The Philippine Statistician
1990, Vol. 39, Nos. 1-2 & 3-4
pages 61-70

MEASURES OF PEAKEDNESS IN PARETO MODEL

by

Syed Afroduzzin Ahmed

Masood-Ul-Haq

and

Anwer Khurshid

University of Karachi
Karachi-32
Pakistan

Abstract

Horn (1983) defined the measure of Peakedness for all symmetric unimodal densities. In this article, an easy modification of Horn's measure of Peakedness is being used for Pareto density. A procedure is also proposed to obtain the Horn's modified measure of Peakedness when sample is available.

Keywords:

kurtosis, Pareto distribution, peakdeness

1. INTRODUCTION

The distributional shape of probability distributions have been studied through the concepts of kurtosis, peakedness and tail weight. Kurtosis has been described by Mosteller and Tukey (1977) as a "vague concept" The oldest and most commonly used definition of kurtosis $\beta_2(F)$ of a probability distribution F is:

$$\beta_2 (F) = \mu_4(F) / (\mu_2(F))2 \qquad (1)$$

where $\mu_\alpha(F)$ stands for the α^{th} central moment of the distribution F The value of kurtosis for some well studied symmetric distribution are given in Table (1).

Table (1)

Kurtosis for some symmetric distributions

Distribution	Kurtosis
Normal	3.000
t(6)	6.000
Double Exponential	6.000
t(5)	9.000
Cauchy	Does not exist

It is clear from Table (1), that kurtosis does not exist for the Cauchy distribution although it is more peaked than the normal which has a kurtosis value of 3. Besides, kurtosis value for the double exponential distribution is 6 and that for t(6) is also 6, therby giving an impression that these two curves are equally peaked. This is not wholly so. Due to these limitations of the kurtosis as a measure of peakedness of a probability distribution, some other measures of peakedness have been suggested in current research journals.

2. ALTERNATIVE MEASURES OF KURTOSIS

Groeneveld and Meeden (1984) proposed a number of alternate measures of Kurtosis that have natural interpretations for symmetric distributions in terms of the movement of probability mass from the shoulders of a distribution into its centre or tails. They suggested that for each α in (0, 1/4) the quantity $\beta_2(\alpha, F)$ defined by:

$$\beta_2(\alpha,F) = F^{-1}(7.5+\alpha)-F^{-1}(7.5-\alpha)/F^{-1}(.75+\alpha) - F^{-1}(.75-\alpha) \quad (2)$$

measures the kurtosis of the symmetric distribution.

Another family of quantile-based measures that has been discussed by Balanda (1986) and Ruppert (1987) is $t_P(f)$ for a probability distribution F for $0 < p < 1/2$ and defined by:

$$t_P(F) = \frac{F^{-1}(.5 + p) - F^{-1}(.5 - p)}{F^{-1}(.75) - F^{-1}(.25)} \quad (3$$

Hogg (1974) proposed adaptive location estimators that used statistics like:

$$Q = [U(.2) - L(.2)] / [U(.5) - L(.5)] \quad (4)$$

where U(d) and L(d) denote the average of the largest and smallest 100xd% of the sample. For probability distributions f(x):

$$U(.2) = \int_{q.8}^{\infty} x\, f(x)\, d(x) \quad \text{and}$$

$$L(.2) = \int_{-\infty}^{q.2} x\, f(x)\, d(x) \quad (5)$$

These are thefore "tail means" and measures the weight of the upper and lower tails of a p.d.f. f(x).

Horn (1983) suggested that for $0 < p < 1/4$, the quantity $mt_p(f)$ =

$$1- \{p / (h(m_F) (F^{-1} (.5+p) - m_F)\} \quad (6)$$

be used as a measure of peakedness for a symmetric unimodal density f(x).

In this present paper, the method suggested by Horn (1983) is used with some modifications. The modifications are essential, since Pareto distribution which has been discussed here is a right skewed distribution. There is another difficulty in the method suggested by Horn (1983), and that is that it gives a measure of peakedness at each value of the random variable, unlike kurtosis, which is a single value for the entire probability distribution. Due to this, a comparison of peakedness for different distributions are performed point by point. In this paper a procedure is proposed to obtain Horn's modified measure of peakedness when a sample is available.

3. HORN'S MEASURE OF PEAKEDNESS

Suppose $f(x)$ is a symmetric, unimodal, probability density function, defined on $(-\infty, \infty)$, and let $F(x)$ be the distribution function. Consider the rectangale in the x-y plane formed by the lines: $x = 0$, $y = 0$, $y = f(0)$, zero being the point of symmetry; and $x = F^{-1}(p + 0.5)$, for some p, in the interval $0 < p < 0.5$. The area in the rectangle is denoted by $A_p(f)$ which is:

$$A_p(f) = f(0) \cdot F^{-1}(p + 0.5).$$

Also the area of the portion of the p.d.f. $f(x)$, which lies inside this rectangle is equal to:

$$p = F(p + 0.5).$$

Thus a measure of peakedness would be: (see fig. 1)

$$m_p(f) = 1 - \frac{p}{A_p(f)} \qquad (7)$$

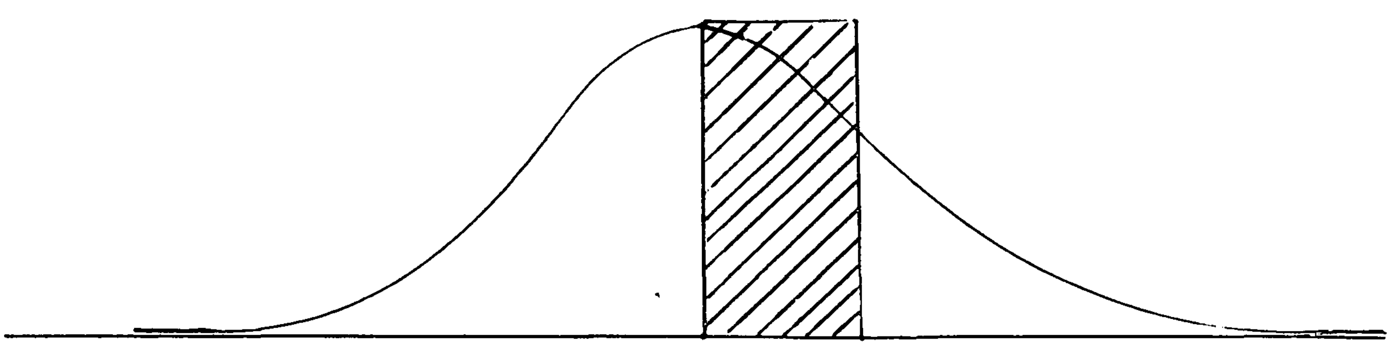


Fig. 1

If $p/A_p(f)$ is close to 1, then most of the density is under the rectangle and therefore the p.d.f. $f(x)$ is not very peaked at this value of $x(= F^{-1}(p + .5))$. The measure $m_p(f)$ has the desirable property that it lies between 0 and 1.

4. MODIFICATION OF HORN'S MEASURE

Hórn's measure of peakedness may be easily extended if we consider a right skewed distribution that exists only in the first quadrant, like the Pareto model which has been discussed here (see fig.2).

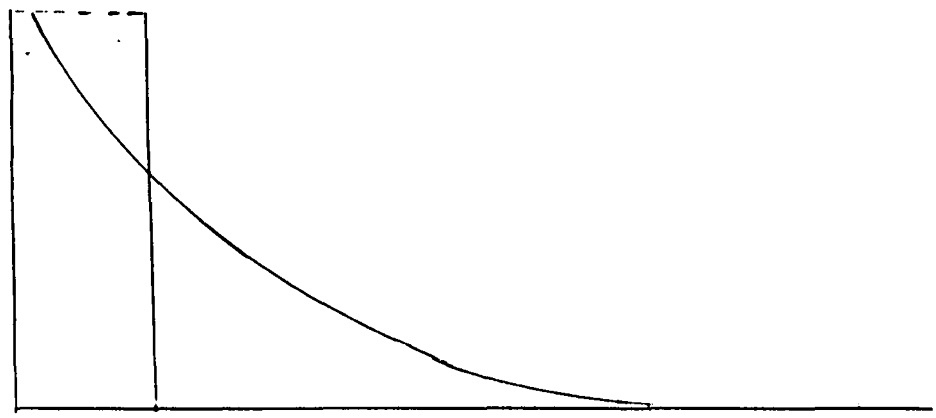


Fig. 2

In such distributions, the mode, instead of the point of symmetry, may be considered as the starting point, so that the sides of the rectangles are:

$$x = a,\ y = 0,\ y = f(a),\ \text{and}\ x = F^{-1}(p).$$

The measure of peakedness $m_p(f)$ is:

$$m_p(f) = 1 - \frac{p}{f(a)\quad F^{-1}(p)} \qquad (8)$$

For the Pareto distribution:

$$f(x) = v \cdot a^{v}\ x^{-v-1}, \quad x>a,\ a>0 \quad (9)$$

The mode is at x = a, hence, f(a) = v/a, so that the area is

$$A_p(f) = (v/a)\ (x-a),$$

where $x = 1(1-p)^{-(1/V)}$, and $p = F(x)$.

The value of $m_p(f)$ is then

$$1 \quad \frac{p}{v(\ (\ 1-p)^{-(1/V)} - 1)} \qquad (10)$$

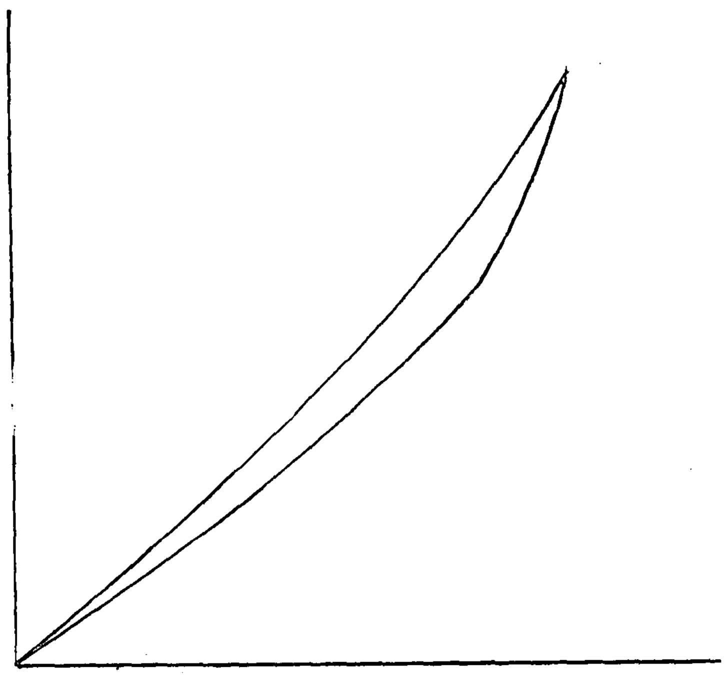
Table (2) shows the values of $m_p(f)$ for different values of v and p.

Table (2)

The values of $m_p(f)$ for different values of p and v = 1.5 in a Pareto distribution.

p	$m_p(f)$ for v=1.5	$m_p(f)$ for v-=2.5
0.10	0.08387	0.07074
0.15	0.12617	0.10671
0.20	0.16878	0.14321
0.25	0.21169	0.18002
0.50	0.43256	0.37403
0.75	0.67104	0.59520
1.00	1.00	1.00

The values of $m_p(f)$ are graphically shown in fig(3).



5. MEASURING $m_p(f)$ FROM SAMPLE:

For the determination of an estimator of $m_p(f)$, when a random sample of size n is available from a Pareto distribution, we first calculate the estimator of v. The maximum likelihood estimator of v is:

$$\hat{v} = [\log (\frac{g}{X_{(1)}})]^{-1} \qquad (11)$$

where g is the sample geometric mean and $x_{(1)}$ is the smallest ordered variable, for the given random sample:

$x_1, x_2, \ldots\ldots\ldots, x_n$, and the maximum likelihood estimator

of a is $\hat{a} = x_{(1)}$. The estimators of v and a generate the Pareto model which may be written as:

$$f(x) = v\, a^v\, x^{-v-1}\ , \ x > a. \qquad (12)$$

The estimator of the peakedness measure $m_p(f)$ is then:

$$\hat{m}_p(f) = 1 - \frac{\hat{p}}{\hat{v}\,[\,(1-\hat{p})^{-(1/\hat{v})}-1]} \qquad (13)$$

where v is obtained form $[\log(g/x_{(1)}]^{-1}$ and p can be obtained corresponding to each $x_{(i)}$ in $x_{(1)} < x_{(2)} < \ldots\ldots < x_{(n)}$.

Thus:

$$\hat{p} = \int_{\hat{a}}^{x_{(i)}} \hat{v}\, \hat{a}^{\hat{v}}.\ x^{-\hat{v}-1}\, dx$$

$$= 1 = (\frac{\hat{a}}{x_{(i)}})^{\hat{v}}$$

$$= 1 - (x_{(1)}/x_{(i)})^{\hat{v}} \qquad (14)$$

As a numerical example, suppose a random sample from Pareto distribution is:

5.4, 10.10, 19.37, 11.96, 14.57, 5.26, 7.84, 7.49, 5.39, 37.10

From the sample:

$\hat{a} = x(i) = 5.26$
$x_{(.25)} = 5.395$

$\hat{p} = 0.03793$

$\hat{v} = 1.52599$

$\hat{m}_p(f) = 0.03144$

6. MONTE-CARLO METHOD FOR $m_p(f)$:

The performance of the measures $m_p(f)$ is now discussed by drawing 500 random samples from Pareto distributions with a parameters (a, v) = (1.5, 1.5), (1.5, 2.0) (1.5, 2.5) (1.5, 3.0). Each random sample is of size 83. The value of $\hat{m}_p(f)$ for $x_{(i)}$ = $x_{(21)}$ is obtained, where $x_{(21)}$ is the 1st quartile of the of the sample. The results are presented in Table(3). The random samples were generated on a Vax 2000 computer. In all 500 * 83 random numbers were required and the computer time needed was approximately two and half minutes. The computer program is available from the authors. It is to be noted that $m_p(f)$ can be obtained for each $x_{(i)}$, i = 1,2,..., 83. However, the ordered values of extremely high rank are far away from the peak of the Pareto distribution and hence may not be so useful in elaborating the peakedness of the distribution. The ordered statistic $X_{(i)}$ = $X_{(.25n)}$ is therefore selected for the interpretation of the peakedness.

Table (3)

Pecentage of values of mp(f) for X(21) lying inside given intervals.
Sample size = 83.

	a=1.5, v=1.5	a=1.5, v=2.0	a=1.5, v=2.5	a=1.5, v=3
mp(f)	% freq.	% freq. %	% freq.	% freq.
0.00--0.25	56.6	66.7	74.6	80.8
0.25--0.50	28.2	24.0	19.2	15.4
0.50--0.75	10.8	7.8	5.4	3.4
0.75--1.00	4.4	1.4	0.8	0.4

It becomes clear form the table that as the parameter v increases from 1.5 to 3.0 the estimator $m_p(f)$, for p=.25 has higher percentages of its values lying within the interval (0,0.25), showing the consistent behaviour of $m_p(f)$ as obtained in (13). It also shows a decrease in the peakedness as v increases.

7. CONCLUSION

The expression $m_p(f)$, for the Pareto distribution is:

$$m_p(f) = 1 - [\ p/v\ \{\ (1-p)^{-(1/v)} - 1\}\ ] \qquad (10)$$

and it gives the index of peakedness for given values of p and v. If, however, a sample is given from a Pareto distribution, then an estimator of $m_p(f)$ can be obtained and the value will indicate the level of peakedness for different values of $X_{(1)}$.

REFERENCES :

Balanda, K. P. (1986). Kurtosis, Peakedness and Tailweight. *M.Sc. Thesis,* University of Queensland, Department of Mathematics.

Groeneveld, R. A. and Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician,* **33**, 391-399.

Hogg, R.V. (1974). Adaptive robust procedures: a Partial review and some Suggestions for future applications and theory. *Journal of the American Statistical Association,* **69**, 909-927.

Horn, P.S. (1983). A measure of peakedness. *The American Statistician*, **37**, 55-56.

Mosteller, F. and Tukey, J.W. (1977). *Data Analysis and Regression,* Reading, M.A.: Addison Wesley.

Ruppert, D. (1987). What is kurtosis? *The American Statistician*, **41**-1-5.