MODELLING OF ECONOMICALLY ACTIVE POPULATIONS USING SKEW ELLIPTICAL DISTRIBUTIONS

Olosunde. A.
Department of Statistics
University of Agriculture P.M.B 2240
Abeokuta, Nigeria
e-mail:- olosundeaa@unaab.edu.ng.

Abstract
An area of distribution theory attracting a lot of contributions and developments, which recent literatures have shown to have applications in modelling environmental and epidemiological data amongst others is the class of skew elliptical distributions. In this paper skew-ness to Kotz-type symmetric distribution was introduced and some mathematical/statistical properties of this distribution were presented. The distribution was applied to univariate skewed data of Nigerian economically active population and its gave a satisfactory fitting.

Introduction

A large body of evidence shows that acute neglect of the set of population that plays an active role in economic activities of a nation has resulted into many crises that could have been averted if attention was given to it. In line with African governments' plans to meet the millennium goals, it is imperative for every country to be able to predict its economically active population (for the purpose of planning). This class of the population constitute the major players in the economic growth of any nation, their neglect has resulted in high crime rates and social vices in both developing and developed nations.

In most literature published by various authors, much efforts were concentrated on modelling of criminal activities (Ojo, 1989). This paper adopted a different approach to addressing stated problem. Specifically, its examined the possibility of extending some of the classical methods of skewing original symmetric distribution (Fernandez and Steel, 1998) for modelling economically active population.

In statistical modelling, the symmetric Kotz-type distribution has played an important role in the class of elliptically contour distributions, the distribution which was first introduced by Kotz (1975) has been studied by many researchers, among whom are Iyengar and Tong (1989), Fang et al (1990) and Kotz and Nadarajah (2001).

Some other methods of introducing skewing mechanism to original symmetric distributions were discussed by Azzalini and Dalle-Valle (1996), Jones (2004) etc.

We demonstrated that this new distribution has reasonable flexibility in real data fitting, while it maintains some convenient formal properties of the Kotz-symmetric distribution. More specifically, the objectives of the paper are:

1. To extend skewness using inverse scale method to Kotz-symmetric distribution, especially when they reproduce or resemble similar properties of the symmetric type;
2. To examine the potential applications of this distribution in statistics, with special emphasis on modelling economically active population (Nigeria as a case study)

Definition 1

An n-dimensional random vector $\psi=(\psi_1,\ldots,\psi_n)^T$ has a Kotz-type distribution if its joint density is given by

$$K_n |\Sigma|^{-1/2}[(x-\mu)^T \Sigma^{-1}(x-\mu)]^{k-1} \exp\{-r[(x-\mu)^T \Sigma^{-1}(x-\mu)]\}, \qquad (1)$$

where $\mu^T=(\mu_1,\ldots,\mu_n)$, $r>0$, $k=1,2,\ldots$, Σ is a positive definite $n\times n$ constant matrix, and

$$K_n = \frac{\Gamma(n/2)}{\pi^{n/2}\Gamma((2k+n-2)/2)} r^{(2k+n-2)/2}$$

is a normalizing constant. Let $\xi \sim Kz_n(\mu,\Sigma,r,k)$ such that when $k=1$ and $r=1/2$, the distribution reduces to a multivariate normal distribution. The characteristics functions of ξ is

$$\exp(it^T\mu)\psi(t^T\sum t,r,k), \quad (2)$$

where

$$\psi(u,r,k)=\exp(-u/4r)\sum_{m=0}^{k-1}\binom{k-1}{m}\frac{\Gamma(n/2)}{\Gamma(n/2+m)}(-\frac{u}{4r})^m$$

for example see Iyengar and Tong (1989).

We derived asymmetric Kotz-type distribution using a method called inverse scale method introduced by Fernandez and Steel (1998). The asymmetric distribution assumes a new distribution but retains its nice original symmetric properties and possessed additional new properties of flexibility in skewness and kurtosis

Asymmetric Kotz-type distribution
Definition 2
Consider the following class of density, we say f(.) is the skewed version of the original symmetric Kotz-type distribution if f(.) is given by

$$f(y\,|\,\gamma,\zeta)=\frac{2}{\zeta+\frac{1}{\zeta}}\{g(\zeta(y-\gamma)\,|\,0,\delta)I_{(-\infty,\gamma)}(y)+g((y-\gamma)/\zeta\,|\,0,\delta)I_{[\gamma,\infty)}(y)\}, \quad (3)$$

with

$$g(y\,|\,\gamma,\delta)=\frac{1}{\delta}h(\frac{y-\gamma}{\delta}) \quad (4)$$

where h(.) is a unimodal density and symmetric about 0 with existing first and second derivatives at 0. Thus f(.) is the skewed version of a location-scale density with the same mode as g(y;γ) and a skewness parameter $\zeta>0$ and such that the ratio of probability masses above and below the mode is

$$\frac{\Pr(Y\geq\gamma\,|\,\varsigma)}{\Pr(Y<\gamma\,|\,\varsigma)}=\varsigma^2 \quad (5)$$

Note that the density f(y|γ,1/ζ) is the symmetric form of f(y|γ,ζ) with respect to mode. Therefore, working with $\zeta^{\text{l}}=\log(\zeta)$ might be preferable to indicate the sign of the skewness.
Taken g(y;γ) to be Kotz-type distribution, we obtained additional parameters having the interpretation:

- μ, as the mode, models the location,
- $\zeta>0$, models the skewness
- $\delta^2>0$, (which is not the variance anymore) is a dispersion parameter which measures the inverse curvature of the density at the mode. So we have

$$(-\frac{\partial^2\log f(y\,|\,\gamma,\delta,\varsigma)}{\partial_L y^2})_{|y=\gamma}^{-1}=\frac{h(0)}{-h''(0)}\frac{1}{\varsigma^2}\delta^2$$

and (6)

$$(-\frac{\partial^2\log f(y\,|\,\gamma,\delta,\varsigma)}{\partial_R y^2})_{|y=\gamma}^{-1}=\frac{h(0)}{-h''(0)}\varsigma^2\delta^2$$

as inverse curvature measure on, respectively, the left and the right of the mode γ (with ∂/∂_L *and* ∂/∂_R denoting respectively left and right derivatives).
The basic idea in the distribution (3) above is to introduce inverse scale factors in the positive and the negative half real lines, then ζ is scalar in (0,∞). If the skewness parameter is unity, we retrieve the original symmetric density.

The mode of the density is unchanged, remaining at zero irrespective of the particular value of ζ, we know that this method always transfers moment characteristics from its original symmetric distribution to its new skewed version. Selecting ζ to be larger (smaller) than unity corresponds to right (left) skewed distributions.

Mathematical properties

In what follows we derived some mathematical properties of this distribution which includes the quantile function, cumulative distribution function and characteristics function (cf) of the distribution.

Definition 3

Assume that $f(y\,|\,\gamma,\delta,\varsigma)$ is a density skewed using the technique proposed by Fernandez and Steel (1998) is applied to Kotz-type distribution which is a continuous and symmetric location-scale density $g(y\,|\,\gamma,\delta)$ with location and dispersion parameters γ and δ. We can relate the cumulative distribution function (cdf) F and the quantile function F^{-1} to the starting cdf of G and quantile function G^{-1}. We have

$$F(y\,|\,\gamma,\delta,\varsigma)=\frac{2}{1+\varsigma^2}G(\varsigma(y-\gamma)\,|\,0,\delta) \qquad if\ \ y<\gamma$$

$$=1-\frac{2}{1+\varsigma^{-2}}G(-\varsigma^{-1}(y-\gamma)\,|\,0,\delta)\ \ if\ \ y\geq\gamma$$

for the cdf and

$$F^{-1}(p\,|\,\gamma,\delta,\varsigma)=\frac{1}{\varsigma}G^{-1}\left(\frac{p}{2}(1+\varsigma^2)\,|\,\gamma,\delta\right)\ \ if\ \ p<\frac{1}{1+\varsigma^2}$$

$$=-\varsigma G^{-1}\left(\frac{1-p}{2}(1+\varsigma^{-2})\,|\,\gamma,\delta\right)\ if\ p\geq\frac{1}{1+\varsigma^2}$$

for the quantile function

Also let $\xi\sim SKz_n(\mu,\Sigma,\varsigma,r,k)$ where SK symbolizes skew kotz type distribution then the characteristics function (cf) is given by;

$$\frac{2}{\varsigma+\frac{1}{\varsigma}}\{\exp(it^T\mu)\Psi(\varsigma t^T\Sigma t,r,k)+\exp(-it^T\mu)\Psi(t^T\Sigma t/\varsigma,r,k)\}$$

where (6)

$$\Psi(u,r,k)=\exp(-u/4r)\sum_{m=0}^{k-1}\binom{k-1}{m}\frac{\Gamma(n/2)}{\Gamma(n/2+m)}\left(-\frac{u}{4r}\right)^m$$

From expression (6) above, we can easily estimate the mean, variance, skewness and the kurtosis of this new distribution by differentiating and setting t=0.

Univariate distribution

Special case (skew version of the normal distribution)

For the purpose of this study we take T=1 in (1) above to narrow down our study to univariate asymmetric Kotz type distribution, with appropriate choice of values of the parameters such that when k=1 and r=1/2 its gives the skewed version of the normal distribution with three parameters SN(μ,σ,ζ) where,

- μ is the mode of the distribution
- σ is the dispersion parameters
- ζ is the shape parameter such that if ζ=1 we have original symmetric normal distribution and when positive (negative) the distribution skew to the right (left).

Note also that the density $f(.|1/\zeta)$ is the mirror of $f(.|\zeta)$ with respect to the (zero) mean and unit variance i.e $f(.|1/\zeta)=f(.|\zeta)$

Application

The MLE (standard errors or s.e.'s) of the parameters in the skew normal models for the data in the Table 4.1 below can be found for normal model for the same location and scale parameters and for the skew version which has an additional shape parameter ζ. A better fit is obtained with the skew version of the normal model with a minus log-likelihood of 1546.0 against 4556.0 for the normal distribution. The location parameters for both models are similar, since we are dealing with mode of the distributions. The skew model gives a better fit for the data because it has a shape parameter which account for the positive skew ness of the data which is lacking in the usual normal model.

Conclusion and Discussion.

From our analysis and also as shown in the Figure 4.1. it is clear that our skewed model gave good fit to the histogram of age grouping of economically active population y. Furthermore, the empirical cumulative distribution function (ecdf) in fig. 4.2, the smooth curve line shows the fitted empirical cumulative normal distribution and the dotted line is the observed values, the distance from the observed values and the fitted model is the deviation from the normal distribution and the Normal q-q plot shows a wide deviation of the data from normality despite large sample involved, the straight line is the fitted normal model while the dotted line is the observed values (in millions).

From the data analysis above we conclude that the skew version of the normal distribution can be used as a better replacement for a normal distribution when a model for predicting economically active population is required. Despite the large sample involved which asymptotic theory will assume normality but in this case, a noticeable deviation from normality is seen from the ecdf in fig 4.2 and normal q-q plot in fig. 4.3. In other words, not every case involving large sample do the assumption of the asymptotic theory of normality works, there are cases such as the one we examined in this paper that the theory has shown a clear failure. In such cases like this we use the skew version as possible replacement for the normal distribution.

Fig. 4.1. Histogram (probability plot) of economically active population (y) with skew-normal density fitted to the data ,ζ=1.5. .

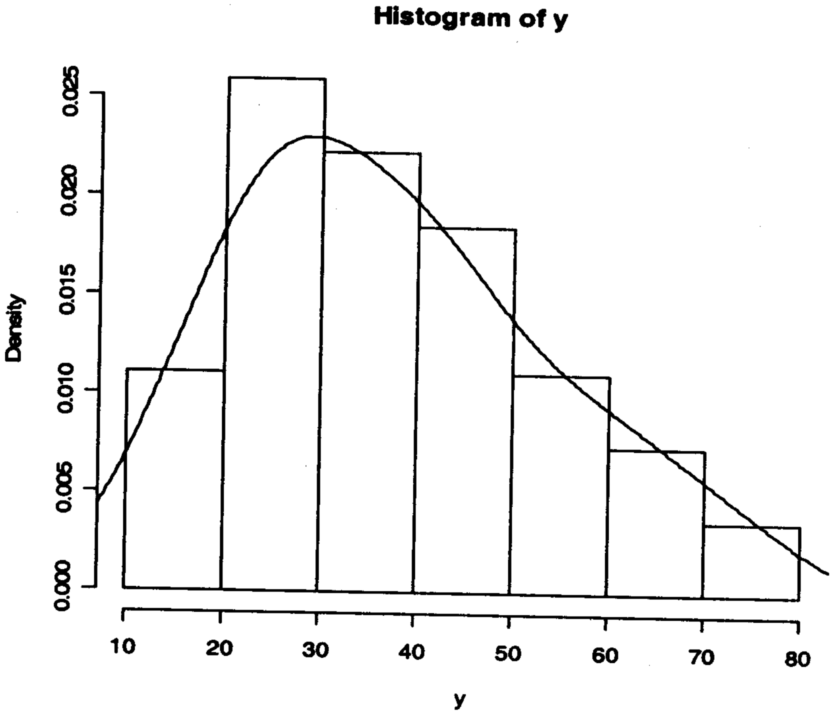


Fig.4.2 The empirical cumulative distribution function fitted to the data in Table 4.1

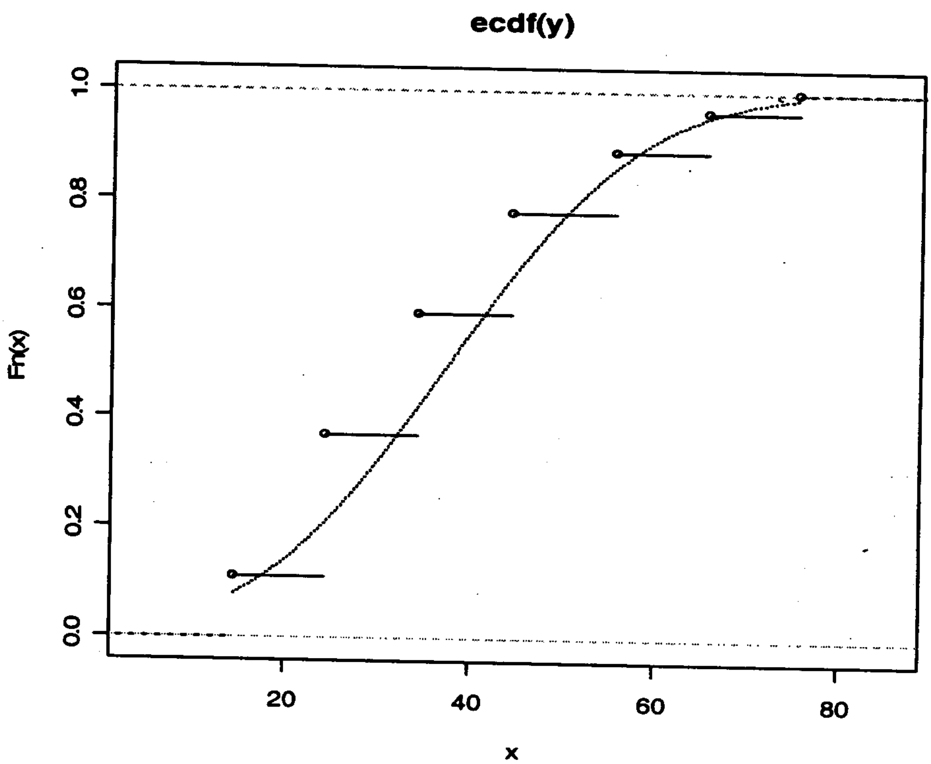
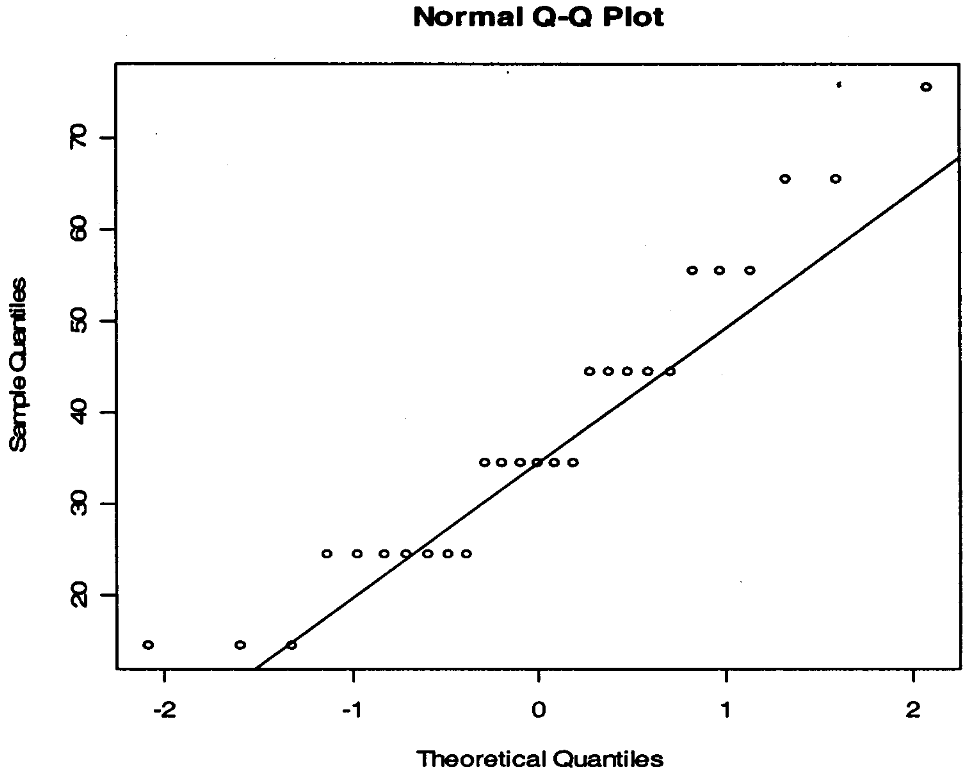


Fig. 4.3. The normal q-q plot which shows the deviation of the data from normality



The author is grateful to Dr. A.O. Adebanji for her useful corrections.

References

Azzalini, A.(1985):"A Class of Distribution which includes the Normal Ones" , scand J Statist. 12: 171-178.

Fang, K.T., Kotz, S.,and Ng, K.W., (1990). Symmetric Multivariate and Related Distributions. Chapman & Hall Ltd., London.

Fernandez, C and M. Steel (1998): "On Bayesian Modelling of Fat Tails and Skewness", Journal of the American Statistical Association, 93, 359-371.

Iyengar, S. and Tong, Y.L., 1989. Convexity Properties of Elliptically Contoured Distributions. Sankhya Ser. A 51, 13-29.

Jones, M.C. (2004). The families of distribution arising from order statistics. Test. Vol. 23, 535-545

Kotz, S. and Nadarajah, S., 2001. Letter to the Editor. Comm. Statist. 30(6), 987-992.

Ojo, M.O. (1989). Analysis of Some Prison data. Journal of Applied Statistics, 16, no.6, 377-383.

APPENDIX A

Table 1

Economically active population grouped by their ages.

Age group(class midpoints)	Observed frequency(n)
14.5	2390641
24.5	7111678
34.5	6833138
44.5	4569953
55.5	2696806
65.5	1730112
75.5	1292598

Sources: FOS/ILO/SIMPOC, National Modular Child Labour Survey, NBS, Abuja Nigeria.