ON CERTAIN RELATIONSHIPS BETWEEN β_1 AND β_2 FOR THE POINT BINOMIAL.*

By MARGARET MERRELL

1. *Introduction.*

The extensive literature on the point binomial covers studies on a variety of its properties, apropos of its use as a discrete probability function and of its approximate representation by certain continuous curves. Investigations on such properties as the sum of its terms within specified limits, the ratio of its ordinates, and the slope of chords connecting successive ordinates have thrown light on the characteristics of the point binomial and have suggested various continuous functions as substitutions for the binomial expansion.

Prominent among such studies have been those dealing with the moments of the binomial and it is with these properties that the present paper is concerned. The first four moments have been used by Pearson[1] as a means of fitting a point binomial to observed data and he has pointed out[2] that these moments expressed in terms of β_1 and β_2 approach the corresponding moments of the normal curve as n becomes indefinitely large. Other papers that especially concern the following discussion are one by Student[3] in which he discussed the relationship between β_1 and β_2 for the point binomial and the Poisson exponential series, and one by Lucy Whitaker,[4] in which the range of β_1 and β_2 and certain relation-

*Paper No. 175 from the Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Md.

[1]Pearson, K. Skew variation in homogeneous material. Phil. Trans. Vol. *186 A,* (1895), pp. 343-414.

[2]Pearson, K. On the curves which are most suitable for describing the frequency of random samples of a population. Biometrika, Vol. *5* (1906) pp. 172-175.

[3]Student. On the error of counting with a haemacytometer. Biometrika, Vol. *5* (1906), pp. 351-360.

[4]Whitaker, Lucy. On the Poisson law of small numbers. Biometrika Vol. *10* (1914), pp. 36-71.

ships between the moments and the constants of the point binomial were discussed.

The present note will give some additional relationships between the third and fourth moments of the point binomial, in terms of β_1 and β_2, and will discuss these relationships in connection with their bearing on the association of the point binomial and the normal curve. The point binomial, $(p+q)^n$, has of course two constants p and n which completely determine its characteristics. Certain of these properties are closely connected with β_1 and β_2, and it is therefore of interest to see how β_1 and β_2 change as p and n take on different values. In order to see the effect of varying each of the constants, the relationship between β_1 and β_2 will be determined for varying values of n when p is held constant, and for varying values of p when n is held constant. In addition to these relationships, it is possible to see how the β's are related when both p and n are allowed to vary while certain functions of these parameters are held constant. In the following discussion the relationship between β_1 and β_2 will be considered for the cases where the mean, np, is held constant, and where the square of the standard deviation, npq, is held constant, n and p being variable.

The moments of the point binomial $(p+q)^n$ are:

$$\text{mean} = np$$

$$\mu_2 = npq$$

$$\mu_3 = npq(q-p)$$

$$\mu_4 = npq\left[1+3pq(n-2)\right]$$

These moments lead to the following values of the β's:

$$\beta_1 = \frac{(q-p)^2}{npq}, \tag{1}$$

$$\beta_2 = \frac{1+3pq(n-2)}{npq}. \tag{2}$$

Although the point binomial is ordinarily applied only to the case where p, q, and n are positive, it should be noted that these formulae for β_1 and β_2 are not limited to this case. The only limitation on these constants is that $p + q = 1$

2. *The relationship between β_1 and β_2 for constant values of p*

If we eliminate n between (1) and (2) we obtain an equation relating β_1, β_2, and p, n being unspecified. This equation is

$$\beta_2 = 3 + \frac{1-6pq}{(q-p)^2}\beta_1 \tag{3}$$

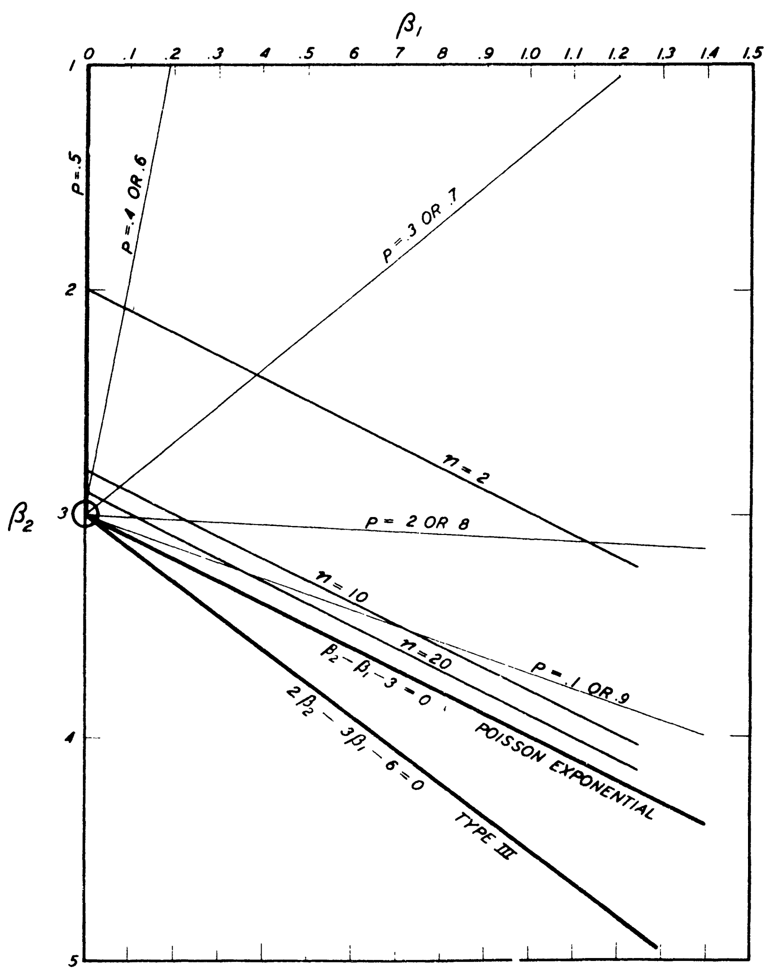

FIG. 1 · THE RELATIONSHIP BETWEEN β_1 AND β_2 FOR POINT BINOMIALS HAS CONSTANT VALUES OF p AND CONSTANT VALUES OF n

for fixed values of p, this represents a family of straight lines with slope $\dfrac{1-6pq}{(q-p)^2}$, all passing through the position $\beta_1 = 0, \beta_2 = 3$. The position of the β's for the normal curve. Figure 1 shows a group of these lines on the $\beta_1 \beta_2$ plane, for the values of p indicated. The various positions on these lines represent β_1 and β_2 for point binomials having the specified p's and varying n's. Only those values of p which are between 0 and 1 are included in this diagram, and these only for positive values of n, since these values cover all the ordinary probability problems. From (1) it can be seen that for such values of p and n, β_1 is positive. The β's for binomials having parameters outside these limits will be discussed in a later paragraph.

The range through which these lines can swing can be determined by substituting the limiting values of p in equation (3). The values of p to be used in determining this range are 0 and $.5$, not 0 and 1. This follows from the fact that, since p and q are interchangeable in the point binomial, and the slope of the lines given by (3) is symmetrical in p and q, the lines obtained for p between $.5$ and 1 would be identical with the lines for the complement of p, $1-p$, between 0 and $.5$. Thus any particular line represents point binomials for two complementary values of p.

If we substitute the value $.5$ for p in (3), the equation becomes

$$\beta_1 = 0 \tag{4}$$

as we would expect, since point binomials having a p or $.5$ are symmetrical. If $p = 0$, equation (3) becomes

$$\beta_2 - \beta_1 - 3 = 0. \tag{5}$$

The latter line is identical with the line giving the relationship between β_1 and β_2 in the Poisson exponential series. This is in harmony with the derivation of the Poisson exponential as the limiting case of the point binomial as p tends to 0, and n tends to ∞, np being finite. If we denote the mean of this series by m, the moments are[5]

$$\mu_2 = m$$

$$\mu_3 = m$$

$$\mu_4 = 3m^2 + m$$

and

$$\beta_1 = \frac{1}{m}$$

$$\beta_2 = 3 + \frac{1}{m}\,.$$

We have thus the equation relating β_1 and β_2 as given by (5).

The radiating lines giving the β's for point binomials having values of p between 0 and 1, will therefore lie in the range between the vertical, $\beta_1 = 0$, and the Poisson exponential line. This range, which is indicated in figure 1, was pointed out by Lucy Whitaker[6] in the paper mentioned above. The Type III line is included in this graph to indicate that portion of the $\beta_1 \beta_2$ plane covered by this family of lines. It is of interest to note that β_1 and β_2 for skew binomials do not approach the β's of the Type III curve, although Pearson[7] has shown that in an important slope property the skew binomial polygon and the Type III curve follow the same law.

3. The relationship between β_1 and β_2 for constant values of n.

The $\beta_1 \beta_2$ equations for constant values of p have been expressed as continuous straight lines, but only certain positions on these lines pertain to binomials having integral values of n. These points are determined by the intersection of these lines with the curve relating β_1, β_2, and n, when n is held constant at integral

[5]Student. Loc. cit. p. 353.

[6]Whitaker, Lucy. Loc. cit. p. 37.

[7]Pearson, K. Skew variation in homogeneous material. Loc. cit. p. 357.

values. The equation of this curve given by eliminating p and q between (1) and (2) is:

$$\beta_2 - \beta_1 - \frac{3n-2}{n} = 0. \tag{6}$$

This is a family of straight lines parallel to the Poisson exponential line, a specified value of n determining a particular line. The intersection of any of these lines with any p line determines the β's for the point binomial of specified p and n. Figure 1 gives the graph of three such lines for $n=2$, $n=10$, and $n=20$. From this graph it can be seen that even with an n as small as 20, β_1 and β_2 for the symmetrical and slightly skew binomials are not far from the position of the β's of the normal curve, but for the highly skew binomials, they are quite far from this position. This is in agreement with the fact that the more skew the binomial, the larger the n required to make the normal curve a good substitute for the binomial expansion.

It is evident from this graph that as n is fixed at increasingly large values, β_1 and β_2 for the point binomials of different p's converge quite rapidly toward the normal position. The limit of equation (6), as n becomes indefinitely large, is equation (5), that is, the line giving β_1 and β_2 for the Poisson exponential. This is to be expected, considering the conditions under which the point binomial approaches the Poisson exponential. For this limiting case the β_1, β_2 line for constant n crosses all the radiating p lines at $0, 3$, except the line for $p=0$, with which it coincides throughout. Thus for all values of p, except 0, β_1 and β_2 for the point binomial agree with the corresponding moments of the normal curve, as n becomes indefinitely great.

4. The relationship between β_1 and β_2 for constant values of np.

In judging how adequate the size of a particular sample is, for a specified value of p, we frequently make our estimate in terms, not of n, but of the mean value, np. This is with the thought that we are in approximately as good a position with a p of $.1$ and an n of 50, for example, as with a p of .01 and an n of 500, since the expected number is the same in both cases. For instance, our knowledge of a penny from 10 tosses is about as complete as that of a dice from 30 tosses. To study this question from the moments of the binomial we can determine the curve relating the β's for binomials of constant np, by replacing np by m in equations (1) and (2), and eliminating the remaining p's and q's between these two equations. This gives the equation

$$(7)\qquad \begin{aligned} &3m^2\beta_1^2 - 5m^2\beta_1\beta_2 + 2m^2\beta_2^2 + 3m(5m-2)\beta_1 \\ &\quad - 4m(3m-1)\beta_2 + 2(3m-1)^2 = 0 \end{aligned}$$

This is the equation of a hyperbola with asymptotes

$$(8)\qquad 2\beta_2 - 3\beta_1 - 6 = 0$$

$$(9)\qquad \beta_2 - \beta_1 - \frac{3m-2}{m} = 0.$$

The substitution of any particular value of m in equation (7) will give the curve of β's for all binomials having this specified mean value. Its intersection with the radiating lines of constant p gives β_1 and β_2 for the point binomial having the specified p and mean value. Figure 2 shows the two hyperbolas for which $m=2$ and 10 respectively.

Turning to the asymptotes of the hyperbola, it will be seen that only one of them varies with m. Thus the various hyperbolas obtained by substituting different values of m in (7) will all be asymptotic to the same line (8). The centers of all the hyperbolas will therefore lie on this line, which, it will be noted, is the Type III line. The other asymptote is parallel to the Poisson exponential line and a comparison of its equation, (9), with equation (6) shows that it represents the same relationship between β_1, β_2, and m, as that previously derived between β_1, β_2, and n. This asymptote is therefore the particular line in the family of n lines for which $n = m$ or np. Since any point on the hyperbola is the

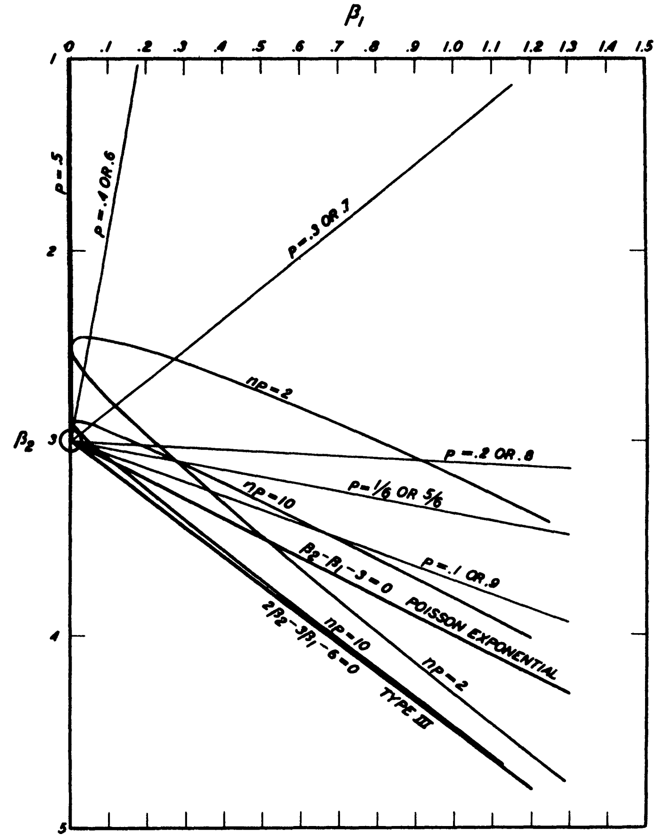

FIG. 2: THE RELATIONSHIP BETWEEN β_1 AND β_2 FOR POINT BINOMIALS HAVING CONSTANT VALUES OF np

intersection of n and p lines such that np has the constant value m, it follows that points farther and farther out on the hyperbola represent the intersection of lines having values of n closer and closer to the mean, np, and values of p approaching the value 1. If we consider this asymptote for hyperbolas of increasing values of m we see that it approaches the Poisson exponential line, and in the limiting case, as m becomes indefinitely large, the hyperbola degenerates into the two lines which represent the β's for the Poisson exponential series, and the Type III curve. These limits will be further discussed in a later paragraph.

All of the hyperbolas in the family represented by equation (7) are tangent to the line $\beta_1 = 0$, which is the value for binomials of $p = .5$. From figure 2 it can be seen that the lines giving the β's for the other values of p are crossed twice by each hyperbola. This is due to the fact that each line represents the β's for two complementary values of p, and for such values of p the same mean value would result from two different values of n, such that

$$m = np = n'(1-p)$$

For example, the binomials $(.2+.8)^{20}$ and $(.8+.2)^{5}$ have p values that lie on the same straight line and mean values that lie on the same hyperbola. It is thus obvious why the hyperbola must be tangent to the line for $p = .5$, since in this case the two complementary values of p are equal and there can be only one value of n which will produce a given m.

From the discussion of lines relating β_1 and β_2 for constant n values, it is evident that of the two crossings of any p line, the one nearer the Gaussian position is for the point binomial with the larger n and therefore for the smaller of the two complementary p values. Furthermore, through the point of intersection of two hyperbolas, only one n line and one p line will pass, and the β's thus determined are for two binomials, one having the smaller mean and the smaller p, and the other the larger mean and p, both having the same n. For example, the two hyperbolas given in figure 2, intersect at the point $\beta_1 = .26667, \beta_2 = 3.1$ The value of n for this position is 12, and for p is $1/6$ or $5/6$. These values of β_1 and β_2 are therefore for the binomials $(1/6+5/6)^{12}$, with the mean value 2, and $(5/6+1/6)^{12}$ with the mean value 10.

From figure 2, it is seen that the hyperbolas extend into the area between the Poisson exponential line and the Type III line. Since, as stated above all p's between 0 and 1 fall in the area between $\beta_1 = 0$ and the Poisson exponential line, it raises the question as to the meaning of the hyperbola outside that area. In the point binomial $[p+(1-p)]^n$, there is nothing to force p to lie between 0 and 1, and n to be positive except the conditions we impose for applications to probability problems. If we consider the general case, without these limitations, an analysis of equation (3) shows that the radiating lines giving β_1 and β_2 for fixed values of p continue into the area between the Poisson exponential and the Type III lines, the values of p in this area being either negative, or the complements of these negative values, that is, positive values greater than 1. This area includes all values of p from 0 to $-\infty$, and from 1 to $+\infty$. Thus, lines giving β_1 and β_2 for all real values of p from $-\infty$ to $+\infty$ are included between the vertical and the Type III lines. Outside this area, the values of p are imaginary. Turning to the values of n, we can see from equation (6) that in the family of parallel lines giving β_1 and β_2 for fixed values of n the lines below the Poisson exponential all have negative values of n. As n approaches $-\infty$,

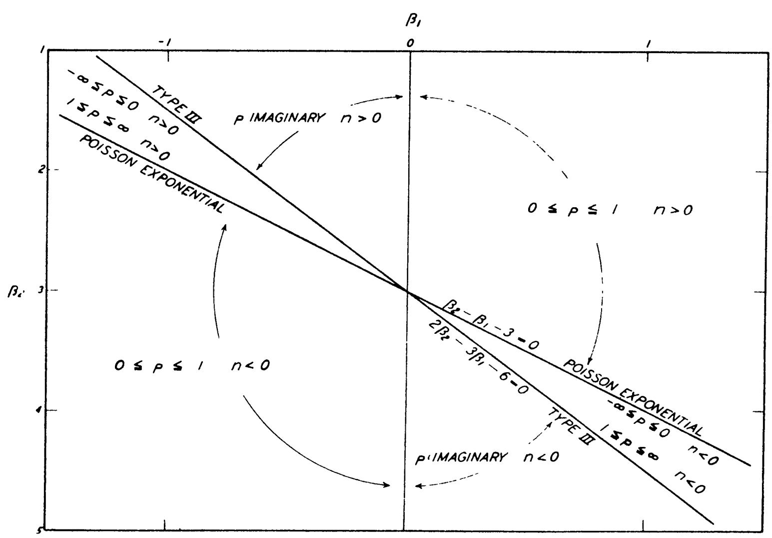

FIG 3 SUBDIVISION OF THE β_1, β_2 PLANE FOR BINOMIALS CLASSIFIED ACCORDING TO THE VALUES OF p AND n

this line approaches the Poisson exponential line from below. Figure 3 shows the subdivisions of the $\beta_1 \beta_2$ plane for the various cases.

It follows from this discussion that the values of the hyperbola in the area between the Poisson exponential and Type III lines give β's for point binomials with negative n's, and since np is a fixed positive value for any hyperbola, negative values of p, ranging from 0 to $-\infty$. These results are in harmony with the case mentioned above, where the hyperbola degenerates into two straight lines (the Poisson exponential line and the Type III line) as the mean becomes indefinitely large, for np approaches ∞ when either n or p approaches ∞. Thus, the Poisson exponential line is the limit when n approaches ∞, and $p = 1$, and the Type III line when p approaches $-\infty$, and n is negative.

5. *The relationship between β_1 and β_2 for constant values of npq.*

A further point of interest is the scatter of the β's for point binomials of varying p's but constant standard deviations. In equations (1) and (2), if we let $npq = \sigma^2$, and eliminate p and q we have the equation

$$(10) \qquad 2\beta_2 - 3\beta_1 - \frac{6\sigma^2 - 1}{\sigma^2} = 0$$

This, like the lines giving β_1 and β_2 for constant values of n, is a family of parallel straight lines, but where the n lines were parallel to the Poisson exponential line, this group is parallel to the Type III line. These lines intersect the radiating lines giving the β's for constant values of p, and as σ^2 increases, the points of intersection of these lines approach the β_1 and β_2 for the normal curve. As σ^2 approaches ∞, the line given by (10) approaches the Type III line, and in the limiting case, crosses all the p lines at the Gaussian position.

6. Summary.

Certain relationships between the third and fourth moments of the point binomial in terms of β_1 and β_2 have been discussed and the following results have been brought out:

A. For fixed values of p, β_1 and β_2 are linearly related, forming a family of radiating lines, all passing through the position of the β's for the normal curve. Each of the lines represents the β's for point binomials having a fixed value p or its complement, $1-p$. The lines for values of p between 0 and 1 are included between the vertical $\beta_1 = 0$, which is the line for $p = .5$ and the Poisson exponential line, $\beta_2 - \beta_1 - 3 = 0$, which is the line for $p = 0$ or 1. The lines for negative values of p or positive values greater than 1, fall between the Poisson exponential line and the Type III line, $2\beta_2 - 3\beta_1 - 6 = 0$. For the rest of the plane; the values of p are imaginary.

B. Although it has been shown by Pearson that in certain slope properties the skew point binomial resembles the Type III curve, none of the binomials which we interpret as probability functions, that is, those having p between 0 and 1 and n positive, has β_1 and β_2 approaching those of the Type III curve, except for the special case where the Type III curve becomes identical with the normal curve.

C. For fixed values of n, β_1 and β_2 determine a series of straight lines parallel to the Poisson exponential line. For positive values of n, these lines are above the Poisson exponential line, and for negative values below this line. Intersections of these lines with the radiating p lines determine β_1 and β_2 for the point binomial of specified p and n. As n is held constant at increasingly great values, the points of intersection are closer and closer to the position of the β's for the normal curve, and in the limiting case the line of β's for constant n intersects at the normal position all of the family of p lines except the line for $p = 0$, with which it coincides.

D. β_1 and β_2 for point binomials of varying p and n, but constant mean values, lie on a family of hyperbolas, a particular hyperbola being determined by a specified mean. One of the asymptotes of all these hyperbolas is the Type III line, and the other is a line parallel to the Poisson exponential line, at a distance from it, depending on the value of the mean. The limit of this asymptote as the mean approaches ∞ is the Poisson exponential line. These hyperbolas are tangent to the line $\beta_1 = 0$, (the line for $p = .5$) and cross the other p lines twice, the intersection of the hyperbola and any p line or any n line determining the β's for the point binomial whose p and n are defined by the intersection.

E. For varying p and n, but fixed npq, β_1 and β_2 lie on a family of straight lines parallel to the Type III line, one line in the group being determined by a particular npq. The limit of these lines as npq is held constant at increasingly large values is the Type III line.

Margaret Merrell.