SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Editors: P. C. MAHALANOBIS, C. R. RAO

SERIES B, VOL. 25 DECEMBER 1963 PARTS 3 & 4

A PROPERTY OF THE PARETO DISTRIBUTION

By N. BHATTACHARYA

Indian Statistical Institute

SUMMARY. Suppose we have an income distribution over the range x_0 to ∞ ($x_0 > 0$), and suppose we consider truncated forms of this distribution over (x_l, ∞) where $x_l \geqslant x_0$. It is proved that the concentration curve (and the concentration coefficient) for this truncated distribution will be independent of x_l if and only if the income distribution has the Pareto form.

Consider a Pareto distribution of income (x) over (x_0, ∞). (There may or may not be a part of the distribution below x_0 which does not follow the Pareto equation.) Then we have

$$N_x = Ax^{-\nu} \qquad \dots \quad (1)$$

where N_x is the number of persons having income $\geqslant x$, and A, ν are parameters. Since $N_{x_l} = Ax_l^{-\nu}$, we have the distribution function $F(x|x \geqslant x_l)$ of the truncated distribution

$$F(x|x \geqslant x_l) = 1-\left(\frac{x}{x_l}\right)^{-\nu} \qquad \dots \quad (2)$$

The concentration curve of this distribution is the same as that of $y = \frac{x}{x_l}$, since x_l is a constant for the time being. But the distribution function of y is given by

$$\begin{aligned} G(y|x \geqslant x_l) &= F(x_l y|x \geqslant x_l) \\ &= 1-y^{-\nu} \,(y \geqslant 1). \end{aligned} \qquad \dots \quad (3)$$

Since this distribution is independent of x_l, the concentration curve for y and hence that of x is independent of x_l. This proves the first part of the theorem. It follows that the Gini concentration coefficient (also called the Lorenz ratio) is also independent of the point of truncation x_l.

As regards the converse, suppose $H(x)$ is the general form of the distribution function over (x_0, ∞) which possesses the above property. The d.f. of the truncated distribution over (x_l, ∞) with $x_l \geqslant x_0$ is given by

$$F(x|x \geqslant x_l) = \frac{H(x)-H(x_l)}{1-H(x_l)}. \qquad \dots \quad (4)$$

Consider the variate $y = \frac{x}{x_l}$ again. The d.f. of y corresponding to $F(x|x \geqslant x_l)$ is

$$G(y|x \geqslant x_l) = \frac{H(x_l y)-H(x_l)}{1-H(x_l)} \text{ for all } y \geqslant 1. \qquad \dots (5)$$

Since the concentration curve for the truncated x-distribution is the same as the concentration curve for the corresponding distribution of $\frac{x}{x_l} = y$, and in virtue of the one-to-one relationship between the distribution function of y and its concentration curve,* our requirement becomes that $G(y|x \geqslant x_l)$ should be independent of x_l. Thus we get the condition that

$$\frac{H(xy)-H(x)}{1-H(x)} = G(y) \text{ independent of } x. \qquad \dots (6)$$

Here $1 \leqslant y \leqslant \infty$, and $x_0 \leqslant x < \infty$.

Assuming differentiability properties we get the frequency function of y as

$$\frac{\frac{\partial H(xy)}{\partial y}}{1-H(x)} = \frac{x\frac{dH(z)}{dz}\Big|_{z=xy}}{1-H(x)} = g(y), \text{ independent of } x. \qquad \dots (7)$$

Now it appears that the condition that $g(1)$ is independent of x ensures that $H(x)$ is of the Pareto form. For let $y \to 1$ from above in (7). We get

$$\frac{\frac{xdH(x)}{dx}}{1-H(x)} = g(1) = c \text{ say } (c > 0). \qquad \dots (8)$$

The differential equation $\quad \frac{dH(x)}{dx}+\frac{c}{x} H(x) = \frac{c}{x} \qquad \dots (9)$

is one of a standard form, and the general solution is

$$1-H(x) = ax^{-c} \text{ (where } a > 0) \qquad \dots (10)$$

which is of the Pareto form.

It has already been seen that for the Pareto distribution $g(y|x \geqslant x_l)$ is independent of x_l for all $y \geqslant 1$.

Observations. Income distributions emanating from income-tax statistics are generally of the truncated form (ignoring some complications), but concentration curves based on them are quite often used in inequality studies. If the entire distribution is of the Pareto form, the truncated distribution can be safely used for inferences about the inequality of the entire distribution. But since the Pareto curve rarely fits over the whole range, one should be careful to distinguish between concentration curves based on truncated distributions from those based on complete ones. [It would be interesting to investigate the effect of truncation on the concentration curve, for instance, when the complete distribution is lognormal.] What is more important is that when truncated distributions of different years are compared for inequality, the points of truncation should be made equal in the real income sense, because differences in the point of truncation may lead to fictitious differences in inequality.

Paper received : February, 1963.

* This is due to the fact that whatever be x_l, the distribution of y starts from the value 1.