Jour. Ind. Soc. Ag. Statistics
50(1), 1997 : 95–102

On a New Symmetrical Distribution

K. Srinivasa Rao, C.V.R.S. Vijaya Kumar, J.L. Narayana
Andhra University, Visakhapatnam - 530003
(Received : May, 1996)

SUMMARY

In this paper, we introduce a family of symmetric bellshaped distributions. The symmetric bellshaped distributions are useful in analysing the data that are arising from biological, sociological, agricultural and environmental experiments. The distributional properties of this family are derived. Various inferential aspects of the parameters of this family of distributions are discussed. This family also includes various platykurtic distributions.

Key words: Symmetrical distribution, Platykurtic distribution, Estimation.

1. Introduction

Applied statisticians must balance a number of issues when searching for an appropriate distribution in approximating a random phenomenon. Frequently the random situation may be modeled with normal distribution if the histogram of the data collected on the phenomenon is symmetric and bellshaped. However, in some situations, even though the shape of the sample frequency curve is symmetric and bellshaped the normal approximation may badly fit the distribution (see Table 1, Fig. 1). This may be due to the peakedness of the data which may not be mesokurtic. To model such situations a more appropriate distribution of symmetric nature is needed. So to give a suitable probability modeling of these samples a new family of distributions is introduced for which the kurtosis (β_2) is less than or equal to three, depending on the index (shape) parameter of the family. This family includes normal distribution as a particular case.

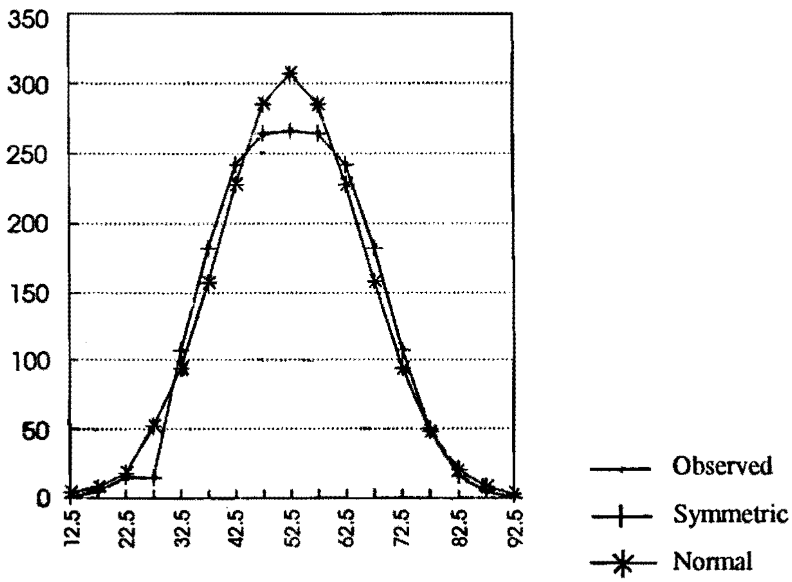

Fig. 1. Frequency Curves

Table 1 : Estimated and observed frequency distribution

Length of the fish (cm)	Observed frequencies	Estimated frequencies	
		Symmetric (r = 1)	Normal (r = 0)
10-15	1	0	4
15-20	5	5	8
20-25	14	16	18
25-30	50	50	52
30-35	107	107	94
35-40	183	182	157
40-45	242	242	228
45-50	263	264	285
50-55	266	266	307
55-60	264	264	285
60-65	242	242	228
65-70	184	182	158
70-75	108	107	94
75-80	48	50	48
80-85	16	16	21
85-90	4	5	8
90-95	1	0	3

2. *A Family of Symmetric Distributions*

A family of symmetric distributions is composed of distributions having the probability density function of the form

$$f_r(x, \mu, \sigma) = \frac{\left[2r + \left(\frac{x-\mu}{\sigma}\right)^2\right]^r e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma(2r)^r(2\pi)^{1/2} + \sum_{j=1}^{r}\binom{r}{j}(2r)^{r-j}\,2^{j+1/2}\,\Gamma(j+1/2)\sigma}$$

$$-\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma > 0 \qquad (2.1)$$

Each value of the shape parameter r (= 0, 1, 2, ...) gives a bell shaped distribution. For r = 0 the equation (2.1) reduces to normal probability density function with parameters μ and σ (Johnson and Kotz [1]).

The characteristics of the distribution are :

(i) symmetric about μ, which is the mean, median and mode.

(ii) the distribution function (d.f) is

$$F(x) = \frac{\sum_{j=1}^{r}\binom{r}{j}(2r)^{r-j}\,2^{(j+1/2)}\,\Gamma\left(j+\frac{1}{2}\right)F_j(x) + (2\pi)^{1/2}F_0(x)}{\sum_{j=1}^{r}\binom{r}{j}(2r)^{r-j}\,2^{(j+1/2)}\,\Gamma\left(j+\frac{1}{2}\right) + (2\pi)^{1/2}(2r)^r}$$

where

$$F_j(x) = \phi\left(\frac{x-\mu}{\sigma}\right) - e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\sum_{s=1}^{r}\frac{(x-\mu)^{2s-1}}{2^{s+\frac{1}{2}}\Gamma\left(s+\frac{1}{2}\right)\sigma^{2s-1}} \qquad (2.2)$$

where $\phi((x-\mu)/\sigma)$ is the d.f. of the standard normal variant.

(iii) The central moments are

$$\mu_{2n} = \left[\frac{\Gamma(n+1/2) + \sum_{j=1}^{r}\binom{r}{j}r^{-j}\Gamma\left(n+j+\frac{1}{2}\right)}{(\pi)^{1/2} + \sum_{j=1}^{r}\binom{r}{j}r^{-j}\Gamma\left(j+\frac{1}{2}\right)}\right]2^n\sigma^{2n}$$

$$\mu_{2n+1} = 0 \qquad (2.3)$$

(iv) The kurtosis of the distribution is

$$\left[\frac{(3\sqrt{\pi})}{4}+\sum_{j=1}^{r}\binom{r}{j}r^{-j}\left(j+\frac{3}{2}\right)\left(j+\frac{1}{2}\right)\left(\Gamma\left(j+\frac{1}{2}\right)\right)\right]\times$$

$$\frac{\left[(\pi)^{1/2}+\sum_{j=1}^{r}\binom{r}{j}r^{-j}\Gamma\left(j+\frac{1}{2}\right)\right]}{\left[(\sqrt{\pi})/2+\sum_{j=1}^{r}\binom{r}{j}r^{-j}\Gamma\left(j+\frac{1}{2}\right)\right]^{2}}=\beta_2 \qquad (2.4)$$

for $r = 0$, $\beta_2 = 3$. As r increases β_2 decreases. The values of β_2 for $r = 0, 1, 2, \ldots$ are given in Table 2.

Table 2 : Values of β_2 for various values of r

r	β_2	r	β_2
0	3.0000	11	1.0869
1	2.5200	12	1.0800
2	2.0260	13	1.0740
3	1.5105	14	1.0689
4	1.2398	15	1.0645
5	1.1826	16	1.0606
6	1.1539	17	1.0571
7	1.1333	18	1.0540
8	1.1176	19	1.0512
9	1.1052	20	1.0487
10	1.0952		

(v) The recurrence relation of central moments is

$$\mu_{2n} = (2n-1)\sigma^2\mu_{2n-2} + \frac{2\sigma^2 \sum_{j=1}^{r} j\binom{r}{j} r^{-j}\Gamma\left(n+j-\frac{1}{2}\right)}{\Gamma\left(n+\frac{1}{2}\right)+\sum_{j=1}^{r}\binom{r}{j} r^{-j}\,\Gamma\left(j-\frac{1}{2}+n\right)}\,\mu_{2n-2}$$

(vi) The characteristic function is of the form

$$\phi_x(t) = e^{it\mu} - \frac{t^2\sigma^2}{2}\left[\frac{\sqrt{\pi}+\sum_{j=0}^{r}\binom{r}{j} r^{-j}\sum_{k=0}^{2j} 2^{k/2}\,(it\sigma)^{2j-k}\,\Gamma\{(k+1)/2\}}{\sqrt{\pi}+\sum_{j=0}^{r}\binom{r}{j} r^{-j}\,2^j\,\Gamma\{(j+1)/2\}}\right]$$

3. *Estimation of the Parameters*

The shape parameter r can be estimated through sample kurtosis by solving the equation for r in which the sample β_2 is equated to the theoretical β_2.

From equation (2.4)

$$\frac{\left[\frac{3\sqrt{\pi}}{4}+\sum_{j=1}^{r}\binom{r}{j} r^{-j}\left(j+\frac{1}{2}\right)\left(j+\frac{3}{2}\right)\Gamma\left(j+\frac{1}{2}\right)\right]\times\left[(\pi)^{1/2}+\sum_{j=1}^{r}\binom{r}{j} r^{-j}\,\Gamma\left(j+\frac{1}{2}\right)\right]}{\left[(\sqrt{\pi})/2+\sum_{j=1}^{r}\binom{r}{j} r^{-j}\left(j+\frac{1}{2}\right)\Gamma\left(j+\frac{1}{2}\right)\right]^2} = \frac{\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})^4}{n}\right]}{\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\right]^2}$$

and solving this equation numerically we can find the value of r. By Rochies theorem there is one and only one real root to this equation. We can take the nearest integer to this real root as an estimator to the shape parameter (r). Using Table-2 one can estimate r by calculating the sample kurtosis and considering the closer value of population kurtosis given in the Table-2.

Knowing the shape parameter r one can obtain the estimators of the parameters μ and σ by method of moments as

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{n}{n-1}\left[1 + \frac{\sum_{j=1}^{r}\binom{r}{j} r^{-j}\,\Gamma\left(j+\frac{1}{2}\right)}{2\sum_{j=1}^{r}\binom{r}{j} r^{-j}\left(j+\frac{1}{2}\right)\Gamma\left(j+\frac{1}{2}\right)}\right] S^2$$

$$= s_r^2$$

where S^2 is the sample variance. Clearly $\hat{\mu}$ is the best linear unbiased estimator (BLUE) of μ and $\hat{\sigma}^2$ is the best quadratic unbiased estimator of σ^2, Rao *et al.* [4].

The variance of the estimators are

$$\text{Var}(\hat{\mu}) = \frac{2\sigma^2}{n}\left[\frac{1}{2} + \frac{\sum_{j=1}^{r}\binom{r}{j} r^{-j}\left(j+\left(\frac{1}{2}\right)\right)\Gamma\left(n+j+\frac{1}{2}\right)}{\frac{(\sqrt{\pi})}{2} + \sum_{j=1}^{r}\binom{r}{j} r^{-j}\,\Gamma\left(j+\frac{1}{2}\right)}\right]$$

$$\text{Var}(\hat{\sigma}^2) = \left(\frac{1}{n}\right)\frac{\sigma^4}{4\left(1-\frac{1}{n}\right)^2}\left[\frac{(\pi)^{1/2} + \sum_{j=1}^{r}\binom{r}{j} r^{-j}\,\Gamma\left(j+\frac{1}{2}\right)}{\frac{(\sqrt{\pi})}{2} + \sum_{j=1}^{r}\binom{r}{j} r^{-j}\left(j+\frac{1}{2}\right)\Gamma\left(j+\frac{1}{2}\right)}\right]^2 \times$$

$$\left[\frac{\Gamma\left(\frac{5}{2}\right) + \sum_{j=1}^{r}\binom{r}{j} r^{-j}\,\Gamma\left(j+\frac{5}{2}\right)}{\Gamma\left(\frac{1}{2}\right) + \sum_{j=1}^{r}\binom{r}{j} r^{-j}\,\Gamma\left(j+\frac{1}{2}\right)}\right]\left[\left(1-\frac{2}{n}+\frac{1}{n^2}\right) - \left(1-\frac{4}{n}+\frac{3}{n^2}\right)\right]$$

As each of the variances tends to zero as $n \to \infty$ the estimators $\bar{x}$ and s_r^2 are consistent for μ and σ^2 respectively, Rao [3].

The maximum likelihood estimate of the parameters μ and σ can be obtained by solving the following likelihood equations through iterative procedures by considering the moment estimators as first order approximation, Scarborough [5].

The maximum likelihood equations are

$$\bar{x} = \mu + \frac{2r\sigma^2}{n} \sum_{i=0}^{n} \left[\frac{x_i - \mu}{2r\sigma^2 + (x_i - \mu)^2} \right]$$

$$1 = \frac{1}{\sigma^2} \left[\left(\frac{1}{n} \right) \sum_{i=1}^{n} (x_i - \mu)^2 \right] - \frac{2r}{n} \sum_{i=1}^{n} \left[\frac{x_i - \mu}{2r\sigma^2 + (x_i - \mu)^2} \right]$$

4. *Fitting of the Distribution*

We have fitted a distribution with r = 1 to a data set on length of the fish (Rao *et al.* [4]). The estimated and observed frequencies are given in Table 1.

The calculated values are

Normal	New Symmetric
$\hat{\mu} = 52.507$	$\hat{\mu} = 52.507$
$\hat{\sigma}^2 = 12.906$	$\hat{\sigma}^2 = 9.976$
	$r = 1$

$$\beta = 2.530$$

$$\text{Observed } \chi^2_{\text{Normal}} = 37.154871$$

$$\chi^2_{\text{New symmetric}} = 0.8083$$

The 5% significance value of $\chi^2_{12} = 21.03$

Thus at 5% level of significance the distribution with index parameter r = 1 gives a good fit to the data and normal is not a good fit to the data. The curves of new symmetric and observed frequencies are almost coincided since the fit is very good.

ACKNOWLEDGEMENT

The authors are very much thankful to the referee for his valuable comments, which have improved the manuscript.

REFERENCES

[1] Johnson, N.L. and Kotz, S., 1972. *Distribution in Statistics: Continuous Multivariate Distributions*. John Wiley and Sons Inc., New York.

[2] Morgenstern, D., 1956. Einfache Beispiele Zwidimensionaler Verteilungen. *Mitteilingsblatt fur Mathematics Statistic*, **8**, 234-235.

[3] Rao, C.R., 1973. *Linear Statistical Inference and its Applications*. Wiley Eastern, New Delhi.

[4] Rao, K.S. *et al.*, 1988. On a bifocal distribution. *Bull. Cal. Mat. Soc.*, **80**, No. 4.

[5] Scarborough, B.J., 1964. *Numerical Mathematical Analysis*. Oxford Book Company, New Delhi.