ON THE PROPERTY OF DULLNESS OF PARETO DISTRIBUTION

by

Wing-Yue Wong
University of Malaya

Technical Report #82-16

Department of Statistics
Purdue University

*This research was supported by the Office of Naval Research contract N00014-75-C0455 at Purdue University during the author's visit in May 1982. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ON THE PROPERTY OF DULLNESS OF PARETO DISTRIBUTION*

Wing-Yue Wong
University of Malaya

SUMMARY

In the analysis of income distribution, it has been observed that the true income is usually underreported. In this paper we study the statistical properties of reported incomes through a multiplicative underreported model. The concept of "locally nearly dullness" of the reported income is introduced and studied. The above property of dullness has been used to characterize the Pareto distribution for the problem of reported incomes.

1. INTRODUCTION

The Pareto distribution plays an important role in the study of many socio-economic problems, especially in the theory of income. It is common that individuals may underreport their true incomes to avoid payment of some portion of their income tax. This phenomenon of under-reporting has been investigated by Krishnaji (1970) and Talwalker (1980).

Let the random variable X and Y represent the actual income and reported income, respectively. Krishnaji and Talwalker both assume that Y is related to X through the relation

$$Y = RX. \tag{1.1}$$

It has been shown by Talwalker (1980) that if $P(Y \leq 1)=0$, then a necessary and sufficient condition for X to have a Pareto distribution of the form

$$F_X(x) = 1 - x^{a}, \ x \geq 1, \ a > 0 \tag{1.2}$$

*This research was supported by the Office of Naval Research contract N00014-75-C-0455 at Purdue University during the author's visit in May 1982.

is that

$$P(X > xt | X > x) = P(X > t), \forall x, t \geq 1. \tag{1.3}$$

However, it should be pointed out that X has distribution (1.2) if and only if $Z = \ell n X$ has an exponential distribution

$$F_Z(z) = 1 - \exp(-az), z > 0.$$

Condition (1.3) can be expressed in terms of Z as the so-called "lack of memory" property:

$$P(Z > u + \nu | Z > u) = P(Z > \nu), \forall u, \nu \geq 0.$$

Talwalker (1980) has also shown that if $F_X(x)$ is absolutely continuous and concave in x, then F_X is given by (1.2) if and only if condition (1.3) holds for some $x_0 \geq 1$ and for all $t \geq 1$.

In [Krishnaji 1970], it is assumed that X and R are stochastically independent and R has distribution function given by

$$F_R(r) = r^p, 0 < r < 1, p > 0 \tag{1.4}$$

and $P(RX>m)>0$ for some $m>0$. Then Krishnaji showed that X has a Pareto distribution of the form

$$F_X(x) = 1 - \left(\frac{m}{x}\right)^a , \; x \geq m, \; a > 0 \tag{1.5}$$

if and only if

$$P(RX > t \mid RX > m) = P(X > t), \; \forall t \geq m. \tag{1.6}$$

However, if incomes fall below the tax exempt level, there should be no incentive to underreport. Moreover, it will be more realistic to assume that the underreporting factor R depends on the true income X. Therefore in this paper, we assume that the conditional distribution of R given X=x is given by

$$F_R(r|x) = \begin{cases} 0 & \text{if } r < \frac{m}{x} \\ r^p & \text{if } \frac{m}{x} \leq r < 1 \\ 1 & \text{if } r \geq 1 \end{cases} \tag{1.7}$$

where m represents the tax exempt level.

In section 2, we study the properties of Pareto distribution by comparing the tail probabilities of RX and X. Some properties of dullness of Pareto distribution analogous to that of "lack of memory" will be discussed in section 3.

2. TAIL PROPERTIES OF THE REPORTED INCOME

Theorem 2.1. Let X be a random variable such that $F_X(x)=0$, $x<m$, $m>0$, and R be a random variable whose conditional distribution is given by (1.7). Then

$$F_X(x) = \begin{cases} 1 - \alpha\left(\frac{m}{x}\right)^a, & x > m,\ 0 < \alpha \le 1,\ a > 0 \\ 1 - \alpha & x = m \end{cases} \tag{2.1}$$

if and only if

$$P(RX > t) = \frac{p}{a+p} P(X > t),\ \forall t \ge m. \tag{2.2}$$

Proof. (Necessity). Suppose the distribution of X is given by (2.1). Then for $t \ge m$

$$P(RX > t) = \int_t^\infty P(R > \frac{t}{x} | X = x) dF_X(x)$$

$$= P(X > t) - \int_t^\infty \left(\frac{t}{x}\right)^p \alpha \cdot \frac{am^a}{x^{a+1}} dx$$

$$= \frac{p}{a+p} P(X > t).$$

(Sufficiency). Let $G(x)=1-F_X(x)$. By using the method of integration by part, we obtain that for $t \ge m$

$$P(RX > t) = pt^p \int_t^\infty \frac{G(x)}{x^{p+1}} dx.$$

Replacing $x^{-p}G(x)$ by $H(x)$, condition (2.2) can now be expressed as

$$\int_t^\infty \frac{H(x)}{x}\,dx = \frac{H(t)}{a+p}\ , \qquad t \geq m. \tag{2.3}$$

This implies that H is a continuous function and hence differentiable. Thus we obtain that

$$H(t) = bt^{-(a+p)},\ t \geq m \text{ and for some constant } b,$$

or

$$1 - F_X(x) = bx^{-a},\ x \geq m.$$

If $F_X(m)=1-\alpha$, then F_X is given by (2.1).

Remark 2.1. It should be pointed out that if we impose an additional assumption that $F_X(m)=0$ or replace condition (2.2) by

$$P(RX \geq t) = \frac{p}{a+p} P(X \geq t),\ t \geq m \tag{2.2}$$

then it can be shown that $\alpha=1$ and (2.1) reduces to (1.5).

The following result shows another property of dullness of Pareto distribution.

<u>Theorem 2.2</u>. Let X and R be the same as stated in Theorem 2.1. Then the distribution of X is given by (2.1) if and only if

$$P(RX > y|RX > x) = P(X > y|X > x),\ y > x \geq m. \tag{2.4}$$

<u>Proof</u>. If F_X is given by (2.1), then

$$P(RX > y) = \frac{p\alpha}{a+p}\left(\frac{m}{y}\right)^a,\ y \geq m.$$

This shows that

$$P(RX > y|RX > x) = \left(\frac{x}{y}\right)^a = P(X > y|X > x) \text{ for } y > x \geq m.$$

Suppose conversely that condition (2.4) holds. Set

$$H(x) = x^p(1 - F_X(x)).$$

Then condition (2.4) reduces to

$$\frac{H(y)}{\int_y^\infty \frac{H(t)}{t}\,dt} = \frac{H(x)}{\int_x^\infty \frac{H(t)}{t}\,dt}, \quad y > x \geq m.$$

This implies that

$$\int_x^\infty \frac{H(t)}{t}\,dt = cH(x)$$

for $x \geq m$ and for some constant $c>0$. By using an analogous argument as in the proof of Theorem 2.1, the result follows.

Remark 2.2. As analogous to remark (2.1), it is easily seen that if $F_X(m)=0$ or condition (2.4) is replaced by

$$P(RX \geq y|RX \geq x) = P(X \geq y|X \geq x),\ y \geq x \geq m, \tag{2.4}$$

then (2.1) reduces to (1.5).

3. PROPERTY OF DULLNESS OF REPORTED INCOME

In this section we shall investigate some property of dullness of the reported income.

Definition 3.1. The reported income RX is said to be nearly dull if

$$P(RX > mxy|RX > mx) = cP(RX > my) \text{ for some } c > 0 \text{ and for all } x, y \geq 1. \tag{3.1}$$

Theorem 3.1. Let X and R be defined as in Theorem 2.1. Then the reported income is nearly dull if and only if the distribution of X is given by (2.1).

Proof. It is easy to verify that if the distribution of X is given by (2.1), then the reported income is nearly dull.

To prove the converse, suppose the reported income is nearly dull. It is easy to see that $c^{-1}=P(RX>m)$. Let $H(x)=P(RX>mx|RX>m)$. Then condition (3.1) becomes

$$H(xy) = H(x)H(y) \text{ for all } x, y \geq 1.$$

This, together with the condition that H(1)=1, imply that

$$H(x) = x^a \text{ for some } a > 0.$$

Let $G(x)=1-F_X(x)$. Condition (3.1) is then reduced to

$$px^p \int_x^\infty \frac{G(t)}{t^{p+t}}\, dt = \frac{1}{c}\left(\frac{m}{x}\right)^a \text{ for all } x \geq m$$

or

$$\int_x^\infty \frac{G(t)}{t^{p+1}}\, dt = \frac{1}{cp}\,\frac{m^a}{x^{a+p}}, \quad x \geq m.$$

This shows that

$$1 - F_X(x) = \frac{a+p}{cp}\left(\frac{m}{x}\right)^a.$$

Thus F_X is given by (2.1) with $a+p=\alpha cp$.

Next we shall use a method analogous to the one employed by Marsaglia and Tubilla (1975) to study the property of dullness of reported income.

Definition 3.2. The reported income RX is said to be nearly dull at $x(x \geq m)$ if there exists a positive constant $c(x)$ such that

$$P(RX > xy \mid RX > x) = c(x)P(RX > my) \text{ for all } y \geq 1. \tag{3.3}$$

Theorem 3.2. If the reported income is nearly dull at two log-incommensurable points, then the distribution of X is given by (2.1).

When we say x_1 and x^2 are log-incommensurable we mean $\frac{\ln x_1}{\ln x_2}$ is irrational. Before proving the Theorem, let us derive some results related to the concept of locally nearly dullness.

It is easy to verify that the reported income is nearly dull at m.

Lemma 3.1. If the reported income is nearly dull at x, then the positive constant $c(x)$ is independent of x. Furthermore, for any positive integer r, the reported income is also nearly dull at $m^{-(r-1)}x^r$.

Proof. If the reported income is nearly dull at x, then it is easy to verify that $c(x) = \{P(RX>m)\}^{-1}$. Henceforth, we shall denote the constant $c(x)$ simply by c.

For any $y \geq 1$, we have

$$\begin{aligned} P(RX > m^{-(r-1)}x^r y) &= cP(RX > x)P(RX > m^{-r+2}x^{r-1}y) \\ &= cP(RX > x)^2 P(RX > m^{-r+3}x^{r-2}y) \\ &= cP(RX > x)^r P(RX > my). \end{aligned} \tag{3.4}$$

This implies that

$$\{P(RX > x)\}^r = c^{-r+1}P(RX > m^{-(r-1)}x^r).$$

Substituting this into (3.4), we obtain that

$$P(RX > m^{-(r-1)}x^r y) = cP(RX > m^{-(r-1)}x^r)P(RX > my).$$

This proves the lemma.

Lemma 3.2. If the reported income is nearly dull at x_1 and x_2, $x_1 < x_2$, then it is also dull at $\frac{mx_2}{x_1}$.

Proof. For any $y \geq 1$

$$\begin{aligned} P(RX > x_2 y) &= P(RX > x_1 \cdot \frac{x_2}{x_1} y) \\ &= cP(RX > x_1)P(RX > m \frac{x_2}{x_1} y). \end{aligned}$$

On the other hand,

$$P(RX > x_2 y) = cP(RX > x_2)P(RX > my).$$

Moreover,

$$P(RX > x_2) = P(RX > x_1 \cdot \frac{x_2}{x_1}) = cP(RX > x_1)P(RX > m \frac{x_2}{x_1}).$$

This implies that

$$P(RX > m \frac{x_2}{x_1}) = P(RX > m \frac{x_2}{x_1})P(RX > my).$$

Proof of Theorem 3.2. Let D denote the set of all points at which the reported income is nearly dull. Suppose the reported income is nearly dull at x_i, i=1,2 and x_1 and x_2 are log-incommensurable. Let $x_i=my_i$, i=1,2. Without loss of genrality, we may assume that $y_1<y_2$. Let r_1 be the largest integer such that

$$y_2 y_1^{-(r_1+1)} < 1 < y_2 y_1^{-r_1}.$$

Denote $z_1=y_2y_1^{-r_1}$. Then we have $1<z_1<y_1$. Moreover z_1 and y_1 are incommensurable. By using the same argument, we can find a positive integer r_2 such that

$$y_1 z_1^{-(r_2+1)} < 1 < y_1 z_1^{-r_2}.$$

Set $z_2=y_1z_1^{-r_2}$. Note that $z_2<z_1$. By repeating the process infinitely many times, we can generate a decreasing sequence of positive numbers

$$z_1 > z_2 > ,\ldots,> 1,$$

with $mz_i \in D$, i=1,2,... . Clearly $\lim_{n\to\infty} z_n=1$. Next we want to prove that D is dense in $[m,\infty)$. Suppose x>m and $\epsilon>0$. Choose $mz \in D$ such that

$$1 < z < \frac{x+\frac{\epsilon}{2}}{x-\frac{\epsilon}{2}}.$$

Let n denote the largest integer such that $z^n \leq \frac{x}{m} + \frac{\epsilon}{2m}$. Then $z^n \geq \frac{x}{m} - \frac{\epsilon}{2m}$. For if $z^n < \frac{x}{m} - \frac{\epsilon}{2m}$,

$$z^{n+1} < \left(\frac{x}{m} - \frac{\epsilon}{2m}\right)\left(\frac{x + \frac{t}{2}}{x - \frac{t}{2}}\right) = \frac{x}{m} + \frac{\epsilon}{2m} .$$

This contradicts to the choice of n. Thus we have proved that

$$x - \epsilon < x - \frac{\epsilon}{2} \leq mz^n < z + \frac{t}{2} < x + \epsilon .$$

But $mz^n \in D$. Thus D is dense in $[m,\infty)$. Since $P(RX>x)$ is right continuous in x, by Theorem 3.1, the result follows.

REFERENCES

[1] Krishnaji, N. (1970). Characterization of the Pareto Distribution through a model of underreported incomes. Econometrica, 38, 251-255.

[2] Marsaglia, G. and Tubilla, A. (1975). Note on the "lack of memory" property of the exponential distribution. Ann. Probability, 3, 353-354.

[3] Talwalker, S. (1980). On the property of dullness of Pareto distribution. Metrika, 27, 115-119.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

On the property of dullness of Pareto distribution.

5. TYPE OF REPORT & PERIOD COVERED

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

Wing-Yue Wong

8. CONTRACT OR GRANT NUMBER(s)

N00014-75-C-0455

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Purdue University
West Lafayette, Indiana 47907

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

12. REPORT DATE

13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Pareto distribution, multiplicative underreported income, locally nearly dullness, reported incomes.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In the analysis of income distribution, it has been observed that the true income is usually underreported. In this paper we study the statistical properties of reported incomes through a multiplicative underreported model. The concept of "locally nearly dullness" of the reported income is introduced and studied. The above property of dullness has been used to characterize the Pareto distribution for the problem of reported incomes.

DD FORM 1 JAN 73 1473

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)