ON PARAMETER ESTIMATION FOR TRUNCATED PEARSON TYPE III DISTRIBUTIONS

By George Gerard den Broeder, Jr.[1]

Purdue University

Summary and introduction. The problem of estimating the parameters of the Pearson Type III probability density function (p.d.f.)

$$\phi(t, \alpha) = \alpha f(\alpha t) = [\Gamma(p)]^{-1}\alpha^p t^{p-1} e^{-\alpha t}, \qquad \begin{cases} 0 \leqq t \\ 0 < \alpha \\ 0 < p \end{cases}$$

assuming various forms of truncation has been considered recently by A. C. Cohen, Jr. [2], Des Raj [4] and others. In this paper we obtain maximum likelihood estimates of the parameter α when p is known apriori. Truncation is at a known point $T > 0$.

Four cases are considered: truncation to the right of T with the number of observations in the region of truncation (1) known, and (2) not known; and truncation to the left of T with the number of observations in the region of truncation (3) known, and (4) not known. The information lost in not knowing the number of observations in the regions of truncation is measured in terms of the R. A. Fisher indices of information.

The study of cases (2), and hence of case (1), is an outgrowth of the author's experience with a life testing program from which, unfortunately, data had been recorded only for those specimens which failed within 100 hours of testing. Despite this anomaly of experimental design a maximum likelihood estimate of α was found to exist. The analysis proceeded on the assumption of an exponential failure law ($p = 1$).

Another instance of case (2) arises naturally in connection with the distribution of population in urban communities. Colin Clark [1] has found urban population density, as a function of radial distance from city center, to be adequately described by the Pearson Type III p.d.f. with $p = 2$. The maximum likelihood estimate of the unspecified parameter, α, for cities with circular peripheries, is contained in Section 2.

Cases (3) and (4) are included for the sake of completeness. A possible area of application is to be found in the field of telemetry where frequently the result of a random experiment is measured by an instrument which responds only to inputs in excess of a fixed magnitude.

1. Number of observations to the right of T known. Suppose we have a random sample of size N from the Pearson Type III distribution with known p.

Received Nov. 12, 1953.

[1] Work done while author at NBS, Corona, California.

Suppose further that $N - n$ observations are known only to be greater than $T > 0$ while the actual values, $t_k \leqq T$, of the remaining n observations are given. The likelihood function of the sample is then

$$L(t_k, n) = \begin{cases} [1 - F(\alpha t)]^N & \text{for } n = 0 \\ \binom{N}{n} [1 - F(\alpha t)]^{N-n} \prod_{k=1}^{n} [\alpha f(\alpha t_k)] & \text{for } n > 0 \end{cases} \tag{1.1}$$

where $0 \leqq t_k \leqq T$ and $F(x) = \int_0^x f(x)\, dx$. Without loss of generality we take $T = 1$. Obviously, for $n = 0$, $L(t_k, n)$ is decreasing in α. For $n > 0$, we obtain

$$\frac{\partial \log L}{\partial \alpha} = n[p/\alpha - \left(\frac{N - n}{n}\right) Z(\alpha) - \bar{t}], \tag{1.2}$$

where $\bar{t} = 1/n(\sum_{k=1}^{n} t_k)$ and where $Z(\alpha) = f(\alpha)[1 - F(\alpha)]^{-1}$ is the reciprocal of Mills' ratio for the Pearson Type III distribution. The maximum likelihood estimate of α, say $\hat{\alpha}$, is thus the unique solution of $\bar{t} = \alpha^{-1}[p - ((N - n)/n)\alpha Z(\alpha)]$.

The uniqueness and existence of $\hat{\alpha}$ are established using the result, shown by the author in an unpublished paper [3], that $\alpha Z(\alpha)$ is increasing with range $(0, \infty)$ for all $p > 0$. Thus, since both α^{-1} and $[p - ((N - n)/n)\alpha Z(\alpha)]$ are decreasing, the right-hand side of the last equation has a unique zero (except, of course, when $n = N$) for some $\alpha_0 > 0$ and is decreasing with range $(0, \infty)$ for $0 < \alpha < \alpha_0$.

2. Number of observations to the right of T not known. When only n is known we are led to consider the likelihood function

$$L^*(t_k \mid n) = [F(\alpha)]^{-n} \prod_{k=1}^{n} [\alpha f(\alpha t_k)], \qquad 0 \leqq t_k \leqq 1 \tag{2.1}$$

from which we obtain

$$\frac{\partial \log L^*}{\partial \alpha} = n[p/\alpha - W(\alpha) - \bar{t}], \tag{2.2}$$

where $W(\alpha) = f(\alpha)[F(\alpha)]^{-1}$. The maximum likelihood estimate of α, say α^*, is thus a solution of $\bar{t} = p/\alpha - W(\alpha)$.

The pertinent properties of $[p/\alpha - W(\alpha)]$ are readily established by considering the random variable ξ with p.d.f.

$$g(t:\alpha) = \alpha f(\alpha t)[F(\alpha)]^{-1}, \qquad 0 \leqq t \leqq 1. \tag{2.3}$$

We find that $E(\xi) = p/\alpha - W(\alpha)$ and that Var $(\xi) = -\partial E(\xi)/\partial \alpha > 0$. Hence, $E(\xi)$ is decreasing in α and the uniqueness of α^* is assured. (We note that α^* is obtained also by the method of moments). The range of $E(\xi)$ may be established by determining its limits from the equivalent expression

$$E(\xi) = p/\alpha - W(\alpha) = \left[\int_0^\alpha x^p e^{-x}\, dx\right]\left[\alpha \int_0^\alpha x^{p-1} e^{-x}\, dx\right]^{-1}. \tag{2.4}$$

Obviously, $\lim_{\alpha\to\infty} E(\xi) = 0$. To determine its limit as α approaches zero, we apply L'Hospital's rule twice obtaining

$$\lim_{\alpha\to 0} E(\xi) = \frac{p}{p+1}. \tag{2.5}$$

This latter result, (2.5), introduces an interesting complication since $0 \leqq \bar{t} \leqq 1$ and $P(\bar{t} \geqq p/(p+1)) > 0$. Thus $\bar{t} = E(\xi)$ appears to have no solution for $\bar{t} \geqq p/(p+1)$. However, if we interpret $L^*(t_k \mid n)$ to be the likelihood function associated with the random variable ξ introduced above and then complete the family of p.d.f.'s (2.3), by adjoining

$$g(t:0) = \lim_{\alpha\to 0} g(t:\alpha) = p^{-1}t^{p-1}, \qquad 0 \leqq t \leqq 1, \tag{2.6}$$

the likelihood function, $L^*(t_k \mid n)$, is seen to be maximized for $\bar{t} \geqq p/(p+1)$ by $\alpha = 0$.

3. Number of observations to the left of T known. In the event that $N - n$ observations are known only to be less than T while the actual values, $t_k \geqq T$, of the remaining n observations are given, we seek to maximize (again taking $T = 1$),

$$S(t_k, n) = \begin{cases} [F(\alpha)]^N & \text{for } n = 0 \\ \binom{N}{n} [F(\alpha)]^{N-n} \prod_{k=1}^{n} [\alpha f(\alpha t_k)] & \text{for } n > 0 \end{cases} \tag{3.1}$$

with $t_k \geqq 1$. For $n = 0$, $S(t_k, n)$ is increasing in α. For $n > 0$ we obtain

$$\partial \log S/\partial\alpha = n\left[p/\alpha + \left(\frac{N-n}{n}\right) W(\alpha) - \bar{t}\right]$$

which has a unique zero since, for all $p > 0$, $W(\alpha)$ is decreasing and $\lim_{\alpha\to\infty} W(\alpha) = 0$, [3].

4. Number of observations to the left of T not known. Knowing only the values of the n observations, $t_k \geqq T$, we are required to consider the likelihood function

$$S^*(t_k \mid n) = [1 - F(\alpha)]^{-n} \prod_{k=1}^{n} [\alpha f(\alpha t_k)], \qquad t_k \geqq 1 \tag{4.1}$$

from which we obtain

$$\frac{\partial \log S^*}{\partial\alpha} = n[p/\alpha + Z(\alpha) - \bar{t}]. \tag{4.2}$$

That (4.2) has a unique zero is established in a manner analogous to that of Section 2. Consider a random variable η with p.d.f.

$$h(t:\alpha) = \alpha f(\alpha t)[1 - F(\alpha)]^{-1}, \qquad t \geqq 1. \tag{4.3}$$

Then $E(\eta) = p/\alpha + Z(\alpha)$, and Var $(\eta) = -\partial E(\eta)/\partial\alpha > 0$. That is, $E(\eta)$ is a decreasing function of α with range $(1, \infty)$, $(\lim_{\alpha\to\infty} Z(\alpha) = 1$, [3]). Hence, since $\bar{t} \geqq 1$, (4.2) has a unique zero.

5. Concerning information losses. The likelihood function, $L^*(t_k \mid n)$, of Section 2 admits two interpretations. In one instance (urban population density) it is the likelihood function of the random variable ξ with p.d.f., (2.4). In another instance (anomalous experimental design) it is a "conditional" likelihood function of $L(t_k, n)$:

$$L(t_k, n) = \binom{N}{n}[1 - F(\alpha)]^{N-n}[F(\alpha)]^n L^*(t_k \mid n), \quad 0 \leqq t_k \leqq 1. \tag{5.1}$$

It is this latter instance which involves loss of information and with which we are concerned in this section.

The information lost in not knowing N may conveniently be measured in terms of R. A. Fisher indices of information. A measure which suggests itself and which is adequate for our purposes is

$$J(L) = E\left[\frac{\partial^2 \log L^*}{\partial\alpha^2} - \frac{\partial^2 \log L}{\partial\alpha^2}\right] \tag{5.2}$$

where the expectation is with respect to the p.d.f., $L(t_k, n)$. The analog of (5.2), $J(S)$, measures the information loss in the case of truncation on the left. Our only justification for employing the difference of the R. A. Fisher indices rather than, say, their ratio rests with the nature of the results obtained.

Denoting $-E(\partial^2 \log L/\partial\alpha^2)$ and $-E(\partial^2 \log L^*/\partial\alpha^2)$ by $I(L)$ and $I(L^*)$ respectively and differentiating (1.2) and (2.2) with respect to α, we obtain

$$I(L) = (p/\alpha^2)E(n) + Z'(\alpha)E(N - n) = N[p/\alpha^2 F(\alpha) + Z'(\alpha)(1 - F(\alpha))],$$

and

$$I(L^*) = [p/\alpha^2 + W'(\alpha)]E(n) = N[p/\alpha^2 + W'(\alpha)]F(\alpha).$$

Hence

$$J(L) = N[Z'(\alpha)(1 - F(\alpha)) - W'(\alpha)F(\alpha)] = NW(\alpha)Z(\alpha).$$

Similarly

$$I(S) = N[p/\alpha^2(1 - F(\alpha)) - W'(\alpha)F(\alpha)],$$

$$I(S^*) = N[p/\alpha^2 - Z'(\alpha)][1 - F(\alpha)],$$

so that $J(S) = NW(\alpha)Z(\alpha)$, and we see that our measure of information loss in not knowing the number of observations in the tail is independent of whether truncation is on the right or on the left.

Finally, we recall that the information index associated with a random sample of size N from the Pearson Type III distribution is Np/α^2. Our intuition suggests that we should have $I(L) + I(S^*) = I(L^*) + I(S) = Np/\alpha^2$, which is easily seen to obtain.

REFERENCES

[1] C. Clark, "Urban population densities," *J. Roy. Stat. Soc.* (General), Vol. CXIV, Part IV (1951), pp. 490–496.

[2] A. C. Cohen, Jr., "Estimating parameters of Pearson Type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp. 411–423.

[3] G. den Broeder, "Concerning Mills' Ratio and related functions for the Pearson Type III distribution," *Statistical Laboratory Technical Note*, (1955), Purdue University.

[4] Des Raj, "Estimation of the parameters of type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 336–349.