King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 132

December 1992

On a Probability Function Useful in Size Modeling

M.A. Chaudhry, M. Ahmad

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

(Mac10902.tex)

ON A PROBABILITY FUNCTION USEFUL IN SIZE MODELING

By

M. Aslam Chaudhry and Munir Ahmad
Department of Mathematical Sciences
King Fahd University of Petroleum and Minerals
Dhahran, Saudi Arabia

Abstract

A probability function has been derived as a solution to a generalized Pearson differential equation. Some statistical properties of the function are investigated. An interesting relationship between the present distribution and the inverse Gaussian distribution is stated. It is demonstrated that the function is more suitable than other probability densities in some applications of size models.

1 INTRODUCTION

Bliss and Reinker (1964) showed that when compared to other probability densities like Weibull and inverse Gaussian, the three parameter-lognormal density function was relatively an appropriate model for the data collected on the diameter of Douglas firs in a variety of even aged stands. In this paper we have derived a probability function as a solution to a generalized Pearson differential equation. Some statistical properties of the function are investigated. We have found an interesting relationship between the present and the inverse Gaussian distribution. It is shown that the model is more suitable than the lognormal for the same data.

2 DERIVATION OF THE MODEL

In this section we derive the probability function as a solution to a generalized Pearson differential equation. It should be noted that the generalized Pearson distributions have been discussed by Dunning and Hansen (1977). They have provided a method of fitting the density function which is the solution of the differential equation

$$\frac{df}{dt} = \frac{c_0 + c_1 t + c_2 t^2 + \cdots + c_m t^m}{c_0' + c_1' t + c_2' t^2 + \cdots + c_n' t^n} f(t), \qquad m, n \geq 1. \tag{1}$$

We consider the differential equation

$$\frac{df}{dt} = \frac{c_0 + c_4 t^4}{c_3' t^3} f(t), \qquad c_3' \neq 0, \tag{2}$$

which is a special case of (1) when $m = 4$, $n = 3$ and $c_0' = c_1' = c_2' = 0$. The solution to the differential equation (2) is

$$f(t) = C \ \exp(-\alpha t^2 - \beta t^{-2}) \tag{3}$$

where, $\alpha = -\frac{1}{2} c_4/2c_3'$ $\beta = \frac{1}{2} c_0/2c_3'$ and C is the normalizing constant which can be determined (see Appendix B),

$$C = \left(\int_0^\infty \exp(-\alpha t^2 - \beta t^{-2}) dt \right)^{-1} = 2\sqrt{\frac{\alpha}{\pi}} \exp(2\sqrt{\alpha\beta}), \qquad \alpha > 0, \ \beta > 0.$$

Therefore,

$$f(t) = 2\sqrt{\frac{\alpha}{\pi}} \exp(2\sqrt{\alpha\beta}) \ \exp(-\alpha t^2 - \beta t^{-2}), \qquad t > 0, \ \alpha > 0, \ \beta > 0. \tag{4a}$$

or

$$f(t) = 2\sqrt{\frac{\alpha}{\pi}} \exp(-(\sqrt{\alpha} t - \sqrt{\beta} t^{-1})^2) \qquad t > 0, \ \alpha > 0, \ \beta > 0. \tag{4b}$$

3 PROPERTIES OF THE PROBABILITY DENSITY FUNCTION

The cumulative probability function of the random variable X is

$$F(x) = 2\sqrt{\frac{\alpha}{\pi}}\exp(2\sqrt{\alpha\beta})\int_0^x \exp(-\alpha t^2 - \beta t^{-2})dt,$$

which can be integrated to give

$$F(x) = 1 - \frac{1}{2}\left\{\exp(4\sqrt{\alpha\beta})Erfc(\sqrt{\alpha}x + \sqrt{\beta}x^{-1}) + Erfc(\sqrt{\alpha}x - \sqrt{\beta}x^{-1})\right\}, \tag{5}$$

and that can further be expressed in terms of the cumulative distribution function Φ of the normal distribution to give

$$F(x) = \Phi(\sqrt{2}(\sqrt{\alpha}x - \sqrt{\beta}x^{-1})) - \exp(4\sqrt{\alpha\beta})\Phi(\sqrt{2}(\sqrt{\alpha}x + \sqrt{\beta}x^{-1})).$$

It should be noted that $\Phi(x)$ is a well tabulated function in the literature.

The r–th moment $(r \geq 0)$ μ_r' about the origin of the random variable X of the probability function $f(t)$ is given by

$$\mu_r' = 2\sqrt{\frac{\alpha}{\pi}}\exp(2\sqrt{\alpha\beta})\int_0^\infty t^r\exp(-\alpha t^2 - \beta t^{-2})dt.$$

Using the result equation (A.11) we get

$$\mu_r' = 2\sqrt{\frac{\alpha}{\pi}}\exp(2\sqrt{\alpha\beta})\ (\beta/\alpha)^{(r+1)/4}\ K_{\frac{r}{2}+\frac{1}{2}}(2\sqrt{\alpha\beta}),$$

where K_ν is the Macdonald function (see Appendix – A).

In particular when $r = 2n$ is an even integer, we get (see equation (A.9)).

$$\mu_{2n}' = (\beta/\alpha)^{n/2}\sum_{m=0}^{n}\frac{(4\sqrt{\alpha\beta})^{-m}}{m!}\frac{\Gamma(n+m+1)}{\Gamma(n+1-m)}. \tag{6}$$

The first four moments about the origin can be obtained as

$$\mu_1' = 2\sqrt{\beta/\pi}\exp(2\sqrt{\alpha\beta})K_1(2\sqrt{\alpha\beta}) \qquad (7)$$

$$\mu_2' = \frac{1}{2\alpha}(1+2\sqrt{\alpha\beta}), \qquad (8)$$

$$\mu_3' = \frac{2\beta}{\sqrt{\alpha\pi}}\exp(2\sqrt{\alpha\beta})K_2(2\sqrt{\alpha\beta}) \qquad (9)$$

and

$$\mu_4' = \frac{1}{4\alpha^2}(3+6\sqrt{\alpha\beta}+4\alpha\beta). \qquad (10)$$

The mode is found to be

$$\mu_0 = (\beta/\alpha)^{1/4}. \qquad (11)$$

It follows from (11) that $\beta = \alpha\mu_0^4$. The substitution of this value of β in (4.a) leads to

$$f(t) = 2\sqrt{\frac{\alpha}{\pi}}\exp(2\alpha\mu_0^2)\exp\left[-\alpha\mu_0^2\left\{(t/\mu_0)^2+(\mu_0/t)^2\right\}\right], \quad (\alpha>0, \mu_0>0, t>0). \qquad (12)$$

It should be noted that for the probability density function in (12), μ_0 is the location parameter and $\alpha\mu_0^2$ is the shape parameter.

If we introduce a new random variable $Y = T/\mu_0$ and substitute $\delta = \alpha\mu_0^2$ in (12) we get

$$f(y) = 2\sqrt{\frac{\delta}{\pi}}\exp(2\delta)\exp\left\{-\delta(y^2+1/y^2)\right\} \quad (\delta>0, y>0). \qquad (13)$$

The graphs of $f(y)$ for fixed value of 'μ_0' and different values of 'δ' are given in Figure 1.

The moment ratios for the distribution are $\sqrt{\beta_1} = \alpha_3 = \mu_3/\mu_2^{3/2}$ and $\beta_2 = \alpha_4 = \mu_4/\mu_2^2$, where μ_2, μ_3 and μ_4 are moments about mean. It may be noted that $\sqrt{\beta_i} \to 0$ and $\beta_2 \to 3$ when $\delta \to 0$ which shows normality. (β_1, β_2)-graphs of the distribution along with that of some Pearson type distributions are given in figure 2. The upside-down presentation of this figure is in accordance with the well established conversion. It may be noted from figure 2 that the present distribution corresponds to the area in the (β_1, β_2)-diagram like the Pearson distributions of the type I, IV and VI (see Johnson and Kotz (1970)).

4 ESTIMATION OF PARAMETERS

4.1 THE METHOD OF MAXIMUM LIKELIHOOD

Consider the likelihood function defined by $L = \prod_{i=1}^{n} f(t_i)$. The objective of the likelihood function approach is to determine those values of the parameters which maximize the function L. It is more convenient in our situation to use the equivalent function $R = \ln(L)$.

Solving the maximum likelihood equations $\frac{\partial R}{\partial \alpha} = 0$, $\frac{\partial R}{\partial \beta} = 0$, we get

$$\frac{n}{2\hat{\alpha}} + n\frac{\sqrt{\hat{\beta}}}{\sqrt{\hat{\alpha}}} - \sum_{i=1}^{n} t_i^2 = 0 \tag{14}$$

and

$$\frac{n\sqrt{\hat{\alpha}}}{\sqrt{\hat{\beta}}} - \sum_{i=1}^{n} t_i^{-2} = 0. \tag{15}$$

A simple algebraic manipulation of (14) – (15) yields the maximum likelihood estimators (M.L.E.) as

$$\hat{\alpha} = \hat{\beta}(m'_{-2})^2 \text{ and } \hat{\beta} = (2(m'_{-2}m'_2 - 1)m'_{-2})^{-1}, \tag{16}$$

where,

$$m'_2 = \frac{1}{n}\sum_{i=1}^{n} t_i^2 \text{ and } m'_{-2} = \frac{1}{n}\sum_{i=1}^{n} 1/t_i^2. \tag{17}$$

The asymptotic variances and covariance of $\hat{\alpha}$ and $\hat{\beta}$ are found to be

$$\mathrm{Var}(\hat{\alpha}) = \frac{2\alpha^2}{n},$$

$$\mathrm{Var}(\hat{\beta}) = \frac{2\beta^2}{n}\left(\frac{1+\sqrt{\alpha\beta}}{\sqrt{\alpha\beta}}\right),$$

and

$$\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = -\frac{2}{n}\alpha\beta.$$

4.2 THE METHOD OF MOMENTS

Let m_2' and m_4' be the two sample moments about the origin. The moment estimation of the parameters α and β can be found from (7) and (8) and are given respectively by

$$\tilde{\alpha} = \frac{m_2' + \sqrt{4m_4' - 3(m_2')^2}}{4(m_4' - (m_2')^2)}, \tag{18}$$

and

$$\tilde{\beta} = m_4'\tilde{\alpha} - \frac{3}{2}m_2'. \tag{19}$$

Since the moment estimators are not efficient as compared to the maximum likelihood estimators, the asymptotic variances and covariances $\tilde{\alpha}$ and $\tilde{\beta}$ have not been obtained.

5 RELATIONSHIP TO OTHER DISTRIBUTIONS

The density function introduced in this paper is closely related to the inverse Gaussian density function. The inverse Gaussian density function which arises as the density of the first passage time of Brownian motion with positive drift is given by

$$f_Y(y;\mu,\nu) = (\frac{\nu}{2\pi y^3})^{1/2}\exp(\frac{-\nu(y-\mu)^2}{2\mu^2 y}), \quad y,\mu > 0. \qquad (20)$$

It can be seen that if a variate Y has the inverse Gaussian distribution (14) then the distribution function of $T = 1/\sqrt{Y}$ is the density function given by (4.a).

6 APPLICATION

To illustrate the applications of the probability function we consider the data collected by Bliss and Reinker (1964) on the diameter of Douglas firs in a variety of even aged stands. The three parameter lognormal was considered as an appropriate model for the data and it was shown (Boswell et al, 1979) that $\chi^2_{11} = 7.07$. If the last three categories of the data are pooled, the χ^2_9 value becomes 6.672. The lognormal model was considered better than many other models (Boswell et al, 1979)/ We have fitted our two-parameter probability function on the same data based on M.L.E. method. The M.L.E. of α and β are $\hat{\alpha} = 0.0217$ and $\hat{\beta} = 7.606$ when $m_2' = 41.727$ and $m_{-2}' = 0.0534$. The χ^2–value of our probability density function is found to be $\chi^2_{(10)} = 3.316$. It shows that our p.d.f. fits the data better than the lognormal p.d.f.

The observed and expected frequencies under lognormal and our distribution are shown in the table below.

Diameter (inches)	Observed Frequency	Expected Frequency (Log-Normal)	Expected Frequency (our p.d.f.)
1.5–	33	36.8	30.3
2.5–	68	58.3	73.1
3.5–	94	74.6	91.9
4.5–	86	94.6	90.4
5.5–	76	87.6	78.6
6.5–	68	70.5	63.1
7.5–	53	51.1	47.5
8.5–	34	34.1	34.0
9.5–	24	21.4	23.0
10.5–	16	12.8	15.0
11.5–	7	7.4	9.2
12.5–	4	4.1	5.5
13.5–	3	2.3	
14.5–	1 5	1.2 4.9	6.4
15.5–	1	1.4	
	$\chi^2_{(9)} = 6.672$,	$\chi^2_{(11)} = 7.07$,	$\chi^2_{(10)} = 3.316$

The asymptotic variances and covariances of $\hat{\alpha}$ and $\hat{\beta}$ are found to be

$$\text{Var}(\hat{\alpha}) = 1.613(E{-}06), \quad \text{Var}(\hat{\beta}) = 0.713, \quad \text{and} \quad \text{Cov}(\hat{\alpha}, \hat{\beta}) = -5.751(E{-}04).$$

REMARK

The inverse Gaussian and related distributions have been discussed in the literature as a possible alternative to log-normal [Folk and Chhikra (1978), Johnson and Kotz (1970)]/ The interesting relationship between the present and the inverse Gaussian distribution is the most straightforward derivation of the distribution introduced in this paper. However, despite the close relationship to the inverse Gaussian distribution, it does not seem to be considered in the open literature. It is anticipated that the present work will provide a motivation to study the class of distributions of $Y^{1/p}$ where Y is the reciprocal of the inverse Gaussian random variable. Moreover, contrary to Dunning and Hanson's (1977), numerical approximations to generalized Pearson's distribution, we provide an analytic solution to the generalized Pearson's equation, which means maximum likelihood can be used to estimate the parameters.

7 ACKNOWLEDGEMENTS

The authors appreciate the helpful comments made by the Associate Editor and referees on the earlier version of the paper. Useful discussions with Professor A.K. Sheikh, Department of Mechanical Engineering, KFUPM, Dhahran are appreciated. Special thanks are to Messrs Shamsuddin Khan and M. Sami for the help in computer graphics.

8 REFERENCES

1. Bliss C.I. and Reinker K.A. (1964), A lognormal approach to diameter distributions in even aged stands. Forest Science 10, 350–360.

2. Boswell, M.T., Ord J.K. and Patil, G.P. (1979), "Chance mechanisms underlying univariate distributions". Statistical Distributions in Ecological Work. Vol. 4, pp. 3–156, Interscience Cooperative Publishing House.

3. Dunning, K.A., and Hanson, J.N. (1977), Generalized Pearson distributions and nonlinear programming. *J. Statist. Comput. Simul.,* **6**, 115-128.

4. Erdelyi A, et al., (1953) Higher Transcendental Functions, vol. II, McGraw-Hill, New York.

5. Folks, J.L. and Chhikara, R.S. (1978), The inverse Gaussian distribution and its statistical application. A Review. *J. Roy. Sta. (B)*, **40**, 263-289.

6. Gradshteyn, I.S. and Ryzhik, I.M. (1980), Tables of Integrals, Series and Products, Academic Press, New York.

7. Johnson, N.L. and Kotz, S. (1970), *Continuous Univariate Distributions - 1. Vol. 2*, Houghton Mifflin Co., Boston.

APPENDIX–A

Properties of the Macdonald Function, $K_\nu(x)$

The modified Bessel differential equation (see Gradshteyn and Ryzhik, 1980)

$$\frac{du^2}{dx^2} + \frac{1}{x}\frac{du}{dx} - \left(1 + \frac{\nu^2}{x^2}\right) u = 0, \tag{A.1}$$

has two linearly independent solutions $I_\nu(x)$ and $K_\nu(x)$. The function

$$I_\nu(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \quad |x| < \infty \tag{A.2}$$

is called modified Bessel function of the first kind and the function

$$K_\nu(x) = \frac{\pi}{2}\frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\nu\pi)} \quad \text{if } \nu \neq 0, \pm 1, \pm 2, \ldots, \tag{A.3}$$

and

$$K_n(x) = \lim_{\nu \to n} K_\nu(x) \quad n = 0, \pm 1, \pm 2, \pm 3, \ldots, \tag{A.4}$$

is called the Macdonald function. The function $K_\nu(x)$ is symmetric in ν; that is,

$$K_\nu(x) = K_{-\nu}(x). \tag{A.5}$$

The asymptotic behaviour of $K_\nu(x)$ is given by

$$K_0(x) \approx \ln(\frac{2}{x}), \quad x \to 0 \tag{A.6}$$

$$K_n(x) \approx \frac{1}{2}(n-1)!(2/x)^n, \quad x \to 0,\ n = 1, 2, 3, \ldots. \tag{A.7}$$

$$K_\nu(x) \approx \sqrt{\frac{\pi}{2x}}\exp(-x), \quad x \to \infty. \tag{A.8}$$

The Macdonald function $K_\nu(x)$ for $\nu = n + \frac{1}{2}$, $n = 0, 1, 2, 3, \ldots$ reduces to elementary functions given by

$$K_{\pm\frac{1}{2}} = \sqrt{\frac{\pi}{2x}}\exp(-x) \tag{A.9}$$

$$K_{n+\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}}\exp(-x)\sum_{m=0}^{n}\frac{(2x)^{-m}}{m!}\frac{\Gamma(n+m+1)}{\Gamma(n-m+1)}. \tag{A.10}$$

An integral representation of the Macdonald function is given by

$$K_\nu(2x) = \frac{1}{2}x^\nu\int_0^\infty \exp(-t - x^2/t)t^{-\nu-1}dt. \tag{A.11}$$

In particular when $\nu = -\frac{1}{2}$, we get from (A.9) - (A.10) that,

$$\int_0^\infty \exp(-t - x^2/t)t^{-1/2}dt = \sqrt{\pi}\exp(-2x). \tag{A.12}$$

APPENDIX - B

Evaluation of the Normalizing Constant, C

According to the definition of the normalizing constant C, we have,

$$\frac{1}{C} = \int_0^\infty \exp(-\alpha t^2 - \beta t^{-2})dt. \tag{B.1}$$

Substituting $t = \sqrt{\tau/\alpha}$, $dt = d\tau/2\sqrt{\alpha t}$ in (B.1) we get

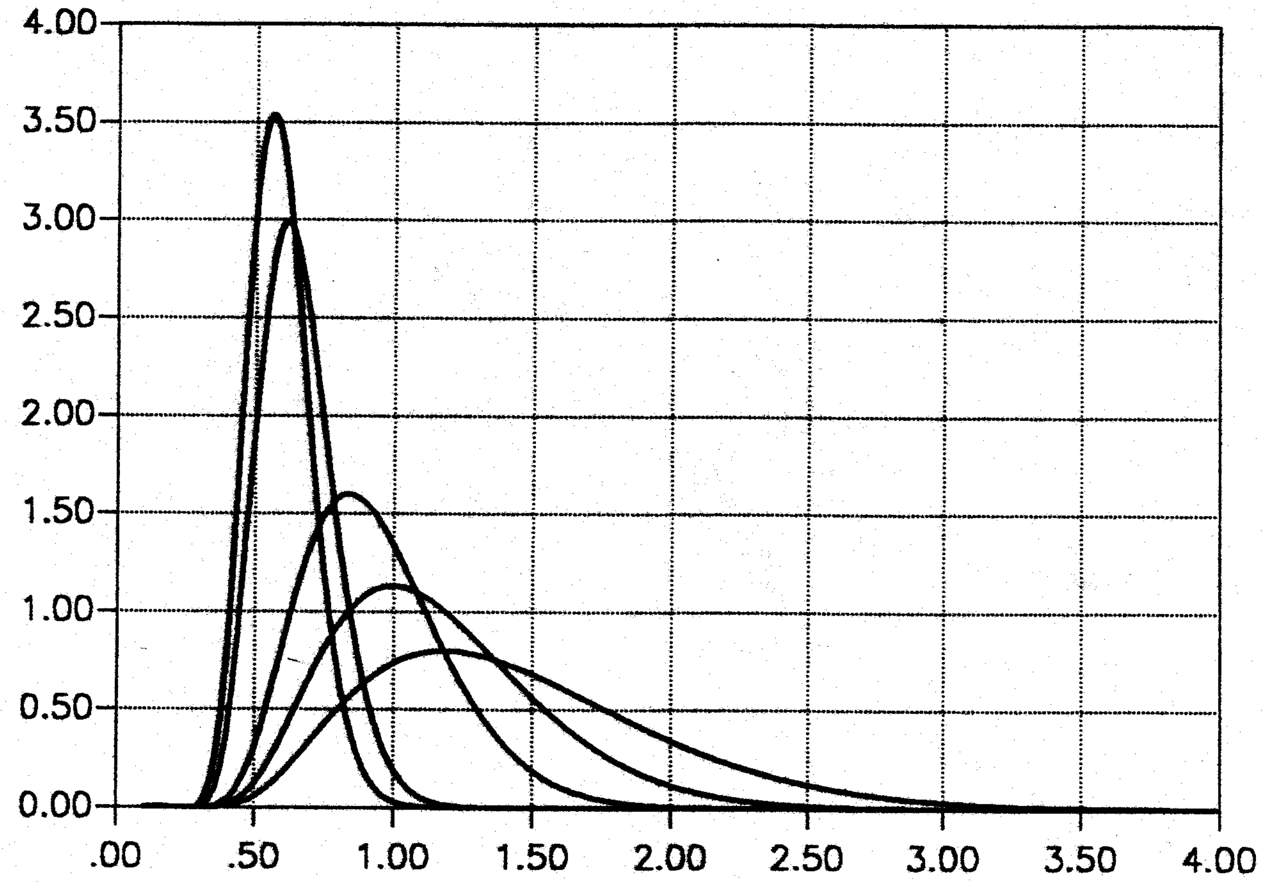
$$\frac{1}{C} = \frac{1}{2\sqrt{\alpha}}\int_0^\infty \exp(-\tau - \alpha\beta/\tau)\tau^{-1/2}d\tau. \tag{B.2}$$

From equations (A.11) – (A.12) and (B.2), we get

$$\frac{1}{C} = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}\exp(-2\sqrt{\alpha\beta}),$$

or

$$C = 2\sqrt{\frac{\alpha}{\pi}}\exp(2\sqrt{\alpha\beta}). \tag{B.3}$$


4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
.00
.50
1.00
1.50
2.00
2.50
3.00
3.50
4.00

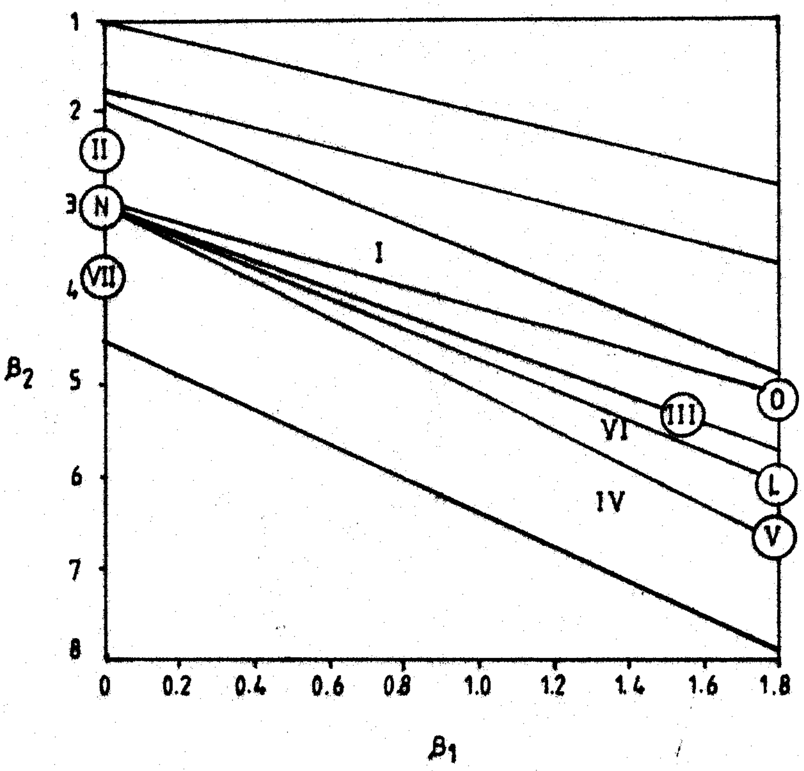


Fig.2

A CHART RELATING OUR DISTRIBUTION (O), LOG NORMAL (L) AND THE TYPE OF PEARSON FREQUENCY CURVE TO THE VALUES OF β_1, β_2