RESEARCH REPORTS

Department of Statistics

University of Leeds

Leeds LS2 9JT

A characterization of the multivariate Pareto distribution

by

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

No. 1. June, 1980

* Now at St. Andrews University

Summary

A characterization of the multivariate Pareto distribution.

A random vector $\underset{\sim}{X}$ on $\underset{\sim}{X} \geq \underset{\sim}{b}$ whose mean exists has a shifted multivariate Pareto distribution if and only if $E(\underset{\sim}{X} - \underset{\sim}{c} \mid \underset{\sim}{X} \geq \underset{\sim}{c})$ is a non-constant linear function of $\underset{\sim}{c}$ on $\underset{\sim}{c} \geq \underset{\sim}{b}$. An interpretation in terms of income-distribution is suggested. Expressions are also given for the MLEs of the parameters of a multivariate Pareto distribution.

A CHARACTERIZATION OF THE MULTIVARIATE PARETO DISTRIBUTION.

BY P.E. JUPP[1] AND K.V. MARDIA

University of St. Andrews and University of Leeds

Running head: MULTIVARIATE PARETO DISTRIBUTION

1. Economic motivation.

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 distribution of type 1 introduced by Mardia (1962) may prove useful for this purpose.

Let the random vector $\underset{\sim}{X} = (X_1,\ldots,X_p)'$ denote income in p successive years. For vectors $\underset{\sim}{x} = (x_1,\ldots,x_p)'$ and $\underset{\sim}{y} = (y_1,\ldots,y_p)'$ we shall use $\underset{\sim}{x} \geq \underset{\sim}{y}$ to mean that $x_i \geq y_i$ for $i=1,\ldots,p$. Then the multivariate Pareto distribution has tail distribution function

$$(1\text{-}1) \qquad P(\underset{\sim}{X} \geq \underset{\sim}{c}) = \left(1 + \sum_{i=1}^{p} (b_i^{-1} c_i - 1)\right)^{-a} \qquad \underset{\sim}{c} \geq \underset{\sim}{b}$$

where $b_i > 0$ for $i=1,\ldots,p$. If $\underset{\sim}{X}$ has distribution (1.1) then each X_i

---

[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 1970 subject classifications: Primary 62H05, 62F10; Secondary 62P20.

Keywords and phrases: multivariate Pareto distribution, characterization theorem, maximum likelihood estimator.

has a Pareto distribution

$$P(X_i \geq c_i) = (b_i^{-1}c_i)^{-a}$$

and, 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.

It seems appropriate to interpret the scale factor b_i in (1.1) as the subsistence level in the i^{th} year. An income of X_i in the i^{th} year is then interpreted as corresponding to a real income of $b_i^{-1}X_i$. If we take $b_1 = \ldots\ldots = b_p = 1$ and so measure all incomes in real terms, then the marginal distributions of the X_i are all equal. The distribution of X_p conditional on $X_1,\ldots,X_{p-1}$ can then be used to construct a $(p-1)^{th}$-order stationary Markov process which could be used to model real income distribution in 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 $X_1,\ldots,X_p$ 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. Our emphasis is rather on savings and subsistence levels and this leads to a characterization of the multivariate Pareto distribution which generalises 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 distributions which have tail distribution functions

$$(2.1)\qquad P(\underset{\sim}{X} \geq \underset{\sim}{c}) = \left(1 + \sum_{i=1}^{p} d_i(c_i - b_i)\right)^{-a} \qquad \underset{\sim}{c} \geq \underset{\sim}{b}$$

where $d_i > 0$ for $i=1,\ldots,p$. These distributions arise also by conditioning the multivariate Pareto distributions (1.1). Note that the mean of a distribution in the family (2.1) exists if and only if $a>1$. The following theorem shows that the shifted p-variate Pareto distributions for which the mean exists are characterized as those income distributions for p successive years satisfying the following natural condition: the expected excess of income above a threshold level in any year is a non-constant linear (more precisely, affine) function of the threshold levels for the p years.

THEOREM. Let $\underset{\sim}{X}$ be a random vector satisfying $\underset{\sim}{X} \geq \underset{\sim}{b}$ for some vector $\underset{\sim}{b}$. Then $\underset{\sim}{X}$ has a shifted multivariate Pareto distribution with exponent $a>1$ if and only if

$$(2.2)\qquad E(\underset{\sim}{X} - \underset{\sim}{c} \mid \underset{\sim}{X} \geq \underset{\sim}{c}) = A\underset{\sim}{c} + \underset{\sim}{f} \qquad \underset{\sim}{c} \geq \underset{\sim}{b}$$

for a constant vector $\underset{\sim}{f}$ and a non-zero matrix A.

PROOF. If $\underset{\sim}{X}$ has a shifted multivariate Pareto distribution then a calculation shows that (2.2) 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 $i,j=1,\ldots,p$. To prove the converse we first define $\underset{\sim}{e}(\underset{\sim}{c}) = A\underset{\sim}{c}+\underset{\sim}{f}$ and $\underset{\sim}{g}=\underset{\sim}{f}-A\underset{\sim}{b}$ and then note that the existence of the conditional expectation in (2.2) implies that each component of $\underset{\sim}{e}(\underset{\sim}{c})$ is positive for $\underset{\sim}{c}\geq\underset{\sim}{b}$. Thus $g_i>0$ and $A_{ij}\geq 0$ for $i,j=1,\ldots,p$. We define

$$G(\underset{\sim}{c}) = P(\underset{\sim}{X}\geq\underset{\sim}{c}) \qquad F(\underset{\sim}{c}) = 1-G(\underset{\sim}{c})$$

$$H_i(\underset{\sim}{c}) = \int_{\underset{\sim}{X}\geq\underset{\sim}{c}} (x_i-c_i)\, dF(\underset{\sim}{x}) \qquad i=1,\ldots,p.$$

Then (2.2) can be rewritten as

$$\text{(2.3)} \qquad H_i(\underset{\sim}{c}) = G(\underset{\sim}{c})\, e_i(\underset{\sim}{c}) \qquad i=1,\ldots,p .$$

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

$$\text{(2.4)} \qquad H_i(\underset{\sim}{c}+\underset{\sim}{\delta}) - H_i(\underset{\sim}{c}) = -\int_I (x_i-c_i)dF(\underset{\sim}{x}) - \delta_i G(\underset{\sim}{c}+\underset{\sim}{\delta})$$

where the region of integration I is defined by $c_j\leq x_j$ for $i\neq j$ and either $c_i\leq x_i<c_i + \delta_i$ or $c_i+\delta_i\leq x_i<c_i$ as δ_i is positive or negative. As the absolute value of each term on the right of (2.4) is at most $|\delta_i|$, $H_i(.)$ is continuous as a function of c_i alone, and so is $G(.)$.

From (2.4) we obtain

$$\text{(2.5)} \qquad (H_i(\underset{\sim}{c}+\underset{\sim}{\delta}) - H_i(\underset{\sim}{c}))/\delta_i + G(\underset{\sim}{c})$$

$$= -\int_I (x_i-c_i)/\delta_i \, dF(\underset{\sim}{x}) - G(\underset{\sim}{c}+\underset{\sim}{\delta}) + G(\underset{\sim}{c}).$$

As the integral is bounded by $F(\underset{\sim}{c}+\underset{\sim}{\delta}) - F(\underset{\sim}{c})$, it follows on letting δ_i tend to zero that

$$\text{(2.6)} \qquad \frac{\partial H_i}{\partial c_i} = - G(\underset{\sim}{c}).$$

We can now differentiate (2.3) to see that $G(.)$ is partially differentiable with respect to c_i and we obtain

$$(2.7)\qquad \frac{\partial}{\partial c_i}(\log G) = -(A_{ii}+1)/e_i(\underset{\sim}{c}) \qquad i=1,\ldots,p.$$

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 (2.7) we can deduce that $A = \underset{\sim}{g}\,\underset{\sim}{d}'$ and $A_{ii} = (a-1)^{-1}$ $i=1,\ldots\ldots,p$ for some vector $\underset{\sim}{d}$ with positive components and some scalar $a>1$. Incorporating this into (2.7) we obtain

$$(2.8)\qquad \frac{\partial}{\partial c_i}(\log G) = -a\frac{\partial}{\partial c_i}(\log u) \qquad i=1,\ldots,p$$

where $u = 1+\underset{\sim}{d}'(\underset{\sim}{c}-\underset{\sim}{b})$. Solving (2.8) and using $G(\underset{\sim}{b})=1$ yields $G(\underset{\sim}{c}) = u^{-a}$ as in (2.1).

A characterization of the multivariate Pareto distribution follows immediately.

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

$$E(b_i^{-1}(X_i-c_i)\,|\,\underset{\sim}{X}\geq\underset{\sim}{c}) = \lambda(1+\sum_{i=1}^{p}(b_i^{-1}c_i-1)) \qquad i=1,\ldots,p$$

for some constant λ.

For an economic interpretation of this result note that $b_i^{-1}(X_i-c_i)$ represents the excess of real income above the threshold in the i^{th} year and that $1+\sum_{i=1}^{p}b_i^{-1}(X_i-c_i)$ represents the real resources (gross income for the p^{th} year plus net income saved from previous years) available for use in the p^{th} 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 p^{th} year.

For any multivariate distribution we can define

$$\underset{\sim}{e}(\underset{\sim}{c}) = E(\underset{\sim}{X} - \underset{\sim}{c} \mid \underset{\sim}{X} \geq \underset{\sim}{c}) \qquad \text{for } G(\underset{\sim}{c})>0.$$

The proof of (2.7) shows that

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

if G(.) 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. It follows that a distribution is determined by $\underset{\sim}{e}(.)$ and its support, provided that the distribution is concentrated on an open connected set inside its support. This is a multivariate generalisation of the familiar univariate result (see for example Cox (1962), page 128) that the expectation of life function determines the distribution of life-times. The above theorem concerns the particular case in which $\underset{\sim}{e}(\underset{\sim}{c})$ is linear. Another case of interest is given by taking $\underset{\sim}{e}(\underset{\sim}{c})$ to be constant. This characterizes the generalized exponential distributions with densities proportional to $\exp(\underset{\sim}{\theta}'\underset{\sim}{x})$ which were introduced by Blaesild (1978).

3. Maximum likelihood estimation. For the univariate Pareto distribution with tail distribution function $P(X \geq c) = (c/b)^{-a}$ the maximum likelihood estimators $\hat{a}$ and $\hat{b}$ of the unknown parameters a and b take a particularly simple form. Indeed for a sample $x_1, \ldots, x_n$ of size n we have $\hat{b} = x_{(1)} = \min(x_1, \ldots, x_n)$ and $\hat{a} = (\prod_{i=1}^{n} x_i)^{1/n}$. Further Malik (1970) showed that $\hat{a}$ and $\hat{b}$ are independently distributed, that $\hat{b}$ has a Pareto distribution with exponent na, and that $2na/\hat{a}$ has a chi-squared distribution with 2(n-1) degrees of freedom.

For the multivariate Pareto distribution (1.1) with $p \geq 2$ the behaviour of the MLEs is more complicated and does not in general take the simple form given by Mardia (1962). For a sample of p-vectors $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ let $\underset{\sim}{x}_{(1)}$ be the vector with j^{th} component $x_{(1)j} = \min(x_{1j}, \ldots, x_{nj})$ where x_{ij} denotes the j^{th} component of $\underset{\sim}{x}_i$. Then $\underset{\sim}{x}_{(1)}$ is not sufficient for $\hat{\underset{\sim}{b}}$ in contrast to the univariate case. Thus $\hat{a}$ and $\underset{\sim}{x}_{(1)}$ are dependent. Define

$$\xi_j(\underset{\sim}{b}) = n^{-1} \sum_{i=1}^{n} x_{ij} b_j^{-1} / (1 + \sum_{k=1}^{p} (x_{ik} b_k^{-1} - 1)) \qquad j=1,\ldots,p$$

and $$\zeta(\underset{\sim}{b}) = n^{-1} \sum_{i=1}^{n} \log(1 + \sum_{k=1}^{p} (x_{ik} b_k^{-1} - 1)).$$

Differentiation of the likelihood function derived from (1.1) shows that at a stationary value of likelihood $\xi_j(\underset{\sim}{b}) = 1/(a+p)$ for $j=1,\ldots,p$. Adding these equations leads to a contradiction, so $\hat{\underset{\sim}{b}}$ does not lie in the interior of the region given by $\underset{\sim}{b} \leq \underset{\sim}{x}_{(1)}$ and $b_i > 0$ for $i=1,\ldots,p$. In general $\hat{a}$ and $\hat{b}$ are given implicitly by

$$\xi_j(\hat{\underset{\sim}{b}}) = 1/(\hat{a}+p) \qquad \text{if } \hat{b}_j < x_{(1)j}$$

$$\zeta(\underset{\sim}{b}) = \sum_{k=1}^{p} 1/(\hat{a}+k-1).$$

However, using the fact that for the multivariate Pareto distribution (1.1) the expected value of $x_j b_j^{-1} / (1 + \sum_{k=1}^{p} (x_k b_k^{-1} - 1))$ is $(a+1)/(a+p)$, it can be shown that asymptotically $\hat{a}$ and $\hat{\underset{\sim}{b}}$ given by

$$(3.1) \qquad \hat{\underset{\sim}{b}} = \underset{\sim}{x}_{(1)} \qquad \zeta(\underset{\sim}{x}_{(1)}) = \sum_{k=1}^{p} 1/(\hat{a}+k-1)$$

are MLEs. In the bivariate case it follows from the the fact that $\xi_j(.)$ is a decreasing function of b_j that asymptotically $\hat{a}$ and $\hat{\underset{\sim}{b}}$ given by (3.1) are the unique MLEs, in agreement with the results given by Mardia (1962).

It can also be shown that MLEs for a and $\underset{\sim}{b}$ always exist but are not in general unique.

REFERENCES

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

[2] BLAESILD, B. (1978). A generalization of the exponential distribution to convex sets in $\mathbb{R}^k$. Scand. J. Statist. 5 189-194.

[3] COX, D.R. (1962). Renewal Theory. Methuen, London.

[4] GIBRAT, R. (1931). Les inegalites economiques. Sirey, Paris.

[5] HUTCHINSON, T.P. (1979). Four applications of a bivariate Pareto distribution. Biom. J. 21, 553-563.

[6] KRISHNAJI, N. (1970). Characterisation of the Pareto distribution through a model of under-reported incomes. Econometrica 38 251-255.

[7] MALIK, H.J. (1970). Estimation of the parameters of the Pareto distribution. Metrika 15 126-132.

[8] MANDELBROT, B. (1961). Stable Paretian functions and the multiplicative variation of income. Econometrica 29, 517-543.

[9] MARDIA, K.V. (1962). Multivariate Pareto distributions. Ann. Math. Statist. 33 1008-1015. Correction Ann. Math. Statist. 34 1603

[10] REVANKAR, N.S., HARTLEY, M.J. & PAGANO, M. (1974). A characterization of the Pareto distribution. Ann. Statist. 2 599-601.