THE DISTRIBUTION OF THE MEAN

By E. L. Welker

University of Illinois

1. Summary. Both population and sample mean distributions can be represented or approximated by Pearson curves if the first four moments of the population are finite. Using the α_3^2, δ chart of Craig [2] to determine the Pearson curve type for the population, an analogous $\bar{\alpha}_3^2$, $\bar{\delta}$ chart is derived for the distribution of the mean. This defines a one to one transformation of α_3^2, δ into $\bar{\alpha}_3^2$, $\bar{\delta}$. The properties of this transformation are used to discuss the approach to normality of the distribution of the mean as dictated by the central limit theorem. This is facilitated by superposing on the α_3^2, δ chart the $\bar{\alpha}_3^2$, $\bar{\delta}$ charts for samples of 2, 5, and 10.

2. Introduction. For any given distribution function of a population, a method is available for finding the distribution function of the mean, when it exists, that depends on characteristic functions and the Fourier integral theorem. For example, characteristic functions have been used to show that the arithmetic means of samples from a normal population is normal, and, with minor restrictions on non-normal populations, that it is asymptotically normal. The method depends, of course, on a knowledge of the exact population distribution. Some authors have discussed the approximation of the distributions of sample means in special cases by one of the Pearson curves. It is the purpose of this paper to consider the complete range of Pearson curves as populations to be sampled, then to give the sampling distributions of the mean as approximated by the Pearson system, and to discuss the manner in which the distribution of the mean approaches the normal curve as dictated by the central limit theorem. Since the choice of a Pearson curve depends only on moment relationships, this will include the approximation of the distribution of the mean for any parent population as based on its moments. Both an algebraic and a graphic analysis will be given.

3. Semivariant and moment relationships. Denote by α_k the kth order moment of the population with zero mean and unit variance. Let λ_k be the kth order seminvariant of the population. Let $\bar{\alpha}_k$ and $\bar{\lambda}_k$ be the same parameters of the distribution of $\bar{x}$, the mean of a random sample of size N drawn from this parent population. Using properties of the seminvariants of linear functions of variables independent in the probability sense, formulas relating these parameters [1] are

$$\bar{\lambda}_k = \lambda_k N^{1-k},$$

$$\bar{\alpha}_3^2 = \bar{\lambda}_3^2 = \alpha_3^2 N^{-1},$$

and

$$\bar{\alpha}^4 = [\alpha^4 + 3(N-1)]N^{-1}.$$

4. The Pearson system of curves and the distribution of the mean. The determination of the Pearson curve will be made in accordance with the scheme discussed by C. C. Craig [2]. In this system the curve type is fixed by the moment α_3 and the constant

$$\delta = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3}.$$

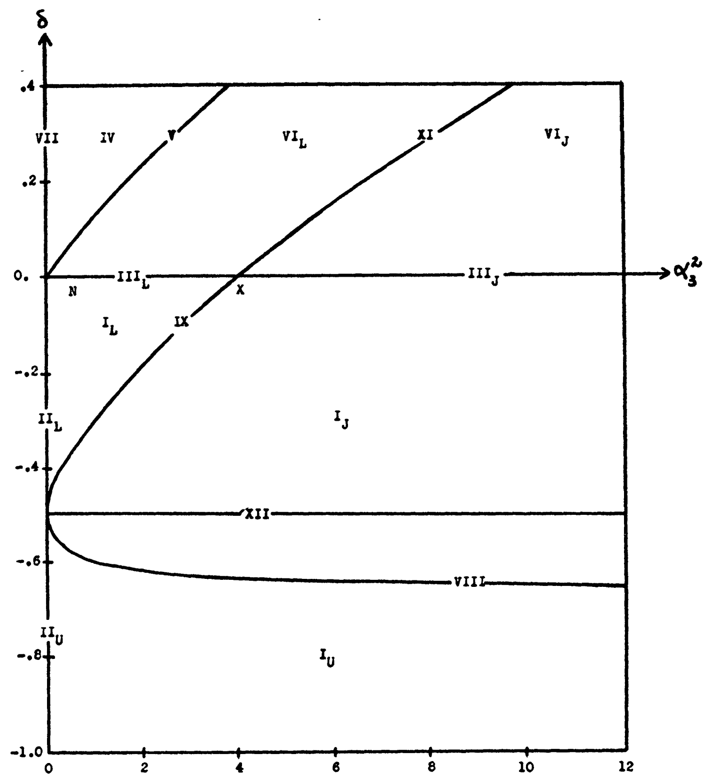

Fig. 1. The α_3^2, δ Chart for Pearson's Curves

The scheme for determining the type of curve is shown graphically in Fig. 1 in which the α_3^2, δ plane is divided into areas in which the Pearson curve types are noted. The bounding α_3^2, δ curves are

$$\delta = -1, \qquad \delta = -\tfrac{1}{2}, \qquad \delta = 0, \qquad \delta = \tfrac{2}{5}, \qquad \alpha_3^2 = 0,$$

$$\alpha_3^2 = 4\delta(\delta + 2), \quad \text{and} \quad (2 + 3\delta)\alpha_3^2 = 4(1 + 2\delta)^2(2 + \delta).$$

Let $\bar{\delta}$ denote the value of the δ function for the distribution of the mean. Then

$$\bar{\delta} = \frac{2\bar{\alpha}_4 - 3\bar{\alpha}_3^2 - 6}{\bar{\alpha}_4 + 3}.$$

In terms of moments of the parent population

$$\bar{\delta} = \frac{2\left[\dfrac{\alpha_4 + 3(N-1)}{N}\right] - 3\dfrac{\alpha_3^2}{N} - 6}{\dfrac{\alpha_4 + 3(N-1)}{N} + 3} = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3 + 6(N-1)}.$$

We see that $\bar{\delta} = \delta$ for $N = 1$, and $\bar{\delta} < \delta$ for $N > 1$. Both $\bar{\delta}$ and $\bar{\alpha}_3^2$ approach zero as N approaches infinity. These are the values of the constants for the normal function. This result is expected from the central limit theorem.

5. The $\bar{\alpha}_3^2$, $\bar{\delta}$ diagram for varying sample size. For every given population with finite moments of orders 1 through 4 there exists a Pearson curve representing or approximating its distribution. This determines a point in the α_3^2, δ plane. For a given sample size, N, there corresponds a point in the $\bar{\alpha}_3^2$, $\bar{\delta}$ plane. If the point $(\bar{\alpha}_3^2, \bar{\delta})$ is now plotted on the α_3^2, δ plane, we can determine the type of Pearson curve which is needed to approximate the distribution of x. The transformation of α_3^2, δ into $\bar{\alpha}_3^2$, $\bar{\delta}$ enables us to analyze the relationship between population distributions and distributions of $\bar{x}$. The transforms of the boundary curves in the α_3^2, δ plane will constitute an $\bar{\alpha}_3^2$, $\bar{\delta}$ chart corresponding to the one for α_3^2, δ shown in Fig. 1. In studying the approach to normality of the distribution of x, it is illuminating to superimpose this $\bar{\alpha}_3^2$, $\bar{\delta}$ chart on the α_3^2, δ chart. In order to do this, it is necessary to make certain algebraic changes in the equations.

First eliminate α_4 from the formula for δ as follows. From

$$\delta = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3} \quad \text{we find} \quad \alpha_4 = \frac{3\delta + 3\alpha_3^2 + 6}{2 - \delta}.$$

Substitute this in the expression for $\bar{\delta}$. Then

$$\bar{\delta} = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3 + 6(N-1)} = \frac{\delta(\alpha_3^2 + 4)}{\alpha_3^2 + 4 + 2(N-1)(2-\delta)}.$$

This formula, in conjunction with

$$\alpha_3^2 = N\bar{\alpha}_3^2$$

enables us to write the transformations of the boundary curves.

Boundary Curve	*Transformed Curve*
$\delta = -1$	$\bar{\delta} = \dfrac{-(N\bar{\alpha}_3^2 + 4)}{N\bar{\alpha}_3^2 + 4 + 6(N-1)}.$
$\delta = -\frac{1}{2}$	$\bar{\delta} = \dfrac{-(N\bar{\alpha}_3^2 + 4)}{2(N\bar{\alpha}_3^2 + 4) + 10(N-1)}.$
$\delta = 0$	$\bar{\delta} = 0.$

$$\delta = \tfrac{2}{5} \qquad\qquad \bar{\delta} = \frac{2(N\bar{\alpha}_3^2 + 4)}{5(N\bar{\alpha}_3^2 + 4) + 16(N - 1)}.$$

$$\alpha_3^2 = 4\delta(\delta + 2) \qquad \bar{\alpha}_3^2[N\bar{\alpha}_3^2 + 4 + 2\bar{\delta}(N - 1)]^2$$
$$= 4\bar{\delta}(\bar{\alpha}_3^2 + 4)[\bar{\delta}(N\bar{\alpha}_3^2 + 8N - 4) + 2N\bar{\alpha}_3^2 + 8].$$

$$(2 + 3\delta)\alpha_3^2 = 4(1 + 2\delta)(2 + \delta) \qquad [\bar{\delta}(16N + 3N\bar{\alpha}_3^3 - 4) + 2N\bar{\alpha}_3^2 + 8]$$
$$[N\bar{\alpha}_3^2 + 4 + 2\bar{\delta}(N - 1)]^2 N\bar{\alpha}_3^2$$
$$= 4[\bar{\delta}(2N\bar{\alpha}_3^2 + 10N - 2) + N\bar{\alpha}_3^2 + 4][\bar{\delta}(N\bar{\alpha}_3^2 + 8N - 4) + 2N\bar{\alpha}_3^2 + 8].$$

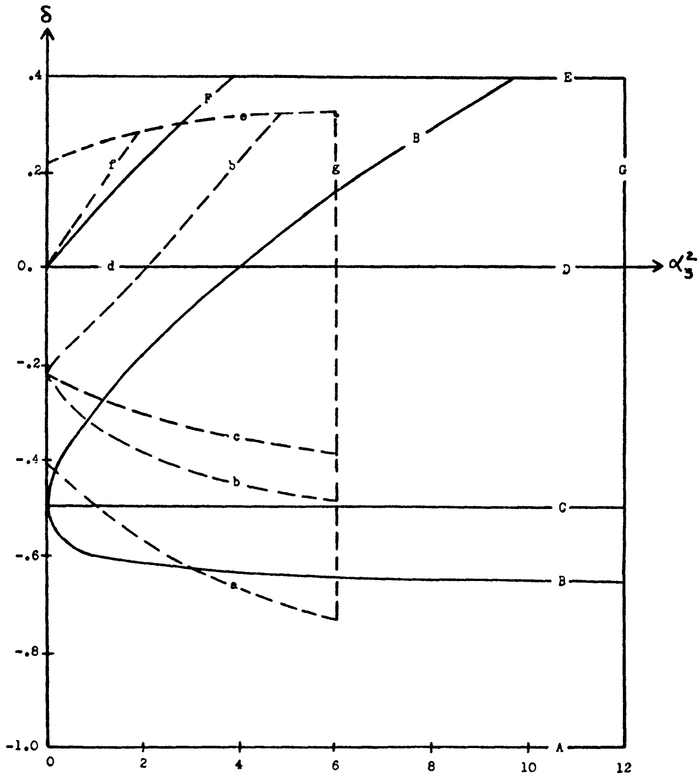

Fig. 2. The $\bar{\alpha}_3^2$, $\bar{\delta}$ and α_3^2, δ Charts. $N = 2$.

Fig. 2 shows the chart for distributions of $\bar{x}$ for $N = 2$ by dashed curves superimposed on the chart for the population shown by the solid curves, and Fig. 3 consists of the same curves for $N = 5$ and $N = 10$. The intervals on the population values are $0 \leq \alpha_3^2 \leq 12$ and $-1 \leq \delta \leq .4$ in Fig. 2, but only part of the α_3^2 range is shown in Fig. 3. In each case the curves for the distribution of $\bar{x}$ cover the interval for α_3^2, δ which corresponds to the entire interval shown for the population in Fig. 2. Population curves are identified by capital letters and the corresponding curves for the distribution of $\bar{x}$ by the corresponding lower case letters.

Before discussing the Pearson curve relationships disclosed by these graphs, let us analyze some of the geometric properties of the transformation itself. Let N be considered as the parameter defining families of curves in the $\bar{\alpha}_3^2$, $\bar{\delta}$ plane corresponding to α_3^2 = constant and δ = constant, the systems of lines parallel to the coordinate axes. The transform of $\alpha_3^2 = k$ is $\bar{\alpha}_3^2 = k/N$, a system of lines perpendicular to $\bar{\delta} = 0$, and approaching $\bar{\alpha}_3^2 = 0$ with increasing N at the rate kN^{-2}. The line $\alpha_3^2 = 0$ is invariant under the transformation, but it is not pointwise invariant.

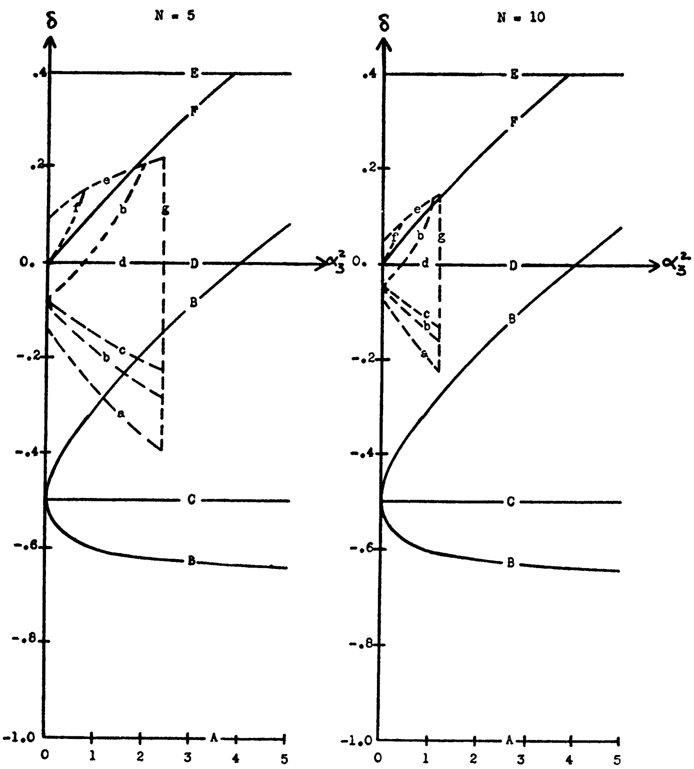

Fig. 3. The $\bar{\alpha}_3^2$, $\bar{\delta}$ and α_3^2, δ Charts

The transform of $\delta = C$ is

$$\bar{\delta} = \frac{C(N\bar{\alpha}_3^2 + 4)}{N\bar{\alpha}_3^2 + 4 + 2(N-1)(2-C)}, \text{ or } \bar{\delta}' = \frac{C(\bar{\alpha}_3^2 + 4/N)}{\bar{\alpha}_3^2 + [4 + 2(N-1)(2-C)]N^{-1}}.$$

Solving for $\bar{\alpha}_3^2$, this becomes

$$\bar{\alpha}_3^2 = \frac{4C - \bar{\delta}[4 + 2(N-1)(2-C)]}{N(\bar{\delta} - C)}.$$

Except for the straight line $\bar{\delta} = 0$, obtained when $C = 0$, this is a system of rectangular hyperbolas with asymptotes

$$\bar{\alpha}_3^2 = -[4 + 2(N - 1)(2 - C)]N^{-1} \quad \text{and} \quad \bar{\delta} = C.$$

We are concerned only with the range $\bar{\alpha}_3^2 > 0$. Hence

$$-[4 + 2(N - 1)(2 - C)]N^{-1}$$

must be positive for the asymptote to show on the diagram. Since $|\delta| < 2$, and thus $|C| < 2$, the expression in brackets is necessarily positive. Hence the vertical asymptote is always outside the range of interest and will not show on the diagram. However the horizontal asymptotes, $\bar{\delta} = C$, do appear in all cases. The hyperbolas are concave downward if $C > 0$ and are concave upward if $C < 0$.

Lines of the pencil $\delta = m\alpha_3^2$ are transformed into the hyperbolas

$$\bar{\delta} = \frac{m\bar{\alpha}_3^2(N\bar{\alpha}_3^2 + 4)}{\bar{\alpha}_3^2 - 2m\bar{\alpha}_3^2(N - 1) + 4}$$

for $N > 1$. It is clear that $(0, 0)$ is the only invariant point. Every point on $\delta = m\alpha_3^2$ is transformed into a point closer to the origin, the square of the distance from the origin changing from

$$(m^2 + 1)\alpha_3^4 \quad \text{to} \quad (m^2 + 1)\alpha_3^4 N^{-2}.$$

It is easily verified that the hyperbolas are asymptotic to

$$\bar{\delta} = \frac{mN\bar{\alpha}_3^2}{1 - 2m(N - 1)} - \frac{(N - 1)(1 + 2m)}{[1 - 2m(N - 1)]^2} \quad \text{and} \quad \bar{\alpha}_3^2 = \frac{-4}{1 - 2m(N - 1)}.$$

As N approaches infinity, these asymptotes approach

$$\bar{\delta} = -\frac{\bar{\alpha}_3^2}{2} \quad \text{and} \quad \bar{\alpha}_3^2 = 0.$$

An area in quadrant one (four) in the α_3^2, δ plane is transformed into an area in quadrant one (four) in the $\bar{\alpha}_3^2$, $\bar{\delta}$ plane. The transformed area is nearer the origin.

6. Types of Pearson curves for distribution of sample means. Examination of the graphs in conjunction with the above described properties of the transformation shows the following facts regarding the distribution of means of samples drawn from populations identified by α_3^2 and δ. First consider the normal function and the three main Pearson types only.

Parent Population	*Distribution of Sample Means*
Normal	Normal
I_L	I_L
I_J	I_J and I_L
I_U	I_U, I_J and I_L
IV	IV
VI_L	VI_L and IV
VI_J	VI_J, VI_L and IV.

The transition types were disregarded completely in the above analysis. It is worth noting that, disregarding type X, III is transformed into III, VII into VII, II_U into II_L, never into II_U, V into IV, but never into V. Type X is transformed into type III, never into X. Others follow a similar pattern.

These moment relationships on the distribution of the mean are not sufficient conditions in general. In special cases they are, for example the normal distribution and the type III (see [3]). They do represent the best approximation curve as specified by the Pearson system. We know that in some cases, for example type II (see [3]), the distribution of means is not described by a Pearson curve. It is clear, however, that the approach to normality is indicated analytically by the transformation α_3^2, δ to $\bar{\alpha}_3^2$, $\bar{\delta}$ and is shown graphically by the $\bar{\alpha}_3^2$, $\bar{\delta}$ diagram. Skewness and kurtosis in the parent population are reflected in the distribution of the mean in small samples. A symmetric distribution of the mean requires a symmetric parent population regardless of sample size, but the degree of skewness decreases rapidly with an increasing number in the sample. The Pearson curve which approximates the distribution of $\bar{x}$ from a bell-shaped parent population is also bell-shaped. The Pearson curve approximating the distribution of $\bar{x}$ for samples of $N = 10$ (Fig. 3) is bell-shaped for any parent population with values of α_3^2 and δ within the intervals considered. For samples of 5 in the same range the approximating curve is either bell-shaped or J-shaped, but it is never U-shaped. For samples of 2, even the U-shaped distribution is possible, but only with extreme values of α_3^2 and δ. The point in the α_3^2, δ plane corresponding to the normal curve is the only invariant point in the transformation. Hence parent populations with parameters not satisfying $\alpha_3^2 = \delta = 0$ cannot yield normal distributions of sample means.

REFERENCES

[1] T. N. THIELE, "The theory of observations", *Annals of Math. Stat.*, Vol. 2 (1931), p. 206.

[2] C. C. CRAIG, "A new exposition and chart for the Pearson system of frequency curves", *Annals of Math. Stat.*, Vol. 7 (1936), pp. 16–28.

[3] J. O. IRWIN, "On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's type II", *Biometrika*, Vol. 19 (1927), pp. 225–239.