182

THE NEGATIVE BINOMIAL DISTRIBUTION

By R. A. FISHER, F.R.S.

Although the algebra of the two cases is equivalent, the positive and negative binomial expansions play very different parts as statistical distributions.

The positive binomial $$(q+p)^n$$
occurs normally with n a known integer, but the fractions p and $q = 1-p$, unknown. The case in which n also is unknown is conceivable, but rather artificial for the following reasons:

If n is not integral the expansion after a certain stage develops negative coefficients; these cannot be interpreted as negative frequencies, so that the expansion does not correspond with any distribution.

There remains the case in which n is necessarily integral, although unknown. A variety of problems may be constructed of this sort, all entirely academic. With a sufficiently large sample n is necessarily one less than the number of frequency classes, and is thus determined without reference to the actual frequencies.

The negative binomial, on the other hand, which, following Haldane (1941), we may write

$$(q-p)^{-k}, \quad q = 1+p, \ k \text{ positive},$$

gives on expansion the term $$q^{-k}\frac{(k+x-1)!}{x!\,(k-1)!}\left(\frac{p}{q}\right)^x,$$

which is positive for all positive values of x, whether k is integral or not. Consequently, in this case, there is a practical problem in the simultaneous estimation of p and k to which the positive binomial offers no analogy.

In experimental sampling the negative binomial with unknown exponent arises in a simple extension of the conditions which give rise to the Poisson Series. The Poisson Series arises when equal samples are taken from perfectly homogeneous material. It is completely determined by the average or expected number, m, of occurrences per sample. If unequal samples were taken, or if the material were not perfectly homogeneous, the value of m would vary from sample to sample. Since m is necessarily positive, the simplest frequency distribution which allows some variation of m is the Eulerian distribution, familiar as that of χ^2, in which the frequency element is

$$df = \frac{1}{(k-1)!}\,p^{-k}m^{k-1}e^{-m/p}\,dm.$$

For χ^2 the parameter k is always the half of a positive integer; in general it may be any number exceeding zero.

When m varies in this way the frequency of occurrence of x units in the sample is

$$\int_0^\infty \frac{1}{(k-1)!}\,p^{-k}m^{k-1}e^{-m/p}\,.\,e^{-m}\frac{m^x}{x!}\,.\,dm.$$

This integral also is of the Eulerian type having the value

$$\frac{(k+x-1)!}{x!(k-1)!}\frac{p^x}{(1+p)^{k+x}},$$

and this is identical with the standard form for the negative binomial. The variance of m always increases the variance of x for a given mean value, so that a positive binomial distribution cannot be obtained in this way, for it would correspond with m having a negative variance.

2. The efficiency of fitting the first two moments

The binomial with known exponent is efficiently fitted by the observed mean; it is therefore rational, and not inconvenient, to fit the negative binomial, using the first two moments. Jeffreys (1939) has pointed out that this process is not efficient.

The expression for the moments of the negative binomial are equivalent to those for the positive binomial, changing the sign of p, and remembering that k corresponds to $-n$, and $q = 1+p$.

$$\mu_1' = pk, \quad \mu_3 = pq(q+p)\,k,$$
$$\mu_2 = pqk, \quad \mu_4 - 3\mu_2^2 = pq(1+6pq)\,k.$$

Consequently, for large samples, for which case alone the method of moments need be investigated, we may use the equations of estimation

$$p = \frac{m_2-\bar{x}}{\bar{x}}, \quad k = \frac{\bar{x}^2}{m_2-\bar{x}},$$

where $\bar{x}$ is the mean, and m_2 the variance as estimated from the sample.

To examine the efficiency of the method we shall need the determinant of the covariance matrix of p and k so estimated; this may be found as follows without determining the covariance matrix itself.

The covariance matrix of $\bar{x}$ and m_2 for large samples of N is in general

$$\frac{1}{N}\begin{Bmatrix}\mu_2 & \mu_3 \\ \mu_3 & \mu_4-\mu_2^2\end{Bmatrix};$$

substituting for p and k, this gives the determinant

$$\frac{1}{N^2}\begin{vmatrix} pqk & pq(q+p)\,k \\ pq(q+p)\,k & pq(1+6pq)\,k+2p^2q^2k^2 \end{vmatrix}$$
$$= \frac{2}{N^2}p^3q^3k^2(k+1).$$

To derive from this the determinant of the covariance matrix for the estimates of p and k, we need only multiply by the square of the Jacobian

$$\frac{\partial(p,k)}{\partial(\bar{x},m_2)},$$

writing for $\bar{x}$ and m_2 their expected values. The Jacobian is

$$\begin{vmatrix} -\dfrac{m_2}{\bar{x}^2} & \dfrac{1}{\bar{x}} \\ \dfrac{\bar{x}(2m_2-\bar{x})}{(m_2-\bar{x})^2} & -\dfrac{\bar{x}^2}{(m_2-\bar{x})^2} \end{vmatrix} = \frac{-1}{m_2-\bar{x}},$$

or $-1/p^2k$ on substitution. The determinant of the covariance matrix of p and k estimated by the first two moments is, therefore,

$$\frac{2q^3(k+1)}{pN^2}.$$

We may compare this with the corresponding determinant for any method of efficient estimation.

The most convenient way of doing this is to calculate the information matrix, which will be the reciprocal of the covariance matrix for efficient estimation.

Taking the general term of the negative binomial,

$$C = \frac{(k+x-1)!}{x!\,(k-1)!} \cdot \frac{p^x}{(1+p)^{k+x}},$$

it appears that

$$-\frac{\partial^2}{\partial p^2}\log C = \frac{x}{p^2} - \frac{k+x}{(1+p)^2},$$

whence, substituting its mean value pk for x, we have

$$i_{pp} = \frac{k}{pq},$$

in accordance with the well-known fact that if k were given, p would be efficiently estimated from the observed mean. Next

$$-\frac{\partial^2}{\partial p\,\partial k}\log C = \frac{1}{q}, \quad \text{or} \quad i_{pk} = \frac{1}{q}.$$

Finally,

$$-\frac{\partial^2}{\partial k^2}\log C = F(k-1) - F(k+x-1) = \frac{1}{k^2} + \frac{1}{(k+1)^2} + \dots + \frac{1}{(k+x-1)^2},$$

and this expression averaged for varying x gives i_{kk} in the form

$$i_{kk} = \sum_{x=0}^{\infty} \frac{(k+x-1)!}{(k-1)!\,x!} \cdot \frac{p^x}{q^{k+x}} \left(\frac{1}{k^2} + \frac{1}{(k+1)^2} + \dots + \frac{1}{(k+x-1)^2}\right).$$

It is a curious fact that this awkward looking expression can be transformed into one suitable for the comparison we have in view. If

$$r = p/q, \quad 1/q = 1-r,$$

and

$$i_{kk} = (1-r)^k \left\{kr\left(\frac{1}{k^2}\right) + \frac{k(k+1)}{2} r^2 \left(\frac{1}{k^2} + \frac{1}{(k+1)^2}\right) + \frac{k(k+1)(k+2)}{k} r^3 \left(\frac{1}{k^2} + \frac{1}{(k+1)^2} + \frac{1}{(k+2)^2}\right) + \dots\right\}$$

$$= \frac{r}{k} + \frac{r^2}{2k(k+1)} + \frac{4r^3}{6k(k+1)(k+2)} + \dots$$

$$= \sum_{x=1}^{\infty} \frac{r^x}{x} \frac{(x-1)!\,(k-1)!}{(k+x-1)!}.$$

In this form it is easily seen that the determinant of the information matrix

$$N \begin{vmatrix} i_{pp} & i_{pk} \\ i_{pk} & i_{kk} \end{vmatrix}$$

is simply

$$\frac{N^2}{pq} \sum_{x=2}^{\infty} \frac{1}{x} \frac{p^x}{q^x} \frac{(x-1)!\,k!}{(k+x-1)!}.$$

If the determinant of the covariance matrix corresponding with any method of estimation is multiplied by this expression, we have the reciprocal of the efficiency. For the method of moments

$$\frac{1}{E} = 1 + \frac{4}{3}\frac{p}{q(k+2)} + 3\frac{p^2}{q^2(k+2)(k+3)} + \dots.$$

In accordance with the general theory E is always less than unity. The expression shows that it is near to unity when

$$\frac{p}{q(k+2)} = \frac{\bar{x}}{(k+\bar{x})(k+2)}$$

is small.

When the mean is small, e.g. $\bar{x} = 0{\cdot}1$, high efficiencies occur even when k is as low as unity, for which value the expression above is $1/33$, low efficiencies are confined to the region where $k \to 0$. At this extreme if $k > 9\bar{x}$ the value is less than $1/20$.

When the mean is $1{\cdot}0$, k must be as high as 3 for the value to fall to $1/20$.

When the mean is 10, k must be 9 for the value to fall to $1/20{\cdot}9$.

However high $\bar{x}$ may be, values of k above 18 will bring this down below $1/20$.

Thus if p is less than $1/9$ for any value of k, or if k exceeds 18 for any value of p, high efficiency is assured; for intermediate values, if the product $(1 + 1/p)(k+2)$ exceeds 20, the efficiency is satisfactorily high.

3. Numerical examples

Example 1. Table 1 gives a sample of sheep classified according to the number of ticks found on each. (Data due to A. Milne, King's College, Newcastle-on-Tyne.)

Table 1

Number of ticks (x)	Number of sheep f	$f(x-3)$	$f(x-3)^2$
0	7	−21	63
1	9	−18	36
2	8	− 8	8
3	13	—	—
4	8	8	8
5	5	10	20
6	4	12	36
7	3	12	48
8	—	—	—
9	1	6	36
10	2	14	98
Total	60	+15	353

Fitting by the first two moments we have

$$\bar{x} = 3{\cdot}25, \quad m_2 = s^2 = 349{\cdot}25 \div 59 = 5{\cdot}9194915,$$

giving the estimates

$$p = 0{\cdot}821382, \quad 1/p = 1{\cdot}217460, \quad k = 3{\cdot}956746,$$

and

$$\left(1+\frac{1}{p}\right)(k+2) = 13{\cdot}21.$$

From this we may guess the efficiency to be about 90 %. The actual terms in $1/E$ are

$$\begin{array}{l} 1 \\ 0{\cdot}1009 \\ 0{\cdot}0147 \\ 0{\cdot}0027 \\ 0{\cdot}0006 \\ \underline{0{\cdot}0001} \\ 1{\cdot}1190 \quad E = 0{\cdot}8937. \end{array}$$

With efficiency below 90 % many workers would think a more accurate fitting desirable. For this purpose the method of Haldane's note in this number may be recommended.

Example 2. As an example with a somewhat heavier rate of infestation we may take the series

Table 2

Ticks	Sheep	Ticks	Sheep	Ticks	Sheep
0	4	9	2	18	—
1	5	10	2	19	1
2	11	11	5	20	—
3	10	12	—	21	1
4	9	13	2	22	1
5	11	14	2	23	1
6	3	15	1	24	—
7	5	16	1	25	2
8	3	17	—	—	—
				Total	82

Here $\bar{x} = 538 \div 82 = 6{\cdot}5609756, \quad s^2 = 34{\cdot}767841.$

The moment estimates are

$$p = 4{\cdot}299188, \quad k = 1{\cdot}526096$$

and

$$\left(1+\frac{1}{p}\right)(k+2) = 4{\cdot}35.$$

In this case no further calculation is needed to show that the method of moments is decidedly inefficient.

The result of fitting this example by maximal likelihood gives, of course, somewhat different estimates

Fitted by	1st and 2nd moments	Likelihood
p	4·299188	3·691175
k	1·526096	1·777476

but such differences would scarcely mislead one as to the level of efficiency. The efficient values obtained by likelihood merely give a value of $(1 + 1/p)(k + 2)$ of 4·80 in place of 4·35.

The efficient solution does not give a bad fit, in spite of the abrupt changes in frequency, e.g. between 1 and 2 ticks, or again between 5 and 6, which the observed series shows.

Grouping in eleven classes we have

Number of ticks	Expected number of sheep (m)	Observed a	Difference $a-m$	$\frac{(a-m)^2}{m}$
0	5·256	4	−1·256	0·3001
1	7·350	5	−2·350	0·7514
2	8·032	11	+2·968	1·0967
3	7·958	10	+2·042	0·5240
4	7·478	9	+1·522	0·3098
5	6·799	11	+4·201	2·5957
6	6·043	3	−3·043	1·5323
7–8	9·844	8	−1·844	0·3454
9–11	9·990	9	−0·990	0·0981
12–15	7·232	5	−2·232	0·6889
16–	6·018	7	+0·982	0·1602
Total	82·000	82	0	8·4026

Since, in addition to the total frequency, two parameters have been efficiently fitted, χ^2 has eight degrees of freedom. The value of χ^2 is thus very near to its expected value. In spite of their apparent regularity the deviations are no larger than might often be due to chance.

4. Summary

The cases of the positive and negative binomial distributions, in spite of their algebraic similarity, are very different in their applications, and in the statistical problems to which they give rise.

With the negative binomial we ordinarily require to estimate the exponent in addition to the mean of the distribution. This can be done from the first two moments, but the process has been recognized as inefficient, and in the present note the theoretical efficiency is calculated so as to make it easy to judge in practical cases whether a more exact fitting by maximal likelihood is required.

REFERENCES

J. B. S. Haldane (1941). 'The fitting of binomial distributions.' *Ann. Eugen., Lond.*, **11**, 179.

H. Jeffreys (1939). *Theory of Probability*, p. 260. Oxford: Clarendon Press.