APPENDIX I

A. Representation of Density function :

Transformation Method.

Let $Z = f(x)$ be the transformation of x such that Z has unit normal distribution. The function f (x) should depend on a certain number of parameters. To control four moments one has to have a tag on at least four independent parameters. Let us write.

$$Z = \gamma + \delta f\left(\frac{x - \xi}{\lambda}\right)$$

If one writes $y = \frac{x-\xi}{\lambda}$ then $Z = \gamma + \delta f(y)$ and $p(y)\, dy = p'(Z)\, dZ$ by law of conservation of probability where p(y) is the probability density function for $y = \frac{x-\xi}{\lambda}$, and $p'(Z)$ is the correspon-ding function for Z. $p'(Z)$ is assumed to be a gaussian.

Therefore $p(y) = p'(Z) \frac{dZ}{dy}$ $Z \doteq \gamma + \delta f(y)$

$$p(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-1/2\, (\gamma + \delta f(y))^2 * \delta * f'(y)\right)$$

Since $x = \xi + \lambda y$, the distribution of x is of the same shape as that of y. It is then obivious that λ is the parameter affecting the variance, and ξ is responsible for any changes in the centroid value of the distribution. The remaining parameters γ and δ determine the shape of the distribution and depend on skewness and excess.

With the following three transformations, it is possible to cover all the possible values of skewness and excess uniquely.

1. Lognormal distribution : $Z = \gamma + \delta \log(\frac{x-\xi}{\lambda})$ transforms a lognormal distribution to normal distribution. There are only three independent parameters in the transformation as it can be written as $Z = (\gamma - \delta \log \lambda) + \delta \log (x-\xi) = \gamma' + \delta \log (x-\xi)$; therefore, the transformation depends only on three moments. There exists a relation between the third and the fourth moment. Figure 11 shows the relation between skewness and excess. The probability density of x is given by

$$p(x) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{(x-\xi)} \left(\exp\left(-1/2\left(\gamma' + \delta \log (x-\xi)\right)^2\right)\right) \quad (x-\xi) > 0.$$

The distribution has the following properties (i) The range of the distribution is bounded at one end and extends indefinitely in the other direction (ii) The sign of skewness decides at which end of the distribution is the range bounded. (iii) Skewness = 0 is a limiting case and gives a gaussian distribution.

The curve in Figure 11 divides the region into two parts. A different transformation is used in each part. The lognormal distribution is the limiting case for both of these regions.

2. S_B type transformation :

The S_B type transformation can be mathematically represented

FIGURE 11

The lognormal line.

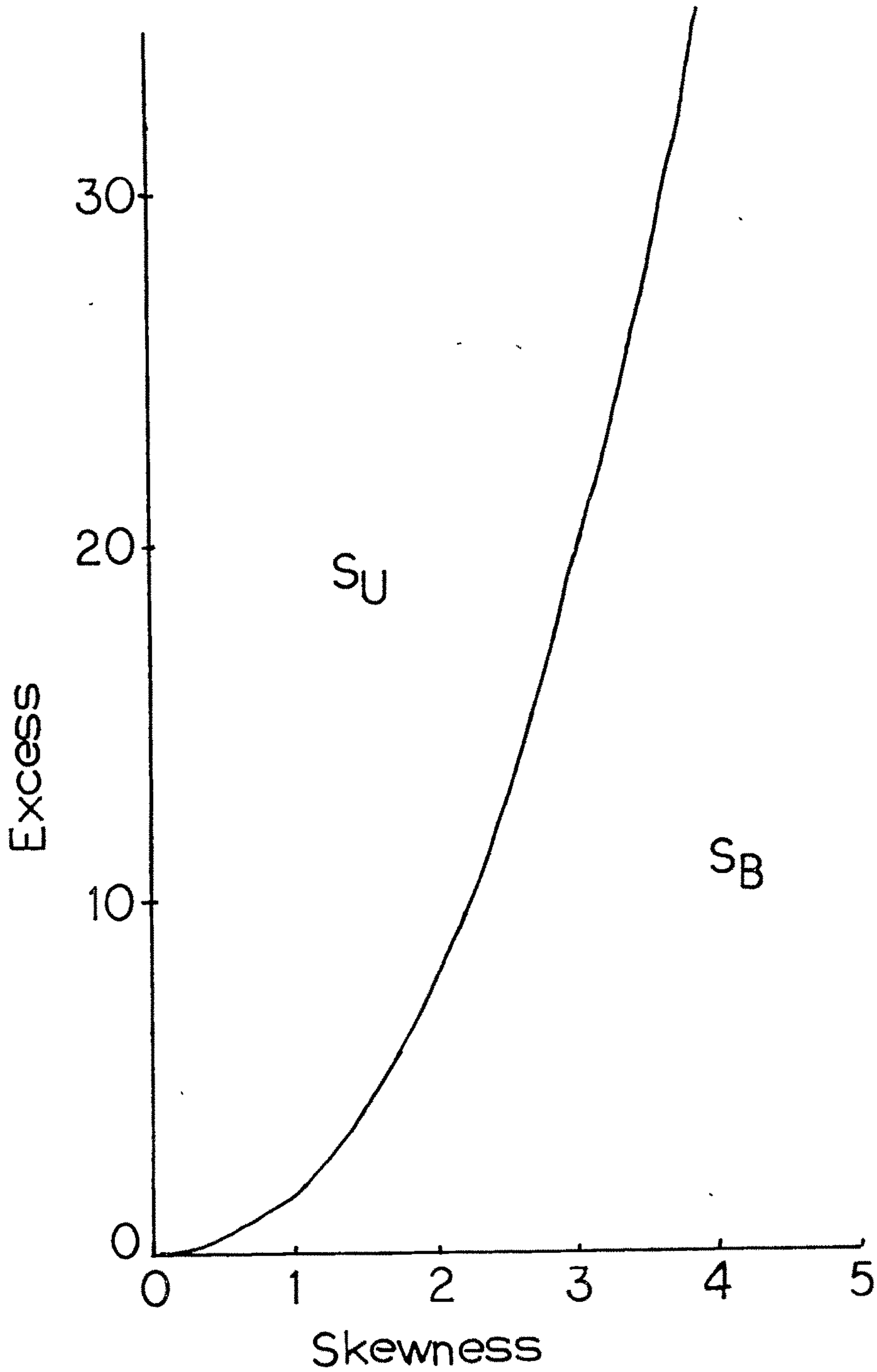


FIGURE 11

as $Z = \gamma_1 + \delta \log \left(\frac{x - \xi}{\xi + \lambda - x} \right) = \gamma + \delta \log \left(\frac{y}{1-y} \right)$

where $y = \frac{x - \xi}{\lambda}$

The probablity density function for y is given by

$$p(y) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{y(1-y)} \exp\left(-1/2 \left(\gamma + \delta \log \left(\frac{y}{1-y} \right) \right)^2 \right)$$

The range of the distribution is bounded at both ends. This transformation can be used in the region S_B in Figure 11. The region is bounded on both sides, on one side by lognormal distribution and on the other side by the limiting value of $(\gamma_2 - \gamma_1{}^2)$.

3. S_U type transformation :

$$Z = \gamma + \delta \sinh^{-1} \left(\frac{x - \xi}{\lambda} \right) = \gamma + \delta \sinh^{-1} y$$

$$= \gamma + \delta \log \left(y + \sqrt{y^2 + 1} \right) \text{ where } y = \frac{x - \xi}{\lambda}$$

The probability density function for y is then given by

$$p(y) = \frac{\delta}{\sqrt{2\pi}} \frac{1}{(y^2+1)^{\frac{1}{2}}} \exp\left(-1/2 \left(\gamma + \delta \log \left(y + \sqrt{y^2 + 1} \right) \right)^2 \right)$$

The range of the distribution is unbounded; the transformation can be used in the region marked S_U in the Figure 11. The region is bounded by lognormal distribution on one side. All the distributions in this region are unimodal. In region S_B however, the distribution could be bimodal. The following steps are required to find the parameters for the transformation. (i) Using values of skewness and excess, determine the distribution system. (ii) Using proper tables, determine the parameters γ and δ to fit the values of skewness and excess. (iii) Parameters λ and ξ are fixed so as to obtain proper values of variance and centroid.

B. Representations of Pearson's distribution curves.

As mentioned in the main text, for every member of Pearson's system, the probability density function p (x) satisfies a differential equation of the form

$$\frac{1}{p} \frac{dp}{dx} = - \frac{a + x}{c_0 + c_1 x + c_2 x^2} \quad .. \quad (1)$$

along with the conditions

$$p(x) \geqslant 0 \quad \text{and} \quad \int_{-\infty}^{\infty} p(x)\, dx = 1 \quad .. \quad (2)$$

Provided 'a' is not a root of the equation $c_0 + c_1 x + c_2 x^2 = 0$, p (x) is finite when x =a and $\frac{dp(x)}{dx} = 0$ when x = a. Also $\frac{dp(x)}{dx} = 0$ when y =0 But if $x \neq a$ and $y \neq 0$ then $\frac{dp(x)}{dx} \neq 0$. Since the conditions (2) must be satisfied, it follows ffom (1) that $p(x) \to 0$ as $x \to \infty$ and so also must $\frac{dp(x)}{dx}$. In such cases when $p(x) \geqslant 0$ is not satisfied, it is necessary to restrict the range of values of x such that $p(x) = 0$ when x is outside this range.

The shape of the curve representing the probability density function varies considerably with the values of the coefficients a, c_0, c_1 and c_2. Furhter, the form of solution of (1) depends on the roots of the equation

$$c_0 + c_1 x + c_2 x^2 = 0 \qquad .. (3)$$

The various curves of Pearson's system correspond to the different forms of the solution. The different shapes have been classified into a number of types by Pearson. Here is a resume of his classification.

TYPE I : Corresponds to the case when both roots a_1 and a_2 of (3) are of real and opposite sign. i.e. $a_1 < 0 < a_2$.

$\therefore c_o + c_1 x + c_2 x^2 = - c_2 (x-a_1)(a_2 - x)$ Equation (1) becomes

$$\frac{d \log p(x)}{dx} = \frac{1}{c_2 (a_2 - a_1)} \left[\frac{a + a_1}{x - a_1} + \frac{a + a_2}{a_2 - x} \right]$$

$$\therefore p(x) = k (x - a_1)^{m_1} (a_2 - x)^{m_2} \qquad ... (4)$$

$$m_1 = \frac{a + a_1}{c_2 (a_2 - a_1)} \quad ; \quad m_2 = - \frac{a + a_2}{c_2 (a_2 - a_1)}$$

The range of x is limited so that both $(x - a_1)$ and $(a_2 - x)$ are positive and $a_1 < 0 < a_2$ is satisfied. Equation (4) can represent a proper probability density provided m_1 , $m_2 > -1$. This is a general form of the Beta Distribution. Type I can be further subdivided into two types depending on values of m.
Type I (U) if $m_1 < 0$ and $m_2 < 0$.
Type I (J) if $m_1 < 0$ and $m_2 > 0$ or if $m_1 > 0$ and $m_2 < 0$.
If m_j (j=1,2) is zero then p(x) approaches a non zero limit when x approaches a_j (j = 1,2).

TYPE II : is a symmetrical form of (4) with $m_1 = m_2$. If $m_1 = m_2 < 0$, the distribution is called Type II (U). TYPE III corresponds to the case $c_2 = 0$, $c_1 \neq 0$. In this case the solution of (1) is given by $p(x) = k\,(c_0 + c_1\,x)^m \exp\,(-x/c_1)$.

$$m = 1/c_1\,\left(\frac{c_0}{c_1} - a\right)$$

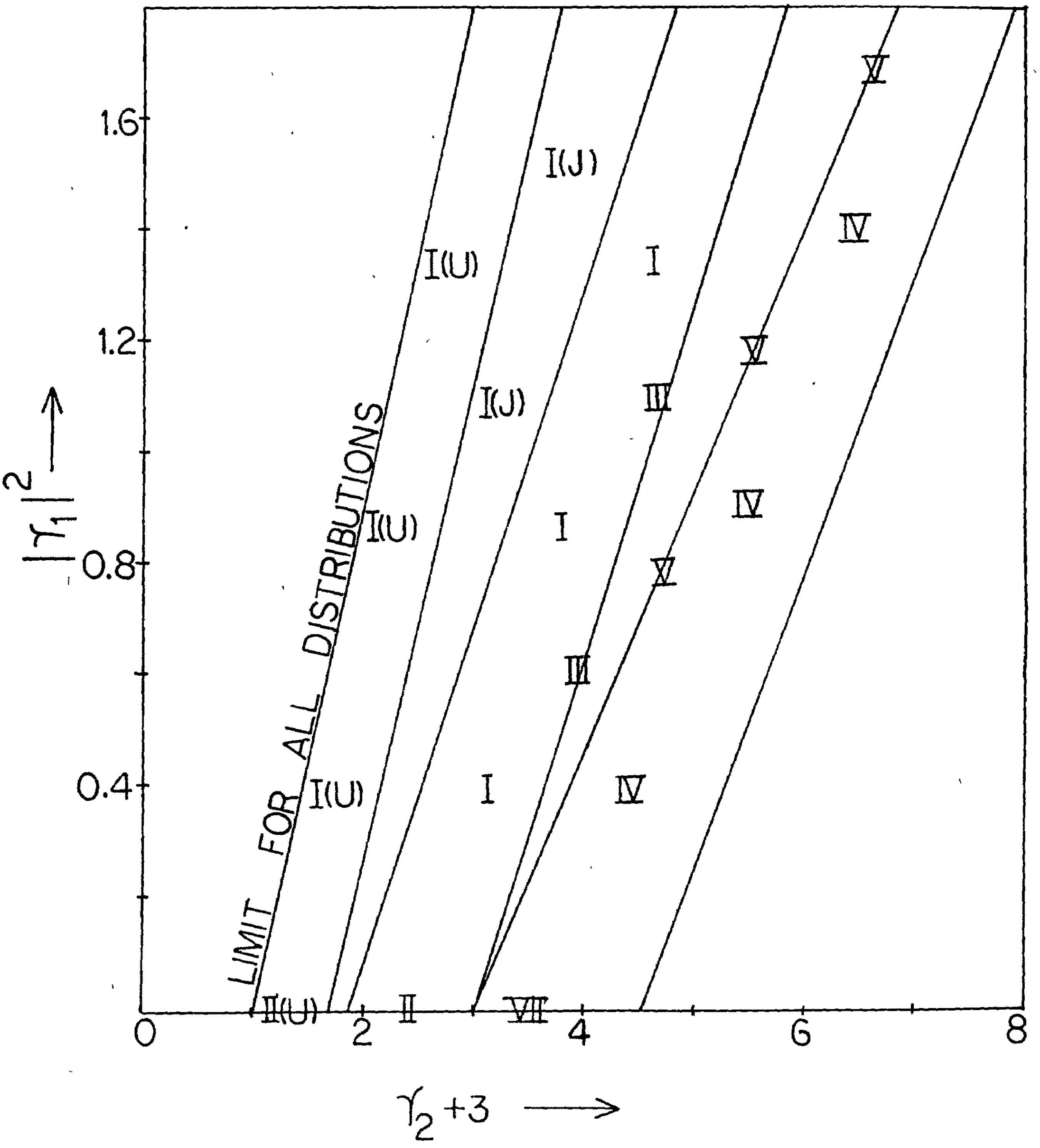
The range of x is greater than or less than $-\,c_0/c_1$ according as $c_1 > 0$ or $c_1 < 0$. The curves represent the general form of Gamma distribution.

TYPE IV : represents the case when (3) does not have any real roots. We use the identity. $c_0 + c_1x + c_2x^2 = C_0 + c_2(x + C_1)^2$

Where $C_o = c_0 - 1/4\,\frac{c_1^2}{c_2}$, $C_1 = 1/2\,\frac{c_1}{c_2}$

FIGURE 12

The Pearson's system of curves in the γ_1, γ_2 plane.
$\beta_1 = |\gamma_1|^2$, $\beta_2 = \gamma_2 + 3$



1.6
1.2
0.8
0.4
$|\gamma_1|^2$
0
2
4
6
8
$\gamma_2 + 3$
LIMIT FOR ALL DISTRIBUTIONS
I(J)
I(U)
I
V
IV
III
II(U)
II
VII

FIGURE 12

$$\therefore p(x) = k(C_0 + c_2 (x + C_1)^2)^{-1/2c_2} \exp\left(-\frac{a-C_1}{\sqrt{c_2 C_0}} * \tan^{-1} \frac{x + C_1}{\sqrt{C_0/c_2}}\right)$$

TYPE V when $c_0 + c_1 x + c_2 x^2$ is a perfect square i.e.
$c_1^2 = 4 c_0 c_2$, we have

$$p(x) = k (x + C_1)^{-1/c_2} \exp\left(\frac{a - C_1}{c_2 (x + C_1)}\right)$$

If $(a - C_1)/c_2 < 0$ then $x > -C_1$, $(a-C_1)/c_2 > 0$ then $x < -C_1$

Special case : $p(x) = k (x + C_1)^{-1/c_2}$ when $a = C_1$ and $|c_2| < 1$.

TYPE VI When roots of (3) are both real and of same sign. If both a_1 and a_2 are negative (say) $a_1 < a_2 < 0$, then similar to the analysis leading to (4), we have
$p(x) = k (x-a_1)^{m_1} (x - a_2)^{m_2}$. This represent a proper density function if $m_2 < -1$ and $m_1 + m_2 < 0$. Also the range of x should be such that $x > a_2$.

TYPE VII relates to the situation when $c_1 = a = 0$; $c_0 > 0$, $c_2 > 0$, then $p(x) = k (c_0 + c_2 x^2)^{-1/2c_2}$

The central t distribution belongs to this class.

The coefficients a, c_0, c_1, c_2 in (1) can be represented in terms of the moments of the distribution. Multiplying (1) by

x^r on both sides, we have

$$x^r\,(c_0+c_1x + c_2x^2)\,\frac{dp(x)}{dx} + x^r\,(a+x)p(x) = 0 \qquad \ldots (5)$$

Integrating both sides from $-\infty$ to $+\infty$ and assuming $x^r p(x) \to 0$ as $x \to \pm\infty$, for $r \leqslant 5$,, we get

$$- rc_0\,\mu'_{r-1} + (-(r+1)c_1+a)\,\mu'_r + (\,-\,(r+2)\,c_2+1)\,\mu'_{r+1} = 0 \;\ldots (6)$$

where μ'_r are the central moments of the distribution. Putting $r = 0,\ 1,2,3$ in (6) and observing that $\mu'_0 = 1$ and $\mu'_{-1} = 0$ and further, if the expected value of the variable (centroid) is zero $\mu_1{}' = 0$ and $\mu_r{}' = {}_r$ fo8 $r \geqslant 2$.

$$\therefore \quad c_0 = (4\beta_2 - 3\beta_1)\,(10\beta_2 - 12\beta_1 - 18)^{-1}\,\mu_2$$

$$a = c_1 = \sqrt{\beta_2}\,(\beta_2 + 3)\,(10\beta_2 - 12\beta_1 - 18)^{-1}\sqrt{\mu_2} \quad \ldots (7)$$

$$c_2 = (2\beta_2 - 3\beta_1 - 6)\,(10\beta_2 - 12\beta_1 - 18)^{-1}$$

From difinitions of various types, we get from (7) for Type I

$$k = 1/4\ \beta_1\,(\beta_2 + 3)^2\,(4\beta_2-3)^{-1}(2\beta_2 - 3\beta_1 - 6)^{-1} < 0.$$

Type II $\quad \beta_1 = 0,\ \beta_2 < 3.$

Type III $\quad 2\beta_2 - 3\beta_1 - 6 = 0.$

Type IV $\quad 0 < k < 1$

Type V $k = 1$

Type VI $k > 1$

Type VII $\beta_1 = 0, \beta_2 = 3.$

The division of (γ_1, γ_2) plane, among various types is shown in figure 12. Only types I, VI and IV correspond to areas in (γ_1, γ_2) plane. The others correspond to lines.

General Properties of Pearson's System :-

1. Mean Deviation $= 2 \left(\frac{1 - 3c_2}{1 - 2c_2} \right) \mu_2 \; p \; (\mu_1')$

for all types of Pearson's system distributions.

2. There is a mode or antimode at $x = -a$.

3. $$\frac{d^2 p(x)}{d x^2} = \frac{p(x)}{(c_0 + c_1 x + c_2 x^2)^2} * \left((x+a)^2 + c_2(x+a)^2 + a^2(1-c_2) - c_0 \right)$$

provided $a^2 (1-c_2) < c_0$ there are points of inflexion equidistant from the mode at

$$x = \text{Mode} \pm \sqrt{(c_0 - a^2 (1-c_2)) \, (1+c_2)^{-1}}$$