NOTE ON THE BINOMIAL DISTRIBUTION

By C. E. Clark

The purpose of this note is to show that

$$(1) \qquad f(x) = (-1)^n \frac{q^n n!}{\pi} \left(\frac{p}{q}\right)^x \frac{\sin \pi x}{x^{(n+1)}}$$

where n is an integer $\geqq 0$, $0 < p < 1$, $p + q = 1$, and $x^{(n+1)} = x(x-1)(x-2) \cdots (x-n)$, is a function whose values at $x = 0, 1, 2, \cdots n$ are the successive terms of the expansion of $(q+p)^n$, and also to consider the problem of fitting $f(x)$ to an observed frequency distribution.

The statement made about (1) can be verified by evaluating (1) as an indeterminate form. On the other hand, (1) can be derived by observing that the x-th term (x an integer) of the expansion of $(q+p)^n$ is

$$(2) \qquad \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{\Gamma(n+1) p^x q^{n-x}}{\Gamma(x+1)\Gamma(n-x+1)};$$

then (1) can be derived from (2) by means of the product expansions for $\Gamma(x)$ and $\sin x$. This derivation of (1) from (2) can also be carried out by expressing (2) as a Beta function and then using

$$B(x+1, n-x+1) = \int_0^\infty \frac{t^x}{(1+t)^{n+2}} dt = (-1)^n \frac{\pi}{(n+1)!} \frac{x^{(n+1)}}{\sin \pi x}.$$

This integration can be performed by means of the theory of residues.

Consider the problem of fitting (1) to an observed frequency distribution. We shall write (1) in the form

$$(3) \qquad F(z) = ab^x \frac{\sin \pi x}{x^{(n+1)}}, \qquad x = \frac{nb}{1+b} + h(z - \bar{z})$$

and determine the constants a, b, n, and h so that, when $\bar{z}$ is the mean of the observed distribution, $F(z)$ will fit the distribution.

The values of a, b, n, and h can be determined by the method of moments. Let ν_2, ν_3, and ν_4, denote the usual second, third, and fourth moments of the distribution, which are calculated in the usual way (as in W. P. Elderton, *Frequency-Curves and Correlation*) and not adjusted by any procedure such as Sheppard's adjustments. Also, use the usual notation $\beta_1 = \frac{\nu_3^2}{\nu_2^3}$ and $\beta_2 = \frac{\nu_4}{\nu_2^2}$.

Then, the method of moments gives

$$(4) \qquad n = \frac{2}{3 + \beta_1 - \beta_2}$$

$$(5) \qquad b = \frac{2 + n\beta_1 \pm \sqrt{n\beta_1(4 + n\beta_1)}}{2}$$

$$h = \sqrt{\frac{nb}{\nu_2}}\left(\frac{1}{1 + b}\right)$$

$a = (-1)^n \dfrac{h(\Sigma f)n!}{\pi(1 + b)^n}$, where Σf is the sum of the frequencies of the distribution. An integer n is chosen nearest the value assigned by (4). The two values of b from (5) determine two curves that are congruent but whose skewnesses are of opposite sign. Hence, b is uniquely determined by (5) and the sign of the skewness of the data.

For a symmetrical distribution, $b = 1$, $\nu_3 = 0$, and

$$n = \frac{2}{3 - \beta_2}$$

$$h = \frac{\sqrt{n}}{2\sqrt{\nu_2}}.$$

We shall consider an illustrative example. In the following table the columns $f(z)$ and $f_2(z)$ are taken from W. P. Elderton, *Frequency-Curves and Correlation* (1906), page 62. $f(z)$ is an empirical frequency distribution, while $f_2(z)$ is obtained by fitting a Pearson Type II curve to the distribution $f(z)$. $f_1(z)$ is computed from

$$f_1(z) = 1624 \frac{\sin \pi x}{x^{(6)}}, \qquad x = 2.0973 + .808z$$

which is determined by the method of this note. $f_3(z)$ is obtained by fitting the normal curve

$$f_3(z) = 485.1e^{-\frac{(z-.4985)^2}{2(1.829)}}$$

z	$f(z)$	$f_1(z)$	$f_2(z)$	$f_3(z)$
−3	11	18	14	19
−2	116	107	109	92
−1	274	281	286	263
0	451	438	433	444
1	432	437	433	444
2	267	267	285	263
3	116	106	109	92
4	16	18	14	19

The coefficients of goodness of fit for $f_1(z)$, $f_2(z)$, and $f_3(z)$ are respectively .35, .58, and .02.