MISCELLANEA

TRUNCATED SAMPLING FROM DISTRIBUTIONS ADMITTING SUFFICIENT STATISTICS

By DES RAJ

Indian Statistical Institute, Calcutta

1. The problem of estimating the parameters of normal and binormal populations from truncated samples has been considered by several authors including Fisher (1931), Stevens (1937), Cochran (1946), Hald (1949), Cohen (1949, 1950) and Des Raj (1952, 1953). The object of this paper is to obtain in a simple way the large sample variances and covariances of the maximum likelihood estimates (from truncated and censored samples) of the parameters of a very general class of distributions, namely those admitting sufficient statistics. The long and tedious calculations involved in obtaining the expected values of the second order partial derivatives of the likelihood function have been by-passed. For simplicity we shall consider univariate distributions only, but the analysis is the same for multivariate distributions linearly truncated in one variate.

2. It has been shown by Koopman (1936) that, under certain regularity conditions, the necessary and sufficient condition that a distribution depending on p parameters should admit a set of p jointly sufficient statistics is that the probability density function of the distribution be of the form

$$f(x, \theta_j) = \exp\{\sum_k u_k(\theta_j)v_k(x)+A(x)+B(\theta_j)\} \quad \dots \quad (1)$$

where u_k's and B are functions of θ's only and v_k's and A are functions of x only and θ_j stands for $\theta_1, \theta_2, \dots, \theta_p$. It may also be noted, as shown by Tukey (1949), that if a family of distributions admits a set of sufficient statistics, then the family obtained by truncation to a fixed set, or by fixed selection also admits the same set of sufficient statistics. We shall consider the problem when the population (1) is truncated at x_0' and $x_0'' = x_0'+R$ and a random sample of size N gives n_0 measured observations in the range R, n_1 unmeasured observations in the lower tail $(-\infty, x_0')$ and n_2 unmeasured observations in the upper tail (x_0'', ∞). We shall consider the cases when (i) n_1 and n_2 are known separately, (ii) $n_1+n_2 = N-n_0$ is known, (iii) the entire sample is taken from the range R.

3. *Case* (i): The likelihood function of the parameters is given by

$$L = \text{Const.} + n_1 \log G_1(\theta_j) + n_2 \log G_2(\theta_j) + \sum_k u_k(\theta_j) \sum_i v_k(x_i) + \Sigma A(x_i) + n_0 B(\theta_j) \quad \dots \quad (2)$$

where $$G_1(\theta_j) = \int_{-\infty}^{x_0'} f(x, \theta_j) dx, \quad G_2(\theta_j) = \int_{x_0''}^{\infty} f(x, \theta_j) dx. \quad \dots \quad (3)$$

Let $\hat\theta_1, \hat\theta_2, \dots, \hat\theta_p$ be a solution of the maximum likelihood equations

$$\frac{\partial L}{\partial \theta_r} = 0 \quad (r = 1, 2, \dots, p) \quad \dots \quad (4)$$

where $$\frac{\partial L}{\partial \theta_r} = n_1 \frac{\partial}{\partial \theta_r} \log G_1 + n_2 \frac{\partial}{\partial \theta_r} \log G_2 + \Sigma \frac{\partial u_k}{\partial \theta_r} T_k + n_0 \frac{\partial B}{\partial \theta_r} \quad \dots \quad (5)$$

and $$T_k = \sum_i v_k(x_i). \quad \dots \quad (6)$$

Now, if $L = \log \phi$, we have

$$\sum_{n_0+n_1+n_2 = N} \int \phi \prod_1^{n_0} dx_i = 1 \quad \dots \quad (7)$$

so that $$E\left(\frac{\partial}{\partial \theta} \log \phi\right) = E\left(\frac{\partial L}{\partial \theta}\right) = 0 \quad \dots \quad (8)$$

assuming that differentiation is possible within the operator $\Sigma \int$. In this case E denotes expectation over two stages, first for fixed values of n_0, n_1 and n_2 and then for all possible samples of size N.

Thus the equations of expectation (8) and the maximum likelihood equations (4) are similar. If a solution of equations (8) be

$$E(T_k) = \lambda_k(\theta_j, E(n_i)), \quad \dots \quad (9)$$

then a solution of the maximum likelihood equations would be

$$T_k = \lambda_k(\hat\theta_j, n_i). \quad \dots \quad (10)$$

Therefore, $$E\left(\frac{\partial^2 L}{\partial \theta_r \partial \theta_s}\right) = \left(\frac{\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{n = E(n),\ \hat\theta = \theta} \quad \dots \quad (11)$$

This gives $$\left[E\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)\right]^{-1} = \left[\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{n = E(n),\ \hat\theta = \theta}\right]^{-1} \quad \dots \quad (12)$$

where the matrices are non-singular.

This shows that the right-hand-side of (12) simply gives the large sample variances and covariances of the maximum likelihood estimates.
Also, it is easy to see that

$$\left[\left\{E\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)\right\}_{\theta = \hat\theta}\right] = \left[\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{n = E(n),\ \theta = \hat\theta}\right]. \quad \dots \quad (13)$$

so that the 'estimated information matrix' evaluated at the maximum likelihood point is simply given by the right hand side of (13).

4. *Case* (ii): The likelihood function of the parameters in this case is given by

$$L = \text{Const.} + (N - n_0) \log (G_1 + G_2) + \sum_{i=1}^{n_2} \log f(x_i, \theta_j). \qquad \dots (14)$$

As before, the equations of maximum likelihood

$$\frac{\partial L}{\partial \theta_r} = 0 \quad (r = 1, 2, \dots, p)$$

and the equations of expectation

$$E\left(\frac{\partial L}{\partial \theta_r}\right) = 0 \quad (r = 1, 2, \dots, p)$$

are similar. In this case E denotes expectation over two stages, first for fixed values of n_0 and then for all possible samples of size N.

Proceeding as before, we establish that

$$\left[E\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)\right]^{-1} = \left[\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{n_0 = E(n_0),\ \hat{\theta} = \theta}\right]^{-1} \qquad \dots (15)$$

and

$$\left[\left\{E\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)\right\}_{\theta = \hat{\theta}}\right] = \left[\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{n_0 = E(n_0),\ \theta = \hat{\theta}}\right]. \qquad \dots (16)$$

The right hand side of (15) gives the large sample variances and covariances of the maximum likelihood estimates.

5. *Case* (iii): In this case a sample of size N is obtained from the truncated population

$$f^*(x, \theta_j)\, dx, \qquad x_0' < x < x_0'' \qquad \dots (17)$$

where

$$f^*(x, \theta_j) = \left[\int_{x_0'}^{x_0''} f(x, \theta_j) dx\right]^{-1} f(x, \theta_j). \qquad \dots (18)$$

The likelihood function is now given by

$$L = \sum_k n_k(\theta_j) T_k + \sum_i A(x_i) + NB(\theta_j) - N \log G(\theta_j) \qquad \dots (19)$$

where

$$G(\theta_j) = \int_{x_0'}^{x_0''} f(x, \theta_j) dx. \qquad \dots (20)$$

We note that the maximum likelihood equations and the equations of expectation are similar. If a solution of the equations of expectation be

$$E(T_k) = \lambda_k(\theta_j),$$

then a solution of the maximum likelihood equations would be

$$T_k = \lambda_k(\hat{\theta}_j).$$

Therefore $$\left(\frac{\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{\hat{\theta}=\theta} = E\left(\frac{\partial^2 L}{\partial \theta_r \partial \theta_s}\right). \quad \ldots (21)$$

This leads to $$\left[E\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)\right]^{-1} = \left[\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{\hat{\theta}=\theta}\right]^{-1} \quad \ldots (22)$$

where the matrices are non-singular.

This relation shows that the information matrix can be simply calculated from the right hand side of (22).

Again, it is of interest to note that

$$\left[\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)_{\theta=\hat{\theta}}\right] = \left[\left\{E\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)\right\}_{\theta=\hat{\theta}}\right]. \quad \ldots (23)$$

The relation (23) has an interesting geometrical interpretation. If the surface generated by the likelihood function be called the 'likelihood surface', then the so-called Gaussian or total curvature K of the surface at the maximum likelihood point is given by

$$K = \frac{\left|\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right|_{\hat{\theta}}}{\left[1 + \Sigma\left\{\left(\frac{\partial L}{\partial \theta_k}\right)^2\right\}_{\hat{\theta}}\right]^2}. \quad \ldots (24)$$

Since $$\left(\frac{\partial L}{\partial \theta_k}\right)_{\hat{\theta}} = 0,$$

we have $$K = \left|\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right|_{\hat{\theta}} = \left|E\left(\frac{-\partial^2 L}{\partial \theta_r \partial \theta_s}\right)\right|_{\hat{\theta}}. \quad \ldots (25)$$

The information function (or the intrinsic accuracy) therefore measures the Gaussian curvature of the likelihood surface at the point represented by the maximum likelihood estimate. Following Huzurbazar (1949), it can be shown that the maximum likelihood estimate actually maximises the likelihood function and that a solution of the system of likelihood equations is unique.

6. As an example, consider the normal population

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(h' + \frac{x'}{\sigma}\right)^2\right\} dx'$$

with origin at the lower truncation point x_0'.

Let the sample of size N consist of n_1 and n_2 unmeasured observations in the lower and upper tails respectively and n_0 measured observations in the truncated range. From the equations of maximum likelihood, we have

$$\sum_{1}^{n_0} x' = \hat{\sigma}\left[n_1\frac{\hat{l}_1}{\hat{a}_1} - n_2\frac{\hat{l}_2}{\hat{a}_2} - n_0\hat{h}'\right] = P(\hat{h}', \hat{\sigma}),$$

$$\sum_{1}^{n_0} x'^2 = \hat{\sigma}^2\left[n_0(1+\hat{h}')^2 - n_2\frac{\hat{l}_2}{\hat{a}_2}\frac{R}{\hat{\sigma}}\hat{h}^{\prime\prime} - n_1\hat{h}'\frac{\hat{l}_1}{\hat{a}_1}\right] = Q(\hat{h}', \hat{\sigma}),$$

where $$l_1 = \frac{1}{\sqrt{2\pi}}\exp(-h'^2/2), \qquad l_2 = \frac{1}{\sqrt{2\pi}}\exp(-h''^2/2),$$

$$a_1 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{h'}\exp(-t^2/2)\,dt, \; a_2 = \frac{1}{\sqrt{2\pi}}\int_{h''}^{\infty}\exp(-t^2/2)\,dt$$

$$h'' = h' + \frac{R}{\sigma}, \quad h' = \frac{x'_0 - m}{\sigma}.$$

Then, $L = \text{Const.} + n_1\log a_1 + n_2\log a_2 - n_0\log\sigma - \frac{n_0}{2}h'^2 - \frac{1}{2\sigma^2}Q(\hat{h}', \hat{\sigma}) - \frac{h'}{\sigma}P(\hat{h}', \hat{\sigma})$..

As an illustration of the calculation of the elements of the information matrix, we take $E\left(\frac{-\partial^2 L}{\partial\sigma^2}\right)$.

Now, $$\frac{-\partial^2 L}{\partial\sigma^2} = n_2\frac{R^2}{\sigma^4}\frac{l_2^2}{a_2^2} + 2n_2\frac{R}{\sigma^3}\frac{l_2}{a_2} - h''\frac{R^2}{\sigma^4}n_2\frac{l_2}{a_2} - \frac{n_0}{\sigma^2} + \frac{3}{\sigma^4}Q(\hat{h}', \hat{\sigma}) + \frac{2}{\sigma^3}h'P(\hat{h}', \hat{\sigma}).$$

Therefore, $$E\left(\frac{-\partial^2 L}{\partial\sigma^2}\right) = \left(\frac{-\partial^2 L}{\partial\sigma^2}\right)_{n=E(n),\,\hat{\theta}=\theta}$$

$$= \frac{N}{\sigma^2}\left\{\frac{R^2}{\sigma^2}l_2\left(\frac{l_2}{a_2} - h''\right) - \frac{R}{\sigma}l_2 + (2+h'^2)(1-a_1-a_2) - h'(l_1-l_2)\right\}.$$

In this connection we note that in the calculation of the elements of the information matrix, as given by Cohen (1950), $E(n_i)$ should be substituted for n_i if the sampling scheme consists in drawing random samples of a fixed size N so that n_0, n_1 and n_2 are random variables. The reason why Cohen has used the equations of maximum likelihood for the calculation of expectations is the similarity in form of the equations of expectation and the equations of maximum likelihood, as shown in this paper.

REFERENCES

COCHRAN, W. G. (1946) : Use of the IBM equipment in an investigation of the truncated normal problem. Proceedings Research Forum, IBM Corp. 40-43.

COHEN, A. C. (1949) : On estimating the mean and standard deviation of truncated normal distributions. *J. Amer. Stat. Ass.*, **44**, 518-525.

——— (1950) : Estimating the mean and variance of normal populations from singly and doubly truncated samples. *Ann. Math. Stat.*, **21**, 557-569.

DES RAJ (1952) : On estimating the parameters of normal populations from singly truncated samples. *Ganita*, **3**, 41-57.

——— (1953) : On moments estimation of the parameters of a normal population from singly and doubly truncated samples. *Ganita*, **4**, 79-84.

——— (1953) : On estimating the parameters of bivariate normal populations from doubly and singly linearly truncated samples. *Sankhyā*, **12**, 277-290.

——— (1953) : On estimating the parameters of bivariate normal populations from linearly truncated samples. *Ganita*, **4**.

FISHER, R. A. (1931) : Properties and applications of Hh functions. *Math. Tables*, **1**, British Assn. for the advancement of Science, XXVI-XXXV.

HALD, A. (1949) : Maximum likelihood estimation of the parameters of a normal population truncated at a known point. *Skandinavisk Aktuariatidskrift*, **32**, 119-134.

HUZURBAZAR, V. S. (1949) : On a property of distributions admitting sufficient statistics. *Biometrika*, **36**, 71-74.

KOOPMAN, B. O. (1936) : On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.*, **39**, 399-409.

STEVENS, W. L. (1937) : The truncated normal distribution. *Ann. Applied Biology*, **24**, 815-852.

TUKEY, J. W. (1949) : Sufficiency, truncation and selection. *Ann. Math. Stat.*, **20**, 309-311.

Paper received: April, 1953.