1-\rho^2)}}\right\}$$
$$+\frac{\rho\sqrt{(1-\rho^2)}}{\sqrt{(2\pi)}}Z\left\{\frac{\sqrt{(h^2-2\rho hk+k^2)}}{\sqrt{(1-\rho^2)}}\right\}. \tag{4}$$

The same transformation, and integration by parts, yields the product moment, which is given by

$$L(h,k;\,\rho)\,m_{11} = \rho L(h,k;\,\rho)+\rho hZ(h)\,Q\left\{\frac{k-\rho h}{\sqrt{(1-\rho^2)}}\right\}+\rho kZ(k)\,Q\left\{\frac{h-\rho k}{\sqrt{(1-\rho^2)}}\right\}$$
$$+\frac{\sqrt{(1-\rho^2)}}{\sqrt{(2\pi)}}Z\left\{\frac{\sqrt{(h^2-2\rho hk+k^2)}}{\sqrt{(1-\rho^2)}}\right\}. \tag{5}$$

3. Estimating the Parameters

The above moments are in standardized form, so that if x, y are the observed quantities and come from distributions with means and standard deviations (μ_1, σ_1), (μ_2, σ_2) respectively, the following relationships also hold:

$$\hat{m}_{10} = (\bar{x}-\mu_1)/\sigma_1, \quad \hat{m}_{20} = \Sigma(x-\mu_1)^2/(n\sigma_1^2),$$

$$\hat{m}_{01} = (\bar{y}-\mu_2)/\sigma_2, \quad \hat{m}_{02} = \Sigma(y-\mu_2)^2/(n\sigma_2^2),$$

$$\hat{m}_{11} = \Sigma(x-\mu_1)(y-\mu_2)/(n\sigma_1\sigma_2).$$

Therefore

$$\hat{\sigma}_1 = \frac{\sqrt{(\Sigma x^2/n-\bar{x}^2)}}{\sqrt{(\hat{m}_{20}-\hat{m}_{10}^2)}}, \tag{6}$$

$$\hat{\mu}_1 = \bar{x}-\hat{m}_{10}\,\hat{\sigma}_1, \tag{7}$$

$$\hat{\sigma}_2 = \frac{\sqrt{(\Sigma y^2/n-\bar{y}^2)}}{\sqrt{(\hat{m}_{02}-\hat{m}_{01}^2)}}, \tag{8}$$

$$\hat{\mu}_2 = \bar{y}-\hat{m}_{01}\,\hat{\sigma}_2. \tag{9}$$

Also

$$\hat{m}_{11} = (\Sigma xy/n-\hat{\mu}_1\bar{y}-\hat{\mu}_2\bar{x}+\hat{\mu}_1\hat{\mu}_2)/(\hat{\sigma}_1\hat{\sigma}_2). \tag{10}$$

The required estimates are obtained by an iterative solution of equations (1) to (10). The recommended procedure is as follows. Approximate values of $\hat{h}, \hat{k}, \hat{\rho}$ are put into equations (1) to (4), where Z and Q can be found from standard tables of the univariate normal distribution, and L from the graphs of Zelen and Severo (1960) or elsewhere. These graphs require the calculation of $k-\rho h, h-\rho k$ and $\sqrt{(h^2-2\rho hk+k^2)}$ which enter into the above equations.

The values of $\hat{m}_{10}, \hat{m}_{01}, \hat{m}_{20}$ and $\hat{m}_{02}$ thus obtained are used in equations (6) to (9) to give the first estimates of the means and standard deviations. From these, new values for $\hat{h}$ and $\hat{k}$ can be derived.

The next step is to obtain an estimate for $\hat{\rho}$ from the following quadratic equation, which is derived from equations (1) to (5):

$$(\hat{h}+\hat{k})\hat{\rho}^2-\{(\hat{h}+\hat{k})\,\hat{m}_{11}-\hat{h}\hat{k}(\hat{m}_{10}+\hat{m}_{01})\}\,\hat{\rho}-(\hat{h}+\hat{k})-\hat{h}\hat{k}(\hat{m}_{10}+\hat{m}_{01})+\hat{k}\hat{m}_{20}+\hat{h}\hat{m}_{02} = 0. \tag{11}$$

The value of $\hat{m}_{11}$ for use in this quadratic is obtained from (10).

From this point the process is repeated until final solutions are obtained.

4. Example

It is known that linear measurements of the body tend to be normally distributed, and this example is of the stature and hip height of a sample of women which appears in Table A 10 of a report prepared by W. F. F. Kemsley, and published by the Board of Trade (1957). The mean stature of the 4,995 subjects, calculated from the table, was 63·06 in., with standard deviation 2·69 in., and the mean height was 31·56 in., with standard deviation 1·84 in. The correlation between the two measurements was 0·79.

If the distribution is truncated at a stature of 61 in. and a hip height of 29 in., it leaves a residue of 3,858 subjects. The true values of h, k are $-0{\cdot}77$ and $-1{\cdot}39$.

Denoting stature by x, and hip height by y, and using the truncation point as origin, we have

$$\Sigma x = 12{,}018, \quad \Sigma y = 12{,}166, \quad \Sigma x^2 = 53{,}146, \quad \Sigma y^2 = 47{,}530, \quad \Sigma xy = 46{,}266.$$

If we start with approximate values $h = k = -1$ and $\rho = 0{\cdot}6$, successive cycles produce the answer given in Table 1.

TABLE 1

Iterative estimation of parameters in truncated bivariate normal distribution

$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{h}$	$\hat{k}$	$\hat{\rho}$
2·20	2·56	2·44	1·97	−0·86	−1·24	0·77
2·14	2·64	2·52	1·91	−0·81	−1·32	0·79
2·09	2·67	2·52	1·90	−0·78	−1·33	0·80
2·07	2·68	2·51	1·90	−0·77	−1·32	0·79

REFERENCES

BOARD OF TRADE (1957), *Women's Measurements and Sizes*. London: H.M. Stationery Office.

COHEN, A. C. (1955), "Restriction and selection in samples from bivariate normal distributions", *J. Amer. statist. Assoc.*, **50**, 884–893.

—— (1957), "Restriction and selection in multinormal distributions", *Ann. math. Statist.*, **28**, 731–741.

—— (1959), "Simplified estimators for the normal distribution when samples are singly censored or truncated", *Technometrics*, **1**, 217–237.

PEARSON, K. (1930), *Tables for Statisticians and Biometricians*, Part I. Cambridge University Press.

SINGH, N. (1960), "Estimation of parameters of a multivariate normal population from truncated and censored samples", *J. R. statist. Soc.* B, **22**, 307–311.

ZELEN, M. and SEVERO, N. C. (1960), "Graphs for bivariate normal probabilities", *Ann. math. Statist.*, **31**, 619–624.