The Annals of Statistics
1982, Vol. 10, No. 3, 1021–1024

A CHARACTERIZATION OF THE MULTIVARIATE PARETO DISTRIBUTION

By P. E. Jupp[1] and K. V. Mardia

University of St. Andrews and University of Leeds

For a random vector **X** on $\mathbf{X} > \mathbf{b}$ whose mean exists, the mean residual lifetime $E(\mathbf{X} - \mathbf{c} \mid \mathbf{X} > \mathbf{c})$ is an affine function of **c** on $\mathbf{c} > \mathbf{b}$ if and only if **X** can be partitioned into independent random vectors which have shifted multivariate Pareto or exponential distributions. An interpretation in terms of income-distribution is suggested for the Pareto case. It is also shown that every multivariate distribution whose mean exists is determined by its mean residual lifetime.

1. Introduction. The Pareto distribution is widely regarded as a suitable model for the distribution of incomes in a given year. Other models which have been suggested include the lognormal distribution considered by Gibrat (1931), the stable distribution proposed by Mandelbrot (1961), and the hyperbolic distribution of Barndorff-Nielsen (1977). These are all asymptotically equivalent to the Pareto distribution in the upper tail. However, little attention has been given to the distribution of incomes in successive years. We suggest that the multivariate Pareto distribution of Type 1 introduced by Mardia (1962) may prove useful for this purpose.

The object of this note is to provide a simple characterization of the multivariate Pareto distribution in terms of the mean residual lifetime.

Let $\mathbf{X} = (X_1, \cdots, X_p)'$ denote a random vector. For vectors **x** and **y** we shall use $\mathbf{x} > \mathbf{y}$ to mean that $x_i > y_i$ for $i = 1, \cdots, p$. Then the multivariate Pareto distribution has tail distribution function

$$P(\mathbf{X} > \mathbf{c}) = \{1 + \textstyle\sum_{i=1}^{p} (b_i^{-1} c_i - 1)\}^{-a}, \quad \mathbf{c} \geq \mathbf{b}, \tag{1.1}$$

where $b_i > 0$ for $i = 1, \cdots, p$ and $a > 0$. If **X** has distribution (1.1) then each X_i has a Pareto distribution

$$P(X_i > c_i) = (b_i^{-1} c_i)^{-a}, \quad c_i > b_i,$$

and when $a > 2$, for $i \neq j$, X_i and X_j have correlation coefficient $1/a$. These desirable properties together with the simple form of (1.1) suggest that this may prove a useful model for the distribution of income in p successive years.

Other applications of the multivariate Pareto distribution have been given by Hutchinson (1979) who shows that this distribution arises as a gamma mixture of distributions in which $X_1, \cdots, X_p$ are independently exponentially distributed.

2. A characterization. The importance of the univariate Pareto distribution has been emphasised by the characterization results of Krishnaji (1970) and of Revankar, Hartley and Pagano (1974) who considered under-reporting of incomes to tax authorities. Another characterization was given by Dallas (1976). Our emphasis is rather on savings and subsistence levels and this leads to a characterization of the multivariate Pareto distribution which generalizes that of Revankar et al. for the univariate case.

By allowing the group of translations to act on the family of distributions (1.1) we obtain the larger family of shifted multivariate Pareto distributions which have tail distribution

Received June 1980; revised February 1982.

[1] Part of this research was carried out while this author was on a Research Fellowship at the University of Leeds funded by the Science Research Council.

AMS 1980 *subject classifications.* Primary, 62H05, 62F10; secondary, 62P20.

Key words and phrases. Characterization, mean residual lifetime, multivariate Pareto distribution.

functions

$$P(\mathbf{X} > \mathbf{c}) = \{1 + \textstyle\sum_{i=1}^{p} d_i(c_i - b_i)\}^{-a}, \quad \mathbf{c} \geq \mathbf{b}, \tag{2.1}$$

where $d_i > 0$ for $i = 1, \cdots, p$. This family is closed under marginalizing and conditioning: if $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ is distributed according to (2.1) then both $\mathbf{X}_2$ and $\mathbf{X}_1 \mid \mathbf{X}_2 = \mathbf{x}_2$ are also distributed according to (2.1). Note that the mean of a distribution in the family (2.1) exists if and only if $a > 1$. We shall need to consider also another family of distributions. It will be convenient to define a multivariate exponential distribution to be that of a random vector $\mathbf{X}$ for which the $X_i - b_i$ have independent exponential distributions.

For any multivariate distribution we can define the mean residual lifetime as

$$\mathbf{e}(\mathbf{c}) = E(\mathbf{X} - \mathbf{c} \mid \mathbf{X} > \mathbf{c}) \tag{2.2}$$

wherever this exists, so generalizing a familiar univariate concept in reliability theory (see, for example, Swartz, 1973). Our main result characterizes those distributions for which the mean residual lifetime is an affine function.

THEOREM 1. *Let* $\mathbf{X}$ *be a random vector satisfying* $\mathbf{X} > \mathbf{b}$ *for some vector* $\mathbf{b}$. *Then*

$$E(\mathbf{X} - \mathbf{c} \mid \mathbf{X} > \mathbf{c}) = A\,\mathbf{c} + \mathbf{f}, \quad \mathbf{c} > \mathbf{b}, \tag{2.3}$$

for some constant matrix A *and constant vector* $\mathbf{f}$ *if and only if (after a possible renumbering of coordinates) there is a partitioning* $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_q)$ *of* $\mathbf{X}$ *into independent vectors* $\mathbf{X}_r$, *where each* $\mathbf{X}_r$ *has either a multivariate exponential distribution or a shifted multivariate Pareto distribution with exponent* $a > 1$.

PROOF. If $\mathbf{X}$ has a shifted multivariate Pareto distribution then a calculation shows that (2.3) holds with $A_{ij} = d_i^{-1} d_j (a-1)^{-1}$ and $f_i = (1 - \sum_{k=1}^{p} b_k d_k) d_i^{-1} (a-1)^{-1}$ for $i, j = 1, \cdots, p$. If $\mathbf{X}$ has a multivariate exponential distribution then (2.3) holds with $A = 0$. It follows that all the distributions considered satisfy (2.3). To prove the converse we first define $G(\mathbf{c}) = P(\mathbf{X} > \mathbf{c})$, $F(\mathbf{c}) = P(\mathbf{X} \leq \mathbf{c})$ and

$$H_i(\mathbf{c}) = \int_{\mathbf{x} > \mathbf{c}} (x_i - c_i)\, dF(\mathbf{x}), \quad i = 1, \cdots, p.$$

Then (2.3) can be rewritten as

$$H_i(\mathbf{c}) = G(\mathbf{c}) e_i(\mathbf{c}), \quad i = 1, \cdots, p, \tag{2.4}$$

where $\mathbf{e}(\mathbf{c}) = A\mathbf{c} + \mathbf{f}$ for $\mathbf{c} > \mathbf{b}$. Note that the existence of the conditional expectation in (2.3) implies that $\mathbf{e}(\mathbf{c}) \geq \mathbf{H}(\mathbf{c}) > \mathbf{0}$ for $\mathbf{c} > \mathbf{b}$, so $A_{ij} \geq 0$ for $i, j = 1, \cdots, p$. Put $\mathbf{g} = A\mathbf{b} + \mathbf{f}$. As $H_i(\cdot)$ is a decreasing function of each c_j, we have

$$g_i = \lim_{\mathbf{c} \to \mathbf{b}} e_i(\mathbf{c}) \geq H_i(\mathbf{c}) > 0, \quad i = 1, \cdots, p,$$

for $\mathbf{c} > \mathbf{b}$.

We must next show that $H_i(\cdot)$ is partially differentiable with respect to c_i. Let $\boldsymbol{\delta}$ be a vector with ith component δ_i and all other components zero. Then

$$H_i(\mathbf{c} + \boldsymbol{\delta}) - H_i(\mathbf{c}) = -\operatorname{sgn}(\delta_i) \int_I (x_i - c_i)\, dF(\mathbf{x}) - \delta_i G(\mathbf{c} + \boldsymbol{\delta}), \tag{2.5}$$

where the region of integration I is defined by $c_j < x_j$ for $i \neq j$ and either $c_i < x_i \leq c_i + \delta_i$ or $c_i + \delta_i < x_i \leq c_i$ as δ_i is positive or negative. As the absolute value of each term on the right of (2.5) is at most $|\delta_i|$, $H_i(\cdot)$ is continuous as a function of c_i alone, and so is $G(\cdot)$.

From (2.5) we obtain

$$\{H_i(\mathbf{c} + \boldsymbol{\delta}) - H_i(\mathbf{c})\}/\delta_i + G(\mathbf{c} + \boldsymbol{\delta}) = -\operatorname{sgn}(\delta_i) \int_I (x_i - c_i)/\delta_i\, dF(\mathbf{x}). \tag{2.6}$$

As the integral is bounded by $|G(\mathbf{c}+\boldsymbol{\delta}) - G(\mathbf{c})|$, it follows on letting δ_i tend to zero that

$$\frac{\partial H_i}{\partial c_i} = -G(\mathbf{c}). \tag{2.7}$$

We can now differentiate (2.4) to see that $G(\cdot)$ is partially differentiable with respect to c_i and we obtain

$$\frac{\partial}{\partial c_i}(\log G) = -(A_{ii}+1)/e_i(c), \quad i = 1, \cdots, p. \tag{2.8}$$

It can now be seen that log G has continuous second-order partial derivatives. Thus the matrix of second derivatives is symmetric. Applying this symmetry to the derivatives of (2.8) we obtain

$$(A_{ii}+1)A_{ij}/e_i^2 = (A_{jj}+1)A_{ji}/e_j^2.$$

If $A_{ij} \neq 0$ it follows that $e_i g_j = e_j g_i$ and so $A_{jj} \neq 0$. Define $I_0 = \{i : A_{ii} = 0\}$. Then the relation on $\{1, \cdots, p\}\backslash I_0$ defined by $i \sim j$ if $A_{ij} \neq 0$ is an equivalence relation. Let the equivalence classes be $I_1, \cdots, I_q$. We can now deduce the existence of constants $d_j > 0$ for $j \in I_r$ with $r > 0$ and $a_r > 1$ for $1 \leq r \leq q$ such that $A_{ij} = g_i d_j$ and $A_{ii} = (a_r - 1)^{-1}$ whenever $i, j \in I_r$ with $r > 0$ and that $A_{ij} = 0$ otherwise.

Incorporating this into (2.8) we obtain

$$\frac{\partial}{\partial c_i}(\log G) = -a_r \frac{\partial}{\partial c_i}(\log u_r), \quad i \in I_r, \tag{2.9}$$

where $u_r = 1 + \sum_{i \in I_r} d_i(c_i - b_i)$. We also have

$$\frac{\partial}{\partial c_i}(\log G) = -1/g_i, \quad i \in I_0. \tag{2.10}$$

Solving (2.9) and (2.10) and using $G(\mathbf{b}) = 1$ yields the required result

$$G(\mathbf{c}) = \left(\prod_{r=1}^{q} u_r^{-a_r}\right) \prod_{i \in I_0} \exp\{-(c_i - b_i)/g_i\}.$$

Every distribution in Theorem 1 is a mixture of multivariate exponential distributions (Hutchinson, 1979). The mixing distribution is unique since it is determined by its Laplace transform which is readily seen to be $G(\mathbf{b}+\cdot)$. The univariate version of Theorem 1 was given by Revankar et al. (1974) and, in the case where only mixtures of exponential distributions are considered, by Morrison (1978).

A characterization of the multivariate Pareto distribution follows immediately from Theorem 1.

COROLLARY. *Let* $\mathbf{X}$ *be a random vector satisfying* $\mathbf{X} > \mathbf{b}$ *for some vector* $\mathbf{b}$ *with* $b_i > 0$ *for* $i = 1, \cdots, p$. *Then* $\mathbf{X}$ *has a multivariate Pareto distribution with exponent* $a > 1$ *if and only if*

$$E\{b_i^{-1}(X_i - c_i) \mid \mathbf{X} > \mathbf{c}\} = \lambda\{1 + \textstyle\sum_{j=1}^{p}(b_j^{-1}c_j - 1)\}, \quad i = 1, \cdots, p, \mathbf{c} > \mathbf{b},$$

for some constant $\lambda > 0$.

For an economic interpretation of this result note that if b_j is regarded as the subsistence level in the jth year then $b_j^{-1}(X_j - c_j)$ represents the excess of real income above the threshold c_j in that year and $1 + \sum_{j=1}^{p} b_j^{-1}(X_j - b_j)$ represents the real resources (gross income for the pth year plus net income saved from previous years) available for use in the pth year. Thus the multivariate Pareto distribution is characterized by the expected real value of excess of income over the threshold value being the same for each year and proportional to the resources from that threshold income available in the pth year.

By applying the proof of (2.7) to the mean residual lifetime of any multivariate distribution we obtain

$$\frac{\partial}{\partial c_i}(\log H_i) = -1/e_i(\mathbf{c})$$

if $G(\cdot)$ is a continuous function of c_i and a refinement of the argument gives analogous versions for left- and right-hand derivatives and limits in the general case. The theorem below follows readily.

THEOREM 2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be random vectors in* $\mathbb{R}^p$ *whose means exist. Then* $\mathbf{X}$ *and* $\mathbf{Y}$ *have the same distribution if and only if*

$$E(\mathbf{X} - \mathbf{c} \mid \mathbf{X} > \mathbf{c}) = E(\mathbf{Y} - \mathbf{c} \mid \mathbf{Y} > \mathbf{c}), \quad \mathbf{c} \in \mathbb{R}^p.$$

This is a multivariate generalization of the familiar univariate result that the mean residual lifetime determines the distribution. See, for example, Cox (1962, page 128), Swartz (1973), or Galambos and Kotz (1978, pages 30–35), where references to related characterizations are given.

Using Theorem 2 it can be shown that a distribution on an open set U has constant mean residual lifetime on U if and only if it has density proportional to $\exp(\boldsymbol{\theta}'\mathbf{x})$ on U and the recession cone of U contains the positive orthant. Distributions on convex sets with densities of this form were introduced by Blaesild (1978).

3. Note on estimation. For the multivariate Pareto distribution (1.1) with $p \geq 2$ the behaviour of the maximum likelihood estimators does not in general have the simple form given by Mardia (1962). If $\mathbf{x}_{(1)}$ denotes the vector of sample minima, then any MLE $\hat{\mathbf{b}}$ of $\mathbf{b}$ must lie on the boundary of the region $\mathbf{0} < \mathbf{b} < \mathbf{x}_{(1)}$; and in general the MLE's $\hat{a}$ and $\hat{\mathbf{b}}$ are not unique. Also, $\mathbf{x}_{(1)}$ is not sufficient for $\mathbf{b}$ and $\hat{a}$ and $\mathbf{x}_{(1)}$ are dependent, in contrast to the univariate case (Malik, 1970). As $n \to \infty$, $\mathbf{X}_{(1)} \to \mathbf{b}$ in probability and the expressions for $\hat{a}$ and $\hat{\mathbf{b}}$ given in Mardia (1962) hold asymptotically.

Acknowledgment. We are grateful to a referee for pointing out a deficiency in our original statement of Theorem 1 and for drawing our attention to several references.

REFERENCES

BARNDORFF-NIELSEN, O. (1977). Exponentially-decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London A* **353** 401–419.

BLAESILD, P. (1978). A generalization of the exponential distribution to convex sets in $\mathbb{R}^k$. *Scand. J. Statist.* **5** 189–194.

COX, D. R. (1962). *Renewal Theory.* Methuen, London.

DALLAS, A. C. (1976). Characterizing the Pareto and power distributions. *Ann. Inst. Statist. Math.* **28** 491–497.

GALAMBOS, J. and KOTZ, S., (1978). *Characterizations of Probability Distributions.* Lecture Notes in Mathematics, 675. Springer Verlag, Berlin.

GIBRAT, R. (1931). *Les Inégalités Economiques.* Sirey, Paris.

HUTCHINSON, T. P. (1979). Four applications of a bivariate Pareto distribution. *Biom. J.* **21** 553–563.

KRISHNAJI, N. (1970). Characterization of the Pareto distribution through a model of under-reported incomes. *Econometrica* **38** 251–255.

MALIK, H. J. (1970). Estimation of the parameters of the Pareto distribution. *Metrika* **15** 126–132.

MANDELBROT, B. (1961). Stable Paretian functions and the multiplicative variation of income. *Econometrica* **29** 517–543.

MARDIA, K. V. (1962). Multivariate Pareto distributions. *Ann. Math. Statist.* **33** 1008–1015. Correction *Ann. Math. Statist.* **34** 1603.

MORRISON, D. G. (1978). On linearly increasing mean residual lifetimes. *J. Appl. Probability* **15** 617–620.

REVANKAR, N. S., HARTLEY, M. J. and PAGANO, M. (1974). A characterization of the Pareto distribution. *Ann. Statist.* **2** 599–601.

SWARTZ, G. B. (1973). The mean residual lifetime function. *IEEE Trans Reliability* **R22** 108–109.

THE MATHEMATICAL INSTITUTE
UNIVERSITY OF ST. ANDREWS
NORTH HAUGH
ST. ANDREWS
SCOTLAND KY16 9SS

DEPARTMENT OF STATISTICS
UNIVERSITY OF LEEDS
LEEDS LS2 9JT
ENGLAND