

**Sampling Theory of the Negative Binomial and Logarithmic Series Distributions**



F. J. Anscombe

*Biometrika*, Vol. 37, No. 3/4. (Dec., 1950), pp. 358-382.

Stable URL:

<http://links.jstor.org/sici?sici=0006-3444%28195012%2937%3A3%2F4%3C358%3ASTOTNB%3E2.0.CO%3B2-L>

*Biometrika* is currently published by Biometrika Trust.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/bio.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## SAMPLING THEORY OF THE NEGATIVE BINOMIAL AND LOGARITHMIC SERIES DISTRIBUTIONS

By F. J. ANSCOMBE, *Statistical Laboratory, University of Cambridge*

### 1. INTRODUCTION

The negative binomial distribution depends on two parameters, which for many purposes may be conveniently taken as the mean  $m$  and the exponent  $k$ . The chance of observing any non-negative integer  $r$  is

$$P_r = \left(1 + \frac{m}{k}\right)^{-k} \frac{\Gamma(k+r)}{r! \Gamma(k)} \left(\frac{m}{m+k}\right)^r. \tag{1.1}$$

Sometimes it is more convenient to replace  $m$  by  $p$  or  $X$  defined by

$$p = \frac{m}{k}, \quad X = \frac{p}{1+p} = \frac{m}{m+k}. \tag{1.2}$$

Thus we may write

$$P_r = (1-X)^k \frac{\Gamma(k+r)}{r! \Gamma(k)} X^r. \tag{1.3}$$

We assume  $k, m, p > 0, 0 < X < 1$ . The factorial-cumulant-generating function is

$$\ln \mathbb{E}\{(1+t)^r\} \equiv \sum_{i=1}^{\infty} \kappa_{[i]} t^i / i! = -k \ln(1-pt), \tag{1.4}$$

and the  $i$ th factorial cumulant is  $\kappa_{[i]} = (i-1)! k p^i. \tag{1.5}$

The generating function of ordinary cumulants\* is

$$\ln \mathbb{E}(e^{tr}) \equiv \sum_{i=1}^{\infty} \kappa_i t^i / i! = -k \ln\{1-p(e^t-1)\}, \tag{1.6}$$

and the first four are

$$\left. \begin{aligned} \kappa_1 &= kp = m, \\ \kappa_2 &= kp(1+p) = m + m^2/k, \\ \kappa_3 &= kp(1+p)(1+2p), \\ \kappa_4 &= kp(1+p)(1+6p+6p^2). \end{aligned} \right\} \tag{1.7}$$

From (1.4) or (1.6) we see that the sum of  $N$  independent observations from the distribution has still a distribution of negative binomial form, with mean  $Nm$  and exponent  $Nk$ .

The logarithmic series distribution of R. A. Fisher is obtained by a limiting process from the negative binomial distribution by considering a sample of  $N$  readings, letting  $N$  tend to infinity and  $k$  to zero, and neglecting the zero readings. It is a multivariate distribution, consisting of a set of independent Poisson distributions with mean values

$$\alpha X, \quad \frac{1}{2}\alpha X^2, \quad \frac{1}{3}\alpha X^3, \quad \dots \tag{1.8}$$

A 'sample' comprises one reading from each Poisson distribution.

\* For a discussion of ordinary and factorial cumulants of a related distribution see Wishart (1947). Aitken (1939) and Haldane (1949) have pointed out that discrete distributions are often more conveniently described by factorial than ordinary cumulants, and this proves to be so for the distributions considered in § 2. The following relations may be noted:

$$\begin{aligned} \kappa_1 &= \kappa_{[1]}, \\ \kappa_2 &= \kappa_{[2]} + \kappa_{[1]}, \\ \kappa_3 &= \kappa_{[3]} + 3\kappa_{[2]} + \kappa_{[1]}, \\ \kappa_4 &= \kappa_{[4]} + 6\kappa_{[3]} + 7\kappa_{[2]} + \kappa_{[1]}. \end{aligned}$$

The main purpose of this paper is to carry somewhat further Fisher's investigations (Fisher, 1941; Fisher, Corbet & Williams, 1943) into the sampling properties of these distributions. The following is a brief summary of contents.

In § 2 the negative binomial form of distribution is compared with seven other two-parameter forms of distribution that have been proposed by various writers. It is shown that they can be arranged in order of increasing skewness and tail length, and that they vary in the number of modes possible in the frequency function. Thus while Neyman's Type A contagious distribution may have an unlimited number of modes, a distribution given by Pólya may have either one or two modes, and the negative binomial and a discrete form of the lognormal distribution have always one mode. The estimation of the distribution of local mean values in heterogeneous Poisson sampling is considered.

In § 3 the estimation of the parameters of a negative binomial distribution from a single large sample is considered.\* Alternatives to the maximum-likelihood method are described and their efficiencies indicated. Three such methods are found to be of practical importance: estimation by the first two sample moments, estimation by the first sample moment and the observed proportion of zero readings, and estimation with the aid of a transformation of the observations which makes the variance independent of the mean.

In § 4 two large-sample tests are described for discriminating between alternative forms of parent distribution. Each test is fully efficient in certain circumstances.

In § 5 the estimation of a common exponent from a series of samples is considered, when the parent populations possibly differ in their means. The results of §§ 3 and 5 have already been summarized by me elsewhere (1949), and their use discussed.

Finally, §§ 6 and 7 deal with the logarithmic series distribution. The estimation of  $\alpha$  by maximum likelihood, and some alternative formulae for its sampling variance, are discussed. Two tests of departure from the logarithmic series form of distribution are considered, one of them being due to Fisher and the other new.

*Notation.* The following notation will be used for the negative binomial distribution:

$m, k, p, X, P_r$  as defined above.

$N$  = total number of observations in sample.

$n_r$  = number of observations equal to  $r$  (for  $r \geq 0$ ).

$$\bar{r} = \sum_{r=0}^{\infty} n_r r / N = \text{mean of sample.}$$

$$s^2 = \sum_{r=0}^{\infty} n_r (r - \bar{r})^2 / (N - 1) = \text{variance estimate.}$$

For other distributions in § 2,  $k$  and  $p$  are defined so that the mean =  $kp$ , variance =  $kp(1 + p)$ .

The notation for the logarithmic series distribution will be:

$\alpha, p, X, n_r$  as defined above.

$$S = \sum_{r=1}^{\infty} n_r.$$

$$I = \sum_{r=1}^{\infty} n_r r.$$

\* The investigation may be compared with that of Shenton (1949) for Neyman's two-parameter Type A distribution.

An estimate of a parameter will be denoted by the same symbol with circumflex added. Different estimates of the same parameter are not distinguished in the notation, but only by context.

## 2. COMPARISON OF NEGATIVE BINOMIAL WITH OTHER DISTRIBUTIONS

A number of ways are known in which the negative binomial distribution can arise:

(1) *Inverse binomial sampling.* If a proportion  $\theta$  of individuals in a population possess a certain character, the number of observations in excess of  $k$  that must be taken to obtain just  $k$  individuals with the character has a negative binomial distribution with exponent  $k$  (Yule, 1910; Haldane, 1945).\*

(2) *Heterogeneous Poisson sampling.* If the mean  $\lambda$  of a Poisson distribution varies randomly from occasion to occasion, a 'compound Poisson distribution' results (Feller, 1943). We obtain a negative binomial with exponent  $k$  if  $\lambda$  has a Type III distribution, proportional to a  $\chi^2$  distribution with  $2k$  degrees of freedom (Greenwood & Yule, 1920).

(3) *Randomly distributed colonies.* If colonies or groups of individuals are distributed randomly over an area (or in time) so that the number of colonies observed in samples of fixed area (or duration) has a Poisson distribution, we obtain a negative binomial distribution for the total count if the numbers of individuals in the colonies are distributed independently in a logarithmic distribution (Lüders, 1934; Quenouille, 1949).

(4) *Immigration-birth-death process.* A certain simple model of population growth, in which there are constant rates of birth and death per individual and a constant rate of immigration, leads to a negative binomial distribution for the population size (McKendrick, 1914; Kendall, 1949). The model has been applied to the growth of some living populations, e.g. populations of bacteria, and to the spread of an infectious disease in a community.

The first of these, inverse binomial sampling, is the simplest mathematically, and is the only one where the mathematical model is likely to hold exactly in practice. While  $\theta$  may be unknown and require estimation,  $k$  is known, and the estimation problems discussed in the present paper are irrelevant. Inverse binomial sampling will therefore not be considered further.

In general heterogeneous Poisson sampling  $\lambda$  may be supposed to have a distribution  $dU(\lambda)$  with mean  $m$ . If  $\lambda$  has a cumulant-generating function

$$\phi(t) = \ln \mathbf{E} e^{t\lambda} = \sum_{i=1}^{\infty} K_i t^i / i!, \quad (2.1)$$

then  $\phi(t)$  is the factorial-cumulant-generating function of the distribution of the observed count  $r$ , and  $\phi(e^t - 1)$  the generating function of ordinary cumulants. Hence if  $\lambda$  has a Type III distribution with cumulant-generating function

$$\phi(t) = -k \ln(1 - mt/k), \quad (2.2)$$

$r$  has the negative binomial distribution required.

With the model of randomly distributed colonies, let the number of colonies observed per sample have a Poisson distribution with mean  $m_1$ , and let the number of individuals  $\rho$  per

\* It appears from I. Todhunter's *History* that the earliest general statement of the negative binomial distribution in this connexion was by Montmort in 1714. As to the other methods of deriving the distribution, a more detailed review has been given by Irwin (1941).

colony have a distribution with frequency function  $u_\rho$ . If the latter distribution has factorial-cumulant-generating function

$$\psi(r(t)) = \ln E(1+t)^\rho = \sum_{i=1}^{\infty} L_i t^i / i!, \tag{2.3}$$

then  $m_1(e^{\psi(t)} - 1)$  is the factorial-cumulant-generating function of the distribution of the total count  $r$ , and the first four factorial cumulants of  $r$  are

$$m_1 L_1, \quad m_1(L_2 + L_1^2), \quad m_1(L_3 + 3L_2 L_1 + L_1^3), \quad m_1(L_4 + 4L_3 L_1 + 3L_2^2 + 6L_2 L_1^2 + L_1^4).$$

We obtain the negative binomial with parameters  $p$  and  $k$  if  $m_1 = k \ln(1+p)$  and if  $\rho$  has the logarithmic distribution

$$u_\rho = \frac{1}{\rho \ln(1+p)} \left( \frac{p}{1+p} \right)^\rho \quad (\rho \geq 1), \tag{2.4}$$

which has mean  $m_2 = p/\ln(1+p)$  and factorial-cumulant-generating function

$$\psi(r(t)) = \ln \left\{ 1 - \frac{\ln(1-pt)}{\ln(1+p)} \right\}. \tag{2.5}$$

Models (3) and (4) for the negative binomial are closely associated, in that we may use (4) to justify (3). But it will be convenient below to consider models of randomly distributed colonies without specifying an evolutionary stochastic process that could give rise to it, and so the two models have been separated.

While we may expect that distributions closely resembling the negative binomial will often in fact be observed in population counts and in the sampling of heterogeneous material, it will not be surprising if sometimes the specific assumptions made in the above models are so wide of the mark that a substantially different form of distribution appears. Before embarking on a detailed study of the sampling properties of the negative binomial it will be as well to consider briefly what other distributions have been proposed that might perhaps fit such observations better. Attention will be confined to distributions having only two adjustable parameters; there seem to be seven of these outstanding in addition to the negative binomial. A convenient method of comparison is to express each distribution in terms of parameters  $k$  and  $p$  such that the mean and variance are  $kp$  and  $kp(1+p)$ , and then evaluate the third and fourth factorial cumulants. Results are shown in Table 1. The distributions can also be compared by computing specimen frequency functions; this is done in Tables 2 and 3.

The two-parameter contagious distribution of Type A of Neyman (1939) arises from the model of randomly distributed colonies in which the number of colonies per sample has a Poisson distribution with mean  $m_1$ , if we assume that the number of individuals per colony also has a Poisson distribution, say with mean  $m_2$ ,  $m_1$  and  $m_2$  being positive constants. A derivation along these lines has been given by Cernuschi & Castagnetto (1946), who, however, appear not to have recognized what they derived. The distribution can also arise from heterogeneous Poisson sampling, if  $\lambda$  has a discrete distribution and is equal to  $m_2 x$ , where  $x$  has a Poisson distribution with mean  $m_1$ . This is more or less the model Neyman used in deriving the distribution. The frequency function is, for  $r \geq 0$ ,

$$P_r = \frac{m_2^r}{r!} e^{-m_1} \sum_{j=0}^{\infty} \frac{j^r}{j!} (m_1 e^{-m_2})^j, \tag{2.6}$$

and its factorial-cumulant-generating function is

$$m_1(e^{m_2 t} - 1). \tag{2.7}$$

To express the mean and variance in the form  $kp$  and  $kp(1+p)$  we must set  $m_1 = k$ ,  $m_2 = p$ .

Neyman's two-parameter contagious distributions of Types B and C were derived from a more complicated model. The factorial-cumulant-generating function of the Type B is

$$m_1 \left\{ \frac{e^{m_2 t} - 1}{m_2 t} - 1 \right\}; \quad (2.8)$$

and we find  $\frac{3}{4}m_1 = k$ ,  $\frac{2}{3}m_2 = p$ . The factorial-cumulant-generating function of the Type C distribution is

$$m_1 \left\{ \frac{e^{m_2 t} - 1 - m_2 t}{\frac{1}{2}(m_2 t)^2} - 1 \right\}; \quad (2.9)$$

and  $\frac{3}{8}m_1 = k$ ,  $\frac{1}{2}m_2 = p$ . These two generating functions, (2.8) and (2.9), can be derived from that of Type A, (2.7), by a suitable integration. Thus to get (2.8) we replace  $m_2$  in (2.7) by  $x$  and  $m_1$  by  $(m_1/m_2) dx$ , and integrate for  $x$  between 0 and  $m_2$ ; while to get (2.9) we do the same except that  $m_1$  is replaced by  $[2m_1(m_2 - x)/m_2^2] dx$ . The observed variable  $r$  can therefore be regarded, in each case, as the limit of the sum of a large number of random variables following independent Type A distributions with values of the second parameter distributed in a range  $(0, m_2)$ .

Recently, Thomas (1949) has proposed a distribution very similar to Neyman's Type A. With the model of randomly distributed colonies, where the number of colonies in the sample has a Poisson distribution with mean  $m_1$ , the number of individuals per colony is assumed to be one plus an observation from a Poisson distribution with mean  $m_2 - 1$ .  $m_2$  is now a constant  $> 1$ . The factorial-cumulant-generating function of the distribution is

$$m_1 \{ (1+t) e^{(m_2-1)t} - 1 \}. \quad (2.10)$$

If  $m_2$  is large the distribution is close to the Neyman Type A with the same values of the parameters  $m_1, m_2$ . For  $m_2 - 1$  small, we may set

$$m_2 = 1 + \frac{1}{2}p + \frac{1}{8}p^2 + O(p^4), \quad m_1 m_2 = kp.$$

Pólya (1930) gives a distribution that arises from the model of randomly distributed colonies when the number of individuals per colony has a geometric distribution with frequency function

$$u_\rho = (1-\tau) \tau^{\rho-1}. \quad (2.11)$$

$\rho$  takes positive integer values, and  $\tau$  is a constant,  $0 < \tau < 1$ . The mean number of individuals per colony is  $m_2 = (1-\tau)^{-1}$ . The frequency function of the total observed count  $r$  per sample is given by

$$P_0 = e^{-m_1}, \quad P_r = e^{-m_1} \tau^r \sum_{j=1}^r \binom{r-1}{j-1} \frac{1}{j!} \left( \frac{m_1(1-\tau)}{\tau} \right)^j \quad (r \geq 1), \quad (2.12)$$

and the factorial-cumulant-generating function is

$$\frac{m_1 t}{1 - \tau - \tau t}. \quad (2.13)$$

We find  $m_1/(2\tau) = k$ ,  $2\tau/(1-\tau) = p$ . Pólya states that the distribution was given by A. Aeppli in a thesis in 1924. It will accordingly be referred to here as the Pólya-Aeppli distribution.

Preston (1948) has considered a distribution derived from the model of heterogeneous Poisson sampling, where it is supposed that  $\lambda$  has a lognormal distribution, i.e. that  $\ln \lambda$  has a normal distribution, say with mean  $\xi$  and variance  $\sigma^2$ . The distribution may be

conveniently referred to as the discrete lognormal distribution.\* It suffers from the disadvantage that its frequency function involves an untabulated integral;† for  $r \geq 0$ ,

$$P_r = \frac{m^r \tau^{\frac{1}{2}r(r-1)}}{r! \sqrt{(2\pi \ln \tau)}} \int_{-\infty}^{\infty} \exp \left[ -\frac{u^2}{2 \ln \tau} - m\tau^{r-\frac{1}{2}} e^u \right] du, \tag{2.14}$$

where  $m = \exp(\xi + \frac{1}{2}\sigma^2)$ ,  $\tau = \exp(\sigma^2)$ . However, the first few factorial cumulants are easily found, for they are the ordinary cumulants of the distribution of  $\lambda$ , and these are obtained (Finney, 1941) from the moments of  $\lambda$  about the origin,

$$\mu'_i = \exp(i\xi + \frac{1}{2}i^2\sigma^2) = m^i \tau^{\frac{1}{2}i(i-1)}. \tag{2.15}$$

The mean and variance of  $\lambda$  are therefore  $m$  and  $m^2(\tau - 1)$ ; and we have

$$(\tau - 1)^{-1} = k, \quad m(\tau - 1) = p.$$

Another rather intractable distribution derived from the model of heterogeneous Poisson sampling has been given by Fisher (1931), who supposed that  $\lambda$  was distributed like the square root of a Type III variable. The frequency function can be expressed in terms of  $Hh$ -functions, which have been tabulated to some extent. The cumulants involve  $\Gamma$ -functions. For the entry in Table 1 the limit  $p \rightarrow 0$  with  $kp$  constant has been considered. The distribution will be referred to as Fisher's  $Hh$ -distribution.

Table 1

Distribution	$\kappa_{[3]}/(kp^3)$	$\kappa_{[4]}/(kp^4)$
Thomas	$\frac{3}{4} + \frac{1}{2}p + O(p^2)$	$\frac{1}{2} + O(p)$
Fisher $Hh$	$1 + k^{-1} + O(p^2)$	$0 + O(p)$
Neyman A	1	1
Neyman B	$\frac{2}{3}$	$\frac{27}{20}$
Neyman C	$\frac{2}{5}$	$\frac{8}{5}$
Pólya-Aeppli	$\frac{2}{3}$	3
Negative binomial	2	6
Discrete lognormal	$3 + k^{-1}$	$16 + 15k^{-1} + 6k^{-2} + k^{-3}$

The third and fourth factorial cumulants of the above distributions are given in Table 1. It will be seen that, apart from Fisher's  $Hh$  distribution, they form a sequence of distributions of increasing skewness and tail length (leptokurtosis) in the order shown. The position of the  $Hh$  distribution relative to the others is ambiguous and variable (ambiguous in that it depends on whether we rank by  $\kappa_{[3]}$  or by  $\kappa_{[4]}$ , variable because it depends on the values of the parameters), but we may say at least that it should come somewhere towards the front of the list.

\* Preston does not give any exact sampling theory. Other writers (e.g. Williams, 1937; Gaddum, 1945) who have alluded to lognormal distributions in connexion with frequency counts have contented themselves with recommending that the data should be transformed by a logarithmic transformation of the form  $y = \log(r + c)$ , so as to appear approximately normal. It should also be noted that Preston is concerned with the situation where zero counts are not recorded and therefore the total sample size  $N$  is unknown. This will be discussed in §§ 6 and 7.

† A usable approximation to the frequency function of the discrete lognormal distribution has been developed by Dr P. M. Grundy.

The difference in shape between the distributions is clearly substantial if  $p$  is large. To demonstrate this further some expected frequencies are given in Table 2 for three distributions having  $p = 10$ , mean = 20, variance = 220, namely,

- (a) Neyman Type A, with  $m_1 = 2$ ,  $m_2 = 10$ ;
- (b) Pólya-Aeppli, with  $m_1 = 3\frac{1}{3}$ ,  $\tau = \frac{5}{6}$ ;
- (c) negative binomial, with  $k = 2$ ,  $m = 20$ .

Also shown is a distribution having mean = 20, variance = 218, namely,

- (d) Thomas, with  $m_1 = 2$ ,  $m_2 = 10$ .

To save space, the frequencies have been grouped. The Neyman distribution (a) has modes or peaks at  $r = 0, 10, 20$ , while at  $r = 30$  there is a mode in first differences which is not large enough to produce a mode in the frequencies themselves. The Thomas distribution (d) is practically indistinguishable from (a), the difference being that the modes of (d) are slightly more pronounced than those of (a). The Pólya-Aeppli distribution (b) has two modes only, at  $r = 0$  and 11. The negative binomial (c) has one mode, at  $r = 9$  and 10 (equal frequencies). If a discrete lognormal distribution were added to Table 2 (with  $m = 20$ ,  $\tau = 1.5$ ,  $\xi = 2.7930$ , and  $\sigma = 0.6368$ ) it would resemble (c) in having only one mode, but would be rather more skew; the frequencies for the first few values of  $r$  would be lower.

Table 2

$r$	Percentage frequency				$r$	Percentage frequency			
	(a)	(b)	(c)	(d)		(a)	(b)	(c)	(d)
0	<b>13.53</b>	<b>3.57</b>	0.83	<b>13.53</b>	17-18	5.04	5.52	5.77	5.09
1-2	0.07	4.18	3.55	0.03	19-20	<b>5.37</b>	5.18	5.28	<b>5.52</b>
3-4	0.72	4.96	5.31	0.54	21-22	5.23	4.81	4.79	5.33
5-6	2.74	5.52	6.34	2.56	23-24	4.74	4.42	4.31	4.74
7-8	5.54	5.89	6.86	5.66	25-26	4.19	4.02	3.86	4.15
9-10	<b>7.01</b>	<b>6.07</b>	<b>7.01</b>	<b>7.29</b>	27-28	3.78	<b>3.63</b>	<b>3.43</b>	<b>3.77</b>
11-12	6.41	<b>6.10</b>	6.90	6.47	29-30	3.48	3.25	3.03	3.51
13-14	5.18	6.00	6.62	5.00	31-32	3.19	2.89	2.67	3.23
15-16	4.73	5.80	6.22	4.59	33-∞	19.05	18.20	17.22	18.98

A less extreme comparison is shown in Table 3, which compares distributions having  $p = 3$ , mean = 6, variance = 24, namely,

- (a) Neyman Type A, with  $m_1 = 2$ ,  $m_2 = 3$ ;
- (b) Pólya-Aeppli, with  $m_1 = 2.4$ ,  $\tau = 0.6$ ;
- (c) negative binomial, with  $k = 2$ ,  $m = 6$ .

(a) and (b) have two modes, (c) has one.

In general, the Neyman Type A and Thomas distributions can have any number of modes from one upwards, and if there are several modes they will occur at values of  $r$  approximately equal to multiples of  $m_2$ . The Pólya-Aeppli distribution, on the other hand, has either one or two modes—two if  $2 < m_1 < (1 - \tau)^{-1}$ , one otherwise. The negative binomial distribution has always one mode. Presumably, by analogy, Fisher's  $Hh$  distribution has one or two modes, and the discrete lognormal one mode.



For the mere purpose of graduating data there is little to choose between the distributions in shape if  $p (=m/k)$  is not large, and then considerations of ease of handling weigh heavily in favour of the negative binomial. Experimental discrimination between the different forms of distribution is practicable, however, if  $p$  is large. An interesting attempt at such discrimination has been made by Beall (1940), who fitted Neyman Types A, B, C, Pólya-Aeppli, and negative binomial distributions to eleven series of counts of insect larvae. Some of the series seemed to indicate a bimodal population, and he concluded that they were well fitted by the Neyman forms of distribution, but not by the other two. Mr D. A. Evans has pointed out to me, however, that Beall fitted the latter distributions incorrectly, having mixed up the two parameters, and he was consequently unfair to them.

Table 3

$r$	Percentage frequency			$r$	Percentage frequency		
	(a)	(b)	(c)		(a)	(b)	(c)
0	<b>14.95</b>	<b>9.07</b>	6.25	9	5.23	4.92	4.69
1	4.47	8.71	9.37	10	4.42	4.12	3.87
2	7.36	9.41	<b>10.55</b>	11	3.67	3.40	3.17
3	8.77	<b>9.49</b>	<b>10.55</b>	12	3.00	2.78	2.57
4	<b>8.83</b>	9.12	9.89	13	2.42	2.25	2.08
5	8.29	8.46	8.90	14	1.92	1.80	1.67
6	7.60	7.63	7.79	15	1.51	1.43	1.34
7	6.86	6.71	6.67	16	1.17	1.13	1.06
8	6.06	5.80	5.63	17-∞	3.48	3.78	3.95

In analysing population counts we may have two quite distinct objects. On the one hand, the counts may have been made on plots in an experiment, and we desire some means of interpreting them so that the effects of treatments can be judged. What is usually done is to apply a transformation to make the method of analysis of variance appropriate. We study the distribution of the original counts in order to find a suitable transformation. It does not matter greatly whether the form of distribution fitted, if any, is very accurate or not.

On the other hand, we may be interested in relating the observed counts to some theory of population growth or spread. In that case we shall endeavour to use only forms of distribution that are 'biologically significant'. Neyman's distributions were intended to represent populations of insect larvae observed shortly after emergence from eggs, the eggs being supposed laid in clusters of a fixed size. The models seem rather special, and not likely to be widely applicable to other sorts of population counts. The characteristic feature of the distributions (at any rate, that of Type A) is the possibility of a series of three or more equally spaced modes. Unless such a series of modes is demonstrated conclusively by observation, one may reasonably feel reluctant to use such a model. Thomas's distribution was intended to represent plant quadrat counts, but no evidence has been adduced to make plausible the special form of  $u_p$  assumed, and again one may reasonably feel reluctant to use it. The derivation of the Pólya-Aeppli distribution from the model of randomly distributed colonies is much more promising. Kendall (1949) has shown that the progeny of a single individual after a fixed lapse of time will follow a geometric distribution with modified zero

term, in certain fairly general conditions of no competition. Hence if progenitors (e.g. plant seeds) are released randomly over an area at one time and their progeny (freely increasing by vegetative reproduction) are observed at a later time, we shall expect the number of individuals per quadrat to follow the Pólya-Aeppli distribution. The parameter  $m_1$  will be the mean number of progenitors per quadrat of which some progeny survive. If, instead of being released all at one time, the progenitors are released with uniform distribution in time from a particular time up to the present, and if the birth- and death-rates per individual are constant, we get the negative binomial, as already remarked at the beginning of this section. We may therefore expect that close approximations to both these distributions will in fact be observed in the study of growing populations. Of course, some population counts will not resemble any of the distributions we have considered, on account of overcrowding or, with mobile fauna, aggregating for reproduction, defence, or other social purposes. It is unlikely that any two-parameter distribution will describe such counts adequately.

In view of the difficulty of discriminating experimentally between forms of distribution arising from different mathematical models, the study of counts made all at one time is not likely to give reliable information on laws of population growth. For this purpose, repeated observations on the same population are needed, if possible with identification of individuals.

Sometimes, in sampling investigations, it is reasonable to suppose that the observations arise from heterogeneous Poisson sampling, but there may be no definite grounds for predicting the distribution of  $\lambda$ . If the observations are sufficiently numerous, the distribution of  $\lambda$  can be estimated (Newbold, 1927). Let  $k_1, k_2, \dots$ , be the  $k$ -statistics calculated from a sample of  $N$  observations on  $r$  (see, for example, Kendall, 1943). Then we can take as unbiased estimates of the first four cumulants  $K_1, K_2, K_3, K_4$ , of  $\lambda$  the following:

$$\left. \begin{aligned} \hat{K}_1 &= k_1, \\ \hat{K}_2 &= k_2 - k_1, \\ \hat{K}_3 &= k_3 - 3k_2 + 2k_1, \\ \hat{K}_4 &= k_4 - 6k_3 + 11k_2 - 6k_1. \end{aligned} \right\} \quad (2.16)$$

The right-hand sides are in fact unbiased estimates of the factorial cumulants of the distribution of  $r$ , analogous to the well-known  $k$ -statistics for ordinary cumulants. It is quite straightforward to calculate the variances (and other cumulants) of these estimates. If, for example,  $K_1 = m$ ,  $K_2 = \sigma^2$ , and all the  $K_i$  for  $i \geq 3$  are zero or negligible, we find (for  $N$  large)

$$\left. \begin{aligned} \text{var}(\hat{K}_1) &= \frac{m + \sigma^2}{N}, \\ \text{var}(\hat{K}_2) &= \frac{2(m + \sigma^2)^2 + 2\sigma^2}{N}, \\ \text{var}(\hat{K}_3) &= \frac{6(m + \sigma^2)^3 + 18\sigma^2(m + 3\sigma^2)}{N}. \end{aligned} \right\} \quad (2.17)$$

One possible application of this method is to estimating the process curve of the output of a production line from inspection records which give the number of defectives found in samples of a fixed size taken from each lot produced. A similar biological problem in surveying a district for presence of potato-root eelworm has been described by me elsewhere (1950). Newbold developed the method in a study of accident-proneness.

3. FITTING A NEGATIVE BINOMIAL DISTRIBUTION TO A LARGE SAMPLE

Haldane (1941) and Fisher (1941) have considered the maximum-likelihood method of fitting a negative binomial distribution to a large sample. If the distribution is expressed in terms of the parameters  $m$  and  $k$ , the maximum-likelihood estimates of the parameters turn out to be independent (asymptotically). For the estimate  $\hat{m}$  of  $m$  we have simply

$$\hat{m} = \bar{r}. \tag{3.1}$$

The estimate  $\hat{k}$  of  $k$  is the root of the following equation in  $x$ :

$$N \ln(1 + \bar{r}/x) = \sum_{j=1}^{\infty} n_j \left( \frac{1}{x} + \frac{1}{x+1} + \dots + \frac{1}{x+j-1} \right). \tag{3.2}$$

It is easy to show that the left-hand side is greater than the right-hand side if  $x$  is large enough, provided  $(N - 1) s^2 > N\bar{r}$  (or, ignoring the difference between  $N$  and  $N - 1$ , if  $s^2 > \bar{r}$ ); and also that the left-hand side is less than the right-hand side if  $x$  is small enough (but positive), provided that  $n_0 < N$ , so that there are some non-zero observations. Since both sides are continuous functions for  $x > 0$ , the equation must have at least one finite positive root. I have not proved that there is only one root in this case, and that if  $s^2 < \bar{r}$  there is none, but after an unsuccessful search for a *gegenbeispiel* I suppose both these statements to be true. If  $s^2 < \bar{r}$  (and if, indeed, there is no finite root), the excess of the right-hand side over the left-hand side tends to zero as  $x \rightarrow \infty$ , and we may say  $\hat{k} = \infty$ , implying that a Poisson distribution is being fitted. If  $n_0 = N$ , we may conventionally take  $\hat{k} = 1$ , say.

For the variance of  $\hat{m}$  we have

$$\text{var}(\hat{m}) = \frac{1}{N} \left( m + \frac{m^2}{k} \right). \tag{3.3}$$

From the matrix of expectations of second derivatives of the log-likelihood function, we obtain the large-sample formulae:

$$\text{cov}(\hat{m}, \hat{k}) \sim 0, \tag{3.4}$$

$$\text{var}(\hat{k}) \sim \frac{2k(k+1)}{NX^2} \left/ \left\{ 1 + \frac{4X}{3(k+2)} + \frac{3X^2}{(k+2)(k+3)} + \dots \right\} \right. \tag{3.5}$$

The summation involved in deriving the second of these is due to Fisher; the series in curly brackets may be written

$$1 + \sum_{j=2}^{\infty} \frac{2}{j+1} \frac{j! X^{j-1}}{(k+2)(k+3)\dots(k+j)}. \tag{3.6}$$

It may be noted in passing that large-sample variances and covariances found in this way relate to the asymptotic normal distribution of the estimates, and are not necessarily asymptotically equal to the actual variances and covariances of the estimates for finite  $N$ . In the present case, in sampling from any negative binomial distribution, there is always a positive chance, albeit perhaps a minute one, of finding  $s^2 < \bar{r}$ , when we should set  $\hat{k} = \infty$ .  $\hat{k}$  does not therefore have a proper distribution, nor variance. It does, however, have an asymptotic distribution with the variance given, as  $N \rightarrow \infty$ .

While equation (3.1) gives  $\hat{m}$  very simply, equation (3.2) for  $\hat{k}$  is tedious to solve, and it is worth while to consider alternative methods of estimating  $k$ . We begin with a general moment method. Let  $f_r$  be any specified convex or concave function (not linear) of the non-negative

integer  $r$ . Then we may consider  $\sum_{r=0}^{\infty} n_r f_r$  as a statistic for  $k$ . Let  $E(f_r)$ , expressed as a function of  $m$  and  $k$ , be denoted by  $F(m, k)$ . Then we shall take as our estimate  $\hat{k}$  the root of the following equation in  $x$ :

$$\frac{1}{N} \sum_{r=0}^{\infty} n_r f_r = F(\bar{r}, x). \tag{3.7}$$

The right-hand side of this equation is a monotone function of  $x$ , with no constant stretches, for  $x$  positive, provided  $\bar{r} > 0$ . For if  $x$  is increased to  $x + \delta x$ , the change  $\delta P_r$  in  $P_r(m, x)$  is negative for  $0 \leq r \leq a$  and for  $r \geq b + 1$ , and non-negative for  $a + 1 \leq r \leq b$ , being strictly positive for one of these values of  $r$  at least, where  $a$  and  $b$  are integers satisfying  $0 \leq a < b$ .

Moreover,  $\sum_{r=0}^{\infty} \delta P_r = \sum_{r=0}^{\infty} r \delta P_r = 0$ . It follows easily that if  $f_r$  is convex from below  $\sum_{r=0}^{\infty} f_r \delta P_r < 0$ , while if  $f_r$  is concave from below the inequality is reversed. Thus in either case the right-hand side of (3.7) is monotone, as stated; and therefore the equation has at most one positive solution  $\hat{k}$ . On the other hand, in repeated samples from the same negative binomial distribution, the probability is small that (3.7) has no positive solution, when  $N$  is large. For the left-hand side has a high probability of being close to  $F(m, k)$ , and  $F(\bar{r}, x)$  is a continuous function of  $\bar{r}$  and  $x$  with a range of values, as  $x$  varies in  $(0, \infty)$ , differing little from the range of values of  $F(m, x)$  if  $\bar{r}$  is near to  $m$ .

We can find the large-sample variance of the estimate  $\hat{k}$  given by (3.7) by treating  $\hat{k} - k$  and  $\bar{r} - m$  as infinitesimals. Denoting the latter by  $\delta k$  and  $\delta m$  respectively, and  $n_r/N - P_r$  by  $\delta P_r$ , we have

$$\begin{aligned} \delta m &= \sum_{r=0}^{\infty} r \delta P_r, \\ A_k \delta k &= \sum_{r=0}^{\infty} (f_r - A_m r) \delta P_r, \end{aligned}$$

where 
$$A_m = \frac{\partial}{\partial m} F(m, k), \quad A_k = \frac{\partial}{\partial k} F(m, k).$$

Now 
$$\text{var}(\delta P_i) = P_i(1 - P_i)/N, \quad \text{cov}(\delta P_i, \delta P_j) = -P_i P_j / N \quad (i, j \geq 0, i \neq j),$$

while from (1.1) we find 
$$\left(m + \frac{m^2}{k}\right) A_m = E(r f_r) - m E(f_r).$$

Hence we obtain the desired results:

$$\text{cov}(\bar{r}, \hat{k}) \sim 0, \tag{3.8}$$

$$\text{var}(\hat{k}) \sim \frac{E(f_r^2) - [E(f_r)]^2 - (m + m^2/k) A_m^2}{N A_k^2}. \tag{3.9}$$

The ratio of the variances (3.5) and (3.9) is the large-sample efficiency of method (3.7) of estimating  $k$ . As already remarked,  $m$  is easily estimated with full efficiency by equation (3.1). Since these estimates are in large samples independent, it is appropriate to consider their efficiencies separately. In general, when considering inefficient estimates of the parameters of a distribution, a reasonable single measure of large-sample efficiency is provided by the square root of the ratio of determinants of sampling variance matrices for the maximum-likelihood estimates and for the alternative estimates under consideration, if there are two parameters to be estimated, or the cube root if there are three parameters, etc. This

measure is the ratio of the numbers of observations required by the maximum-likelihood method and the alternative method to achieve the same error variance determinant. In the present case it would be equal to the square root of the efficiency of estimation of  $k$ .

Let us now consider some examples.

*Method 1.*  $f_r = r^2$ , so that  $k$  is estimated from the sample variance  $s^2$ , i.e.

$$\hat{k} = \frac{\bar{r}^2}{s^2 - \bar{r}}. \tag{3.10}$$

On evaluating  $E(f_r)$ ,  $E(f_r^2)$ ,  $A_k$ ,  $A_m$ , we find the large-sample variance

$$\text{var}(\hat{k}) \sim \frac{2k(k+1)}{NX^2}; \tag{3.11}$$

and the efficiency of estimation of  $k$  is equal to the reciprocal of the expression (3.6).

*Method 2.*  $f_0 = 1, f_r = 0$  for  $r \geq 1$ , so that  $k$  is estimated from the observed proportion of zeros, i.e.

$$n_0/N = (1 + \bar{r}/\hat{k})^{-\hat{k}}. \tag{3.12}$$

The large-sample variance is

$$\text{var}(\hat{k}) \sim \frac{(1-X)^{-k} - 1 - kX}{N[-\ln(1-X) - X]^2}. \tag{3.13}$$

*Method 3.*  $f_r = 1/(r+1)$ .  $\hat{k}$  is given by

$$\frac{1}{N} \sum_{r=0}^{\infty} \frac{n_r}{r+1} = \frac{(1-X) - (1-X)^{\hat{k}}}{(\hat{k}-1)X}, \quad \text{where } X = \frac{\bar{r}}{\bar{r} + \hat{k}}. \tag{3.14}$$

The large-sample variance of  $\hat{k}$  is given by (3.9), where

$$\left. \begin{aligned} E(f_r) &= \frac{(1-X) - (1-X)^k}{(k-1)X}, & E(f_r^2) &= (1-X)^k \sum_{r=0}^{\infty} \frac{k(k+1) \dots (k+r-1)}{(r+1)(r+1)!} X^r, \\ A_m &= \frac{1-X}{k(k-1)X^2} [1 + (k-1)X] (1-X)^k - (1-X), \\ A_k &= \frac{1}{k(k-1)^2 X} [1 - k(k-1)\ln(1-X) - (k-1)^2 X] (1-X)^k - (1-X). \end{aligned} \right\} \tag{3.15}$$

This is for  $k \neq 1$ . Corresponding expressions when  $k = 1$  are easily found directly.

*Method 4.*  $f_r = c^r$ , where  $c$  is a positive constant not equal to 1. We find

$$\frac{1}{N} \sum_{r=0}^{\infty} n_r c^r = [1 + (1-c)\bar{r}/\hat{k}]^{-\hat{k}}; \tag{3.16}$$

and for the large-sample variance

$$\left. \begin{aligned} E(f_r) &= [1 + p(1-c)]^{-k}, & E(f_r^2) &= [1 + p(1-c^2)]^{-k}, \\ A_m &= -(1-c)[1 + p(1-c)]^{-k-1}, \\ A_k &= [1 + p(1-c)]^{-k} \left\{ -\ln[1 + p(1-c)] + \frac{p(1-c)}{1 + p(1-c)} \right\}. \end{aligned} \right\} \tag{3.17}$$

The above seem to exhaust the tractable applications of the moment method. Several other forms for the function  $f_r$  suggest themselves as possibly worth investigating, such as (i)  $f_0 = 0, f_r = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{r}$  for  $r \geq 1$ , (ii)  $f_r = \ln(r+1)$  for  $r \geq 0$ , and (iii)  $f_r = \sqrt{r}$  for  $r \geq 0$ ; but unfortunately these do not lead to a simple expression for  $E(f_r)$ . Their practical use is therefore ruled out unless special tables are constructed for estimating  $k$  from  $\sum n_r f_r / N$ .

A method of a different kind is the following.

*Method 5.* We guess a value of  $k$ , apply a transformation (depending on  $k$ ) to the observations to make the variance independent of the mean, and then obtain an improved estimate of  $k$  by equating the observed variance to the expected; the process is repeated until the answer becomes stable. Suitable transformations were investigated by me (1948) and summarized by me (1949). From a consideration of equation (4.37) of the former paper, which gives an asymptotic expansion for the variance of the transformed variable when  $k = 1$ , it appears that the transformation method is unlikely to be serviceable for  $k < 1$ . For higher values of  $k$ , the method is possible for not-too-low values of  $m$ . In principle the method could be used for any values of  $m$  and  $k$ , if the expected variance of the transformed observations were known as a function of  $m$  and  $k$  in a convenient form, e.g. by an adequate double-entry table. But such a table is not available, and so the method is restricted to those values of  $m$  for which the limiting variance as  $m \rightarrow \infty$  is near enough attained.

The large-sample variance of the estimate of  $k$  derived by the transformation method will now be found. Let  $y$  denote the transformed variable, a function of the observed count  $r$ , when the true value of  $k$  is inserted in the formula for  $y$ ; and let  $\hat{y}$  be the transformed variable when an estimate  $\hat{k}$  of  $k$  is used. Then in samples from a fixed negative binomial distribution with parameters  $m$  and  $k$ , we have, from equation (4.23) of my 1948 paper (setting  $A = \frac{1}{2}\hat{k}$ ),

$$\text{var}(\hat{y}) = \psi'(k) + \frac{(k - \hat{k})k}{(k - 1)^2 m} + O\left(\frac{1}{m^2}\right), \quad (3.18)$$

as  $m \rightarrow \infty$ , if  $k \geq 2$ .\* Let  $s^2$  be the variance estimate calculated from the observations after transformation to  $y$ -variables, and  $\hat{s}^2$  that calculated after the observations have been transformed to  $\hat{y}$ -variables. We choose  $\hat{k}$ , by successive approximation, so that

$$\hat{s}^2 = \psi'(\hat{k}). \quad (3.19)$$

Let  $\partial s^2 / \partial k$  denote the derivative of  $s^2$  (for the given sample of observations) with respect to the parameter  $k$  entering into the  $y$ -transformation, when  $k$  is set equal to its true value. For any arbitrary  $\hat{k}$ , not necessarily close to  $k$ , we have in probability as  $N \rightarrow \infty$

$$\hat{s}^2 = \text{var}(\hat{y}) + O(1/N),$$

and therefore

$$\frac{\partial \hat{s}^2}{\partial \hat{k}} = \left[ \frac{\partial}{\partial \hat{k}} \text{var}(\hat{y}) \right]_{\hat{k}=k} + O\left(\frac{1}{N}\right).$$

Hence, if  $\hat{k}$  is determined by (3.19) and we set  $\hat{k} - k = \delta k$  and  $\hat{s}^2 - \text{var}(y) = \delta \sigma^2$ , we have in probability for large  $N$  and  $m$

$$\psi'(k) + \psi''(k) \delta k = \psi'(\hat{k}) = \hat{s}^2 = s^2 + \frac{\partial s^2}{\partial k} \delta k = \psi'(k) + \delta \sigma^2 - \frac{k \delta k}{(k - 1)^2 m},$$

and therefore

$$\delta \sigma^2 = \left\{ \psi''(k) + \frac{k}{(k - 1)^2 m} \right\} \delta k.$$

The large-sample variance of the left-hand side is found from equations (4.23) and (4.30) of my 1948 paper, and we obtain the result

$$\text{var}(\hat{k}) \sim \frac{\psi'''(k) + 2[\psi''(k)]^2}{N \left[ \psi''(k) + \frac{k}{(k - 1)^2 m} \right]^2}. \quad (3.20)$$

\*  $\psi(t), \psi'(t), \dots$  denote the successive derivatives of  $\ln \Gamma(t)$ .

This should be sufficiently accurate for practical purposes if  $m$  is above 50 and  $k$  above 5, assuming that the appropriate inverse hyperbolic sine transformation is used. It is not clear from existing calculations how far (3.20) may be relied on outside this region.

We thus have an assortment of alternatives to the maximum-likelihood method of estimating  $k$ . Let  $E_1, E_2, E_3, E_4, E_5$  denote the large-sample efficiencies of the five methods. By tabulating these efficiencies we can see under what conditions, if any, each method is likely to prove useful. In the figure are shown 50, 75, 90 and 98 % contours of  $E_1$  and  $E_2$ , and 90 and 98 % contours of  $E_5$ . The 75 % contour of  $E_5$  has been found only for  $m$  very large, when it is close to the line  $k = 1$ ; it is not shown in the figure. Since Method 5 can hardly be used when  $k < 1$ , no attempt has been made to determine the 50 % contour.

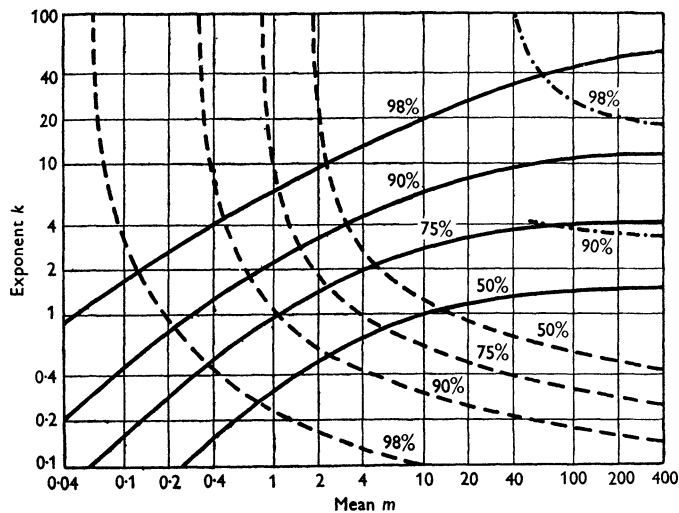


Fig. 1. Large-sample efficiencies of estimation of  $k$ .

Method 1 ——— Method 2 - - - - Method 5 - . - . - .

In considering the lower right-hand region of the figure, where  $m$  is large and  $k$  is small, it is helpful to note\* that if we set  $a = -k \ln(1 - X)$  and let  $k \rightarrow 0, X \rightarrow 1$ , with  $a$  constant, the expression (3.6) is asymptotically equal to  $2(1 - e^{-a})/k$ , whence the limiting efficiencies of Methods 2, 3 and 4 are all equal to  $a^2 e^a / (e^a - 1)^2$ . Other limiting forms of the efficiencies, for movement away from the centre of the figure in various directions, are easily found and need not be quoted.

No contours are shown in the figure for Methods 3 and 4, since, as it turns out, these are nowhere more efficient than the more efficient of Methods 1 and 2, and the latter are more convenient to use. Method 4 becomes equivalent to Method 2 as  $c \rightarrow 0$ , and to Method 1 as  $c \rightarrow 1$ . If we imagine  $c$  increasing continuously from 0 to 1, a constant-value contour of  $E_4$ , such as the 90 % contour, departs from the corresponding contour of  $E_2$  by a translation of the uppermost part of the curve to the right ('east') and a pulling of the middle part of the curve downwards to the left ('south-west'). (The first of these movements is easily expressed. The contour has a vertical asymptote for  $k$  large, of the form  $m = \text{constant}$ . This value of  $m$  is equal to the value for the corresponding contour of  $E_2$  multiplied by  $(1 - c)^{-2}$ .) As  $c$  increases, the contour eventually breaks up into two disjoint parts, and two new parallel

\* Proofs of this and other statements in the remainder of this section are omitted, to save space.

asymptotes appear, of the form  $m/k = \text{constant}$ . As  $c \rightarrow 1$ , the upper part of the curve approaches the  $E_1$  contour, while the lower part recedes and vanishes in the limit. The  $E_3$  contours are similar in character to  $E_4$  contours.

#### 4. TESTS FOR DEPARTURE FROM THE NEGATIVE BINOMIAL FORM OF DISTRIBUTION

We have considered how a negative binomial distribution can be fitted to a large sample. We turn now to testing goodness of fit, in particular to detecting a departure from the negative binomial form towards one of the other forms discussed in § 2. Tests are required that are reasonably convenient to use.

Particular interest attaches to discriminating between the negative binomial and the Pólya-Aeppli forms of distribution. Let us suppose to begin with that  $m/k (=p)$  is small. Then the log-likelihood function of the observations on the hypothesis of a negative binomial distribution is

$$L_0 \equiv \sum_{r=0}^{\infty} n_r \ln P_r = \sum_{r=0}^{\infty} n_r r \ln m - Nm - \sum_{r=2}^{\infty} n_r \ln (r!) + \sum_{r=0}^{\infty} n_r \left[ \frac{(r-m)^2 - r}{m} \right] \frac{p}{2} - \sum_{r=0}^{\infty} n_r \left[ \frac{r(r-1)(2r-1)}{2m^2} - 3r + 2m \right] \frac{p^2}{6} + O(p^3), \quad (4.1)$$

as  $p \rightarrow 0$  with  $m$  fixed. The maximum-likelihood equations for estimating  $k$  and  $p$  are asymptotically those of Method 1 of the last section. Consider now the hypothesis that the observations are drawn from a Pólya-Aeppli distribution. In terms of parameters  $k$  and  $p$  defined as in § 2, the log-likelihood function of the observations,  $L_1$  say, is the same as the above expression for  $L_0$  except that the term in  $p^2$  is

$$- \sum_{r=0}^{\infty} n_r \left[ \frac{r(r-1)^2}{m^2} - \frac{r(r-1)}{m} - r + m \right] \frac{p^2}{4}. \quad (4.2)$$

Again the maximum-likelihood method of estimating  $k$  and  $p$  is asymptotically equivalent to fitting by the first two sample moments. The likelihood ratio criterion for discriminating between the two distributions is found by maximizing  $L_0$  and  $L_1$  separately with respect to  $k$  and  $p$ , and then subtracting them. Its leading term involves the first three sample moments. We are thus led to propose the following test for departure from the negative binomial form towards the Pólya-Aeppli form, to be used when  $p$  is small:

*Test 1.* Estimate the parameters of the negative binomial distribution from the first two sample moments (Method 1), and then compare the third sample moment with the estimated third moment of the distribution.

There is no point in actually evaluating the likelihood ratio criterion just described, since we do not know *a priori* that the parent distribution is necessarily of one or other of the two forms considered. We may, however, apply a test similar to Test 1 to see whether or not the observations are in agreement with the Pólya-Aeppli or any other hypothetical form of distribution.

We can similarly investigate the likelihood ratio criterion in another simple limiting case, namely, for  $P_0 \rightarrow 1$  with  $m$  constant. We find, for both distributions, that the maximum-likelihood estimation of the parameters involves asymptotically only the two statistics  $\bar{r}$  and  $n_0/N$  (Method 2 above). The leading term of the likelihood ratio criterion involves, in addition to these, a further statistic,  $\sum_{r=2}^{\infty} n_r \ln r$ . This is not a convenient statistic on which



to base a test, since its expected value cannot be expressed simply. The only simple test that suggests itself is, in fact, one based on the sample variance, thus:

*Test 2.* Estimate the parameters of the negative binomial distribution from the sample mean and the observed proportion of zeros (Method 2), and then compare the sample variance with the estimated variance of the distribution.

Test 2 arises more directly in another problem of discrimination. We suppose that the parent distribution of the observations is a compound (or heterogeneous) negative binomial, i.e. that each observation is drawn from a negative binomial distribution having the same exponent  $k$  but a mean that varies randomly from observation to observation in a distribution with mean  $m$  and variance  $\epsilon$ , say. The resulting distribution departs from the negative binomial form towards that of the discrete lognormal, having high skewness and kurtosis coefficients. We now test the hypothesis that  $\epsilon = 0$ . If the log-likelihood function of the observations is expanded in ascending powers of  $\epsilon$ , the coefficient of  $\epsilon$  is a function of the first two sample moments (and of the unknown parameters  $m$  and  $k$ ). The optimum large-sample test of the hypothesis thus consists of fitting  $k$  and  $m$  by maximum likelihood, assuming that  $\epsilon = 0$ , and then comparing the observed and estimated variances. In the region where Method 2 of fitting  $k$  is efficient, we would use Test 2 above. In the region where Method 1 is efficient, i.e. for  $k$  large, a small heterogeneity in the mean of the negative binomial has the effect of apparently reducing  $k$  but not otherwise changing the shape of the distribution, so that the heterogeneity is not easily detectable. Test 1 could be used to detect a pronounced degree of heterogeneity in the mean.

It remains to consider how Tests 1 and 2 are carried out. We shall suppose the sample size  $N$  to be large. The criterion of Test 1 is the difference between the third sample moment and its estimated value, i.e.

$$T = \frac{1}{N} \sum_{r=0}^{\infty} n_r (r - \bar{r})^3 - s^2 \left( \frac{2s^2}{\bar{r}} - 1 \right). \tag{4.3}$$

Using the known formulae (quoted by Kendall, 1943) for the variances and covariances of  $k$ -statistics, or, alternatively, working in terms of the sample factorial moments (of which the variances and covariances are easier to find directly), we obtain the large-sample result

$$N \text{ var } (T) \sim 2k(k+1)p^3(1+p)^2 [2(3+5p) + 3k(1+p)]. \tag{4.4}$$

The criterion of Test 2 is the difference between the sample variance and its estimated value, i.e.

$$U = s^2 - \bar{r} - \hat{r}^2 / \hat{k}, \tag{4.5}$$

where  $\hat{k}$  is the estimate of  $k$  by Method 2. We find, for  $N$  large,

$$N \text{ cov } (\hat{k}, \bar{r}) \sim 0, \quad N \text{ cov } (\hat{k}, s^2) \sim -k(k+1)p^2 / \{-\ln(1-X) - X\}, \tag{4.6}$$

from which and (3.13) we obtain

$$N \text{ var } (U) \sim 2k(k+1)p^2(1+p)^2 \left[ 1 - \frac{X^2}{-\ln(1-X) - X} \right] + p^4 \left[ \frac{(1-X)^{-k} - 1 - kX}{(-\ln(1-X) - X)^2} \right]. \tag{4.7}$$

As an example, we may consider a frequency distribution quoted by Williams (1944, Table 6) of the number of head-lice of all stages found in the hair of Hindu male prisoners on admission to Cannanore jail, South India, over a period from 1937 to 1939 (see Buxton, 1940).\*

\* Prof. Buxton has very kindly allowed me to check the frequency distribution against the original records. There seems to be one error in Williams's table, the number of heads without lice being given as 622 instead of 612. Williams has also made a numerical slip in fitting a negative binomial distribution to the observations, so that the fit appears worse than it should.

We have  $N = 1073$ ,  $n_0 = 612$ ,  $\bar{r} = 6.93569$ ,  $s^2 = 583.8$ . By Method 2,  $\hat{k} = 0.144198$ , and hence  $U = 243.3$ , with estimated standard error 38.9. We conclude that the observations are not adequately fitted by a negative binomial distribution, being too skew. The result is not surprising, in view of the heterogeneity of the prisoners in caste and in other respects likely to affect personal hygiene.

5. FITTING A COMMON EXPONENT TO A SERIES OF SAMPLES

We suppose now that we have a series of samples, one from each of  $\nu$  negative binomial distributions. Characteristics of the  $i$ th distribution and the sample from it will be denoted by the usual symbol with suffix  $i$  added. We may be interested in investigating how  $k_i$  is related to  $m_i$ . If the  $k_i$  are not too small we may estimate them by Method 1 and plot them against  $\bar{r}_i$ . We shall obtain a less skew distribution of errors if the reciprocal of the estimate of  $k_i$ , rather than the estimate itself, is considered, i.e.  $(s_i^2 - \bar{r}_i)/\bar{r}_i^2$ . In either case, there is a bias of order  $N_i^{-1}$  which may be worth removing if  $\nu$  is large. The following estimate of  $k_i^{-1}$ ,

$$\frac{1}{\hat{k}_i} = \frac{s_i^2 - \bar{r}_i}{\bar{r}_i^2} \left( 1 + \frac{s_i^2}{N_i \bar{r}_i^2} \right), \tag{5.1}$$

is approximately unbiased, having expected value  $k_i^{-1} + O(N_i^{-2})$ . This is easily proved (dropping now the suffix  $i$ , for convenience) by writing  $\bar{r} = kp + \delta m$ ,  $s^2 = kp(1+p) + \delta\sigma^2$ , and expanding (5.1) in ascending powers of  $\delta m$  and  $\delta\sigma^2$ . To the first order when  $N$  is large, the variance of this estimate of  $k^{-1}$  is equal to the right-hand side of (3.11) divided by  $k^4$ , and its correlation with  $\bar{r}$  is zero, in samples from the same distribution.

Sometimes it is reasonable to suppose that all the  $k_i$  are equal. We then desire an efficient estimate of the common value. A method that suggests itself is to choose a value of  $\hat{k}$  such that the sum (or a weighted sum) of  $s_i^2 - \bar{r}_i - \bar{r}_i^2/\hat{k}$  vanishes. The expected value of this expression, when  $\hat{k}$  is replaced by the true  $k$ , is  $O(N_i^{-1})$ , and it can be reduced to  $O(N_i^{-2})$  if we use instead

$$t_i(\hat{k}) = s_i^2 - (\bar{r}_i + \bar{r}_i^2/\hat{k})(1 - [N_i \hat{k}]^{-1}). \tag{5.2}$$

If weights  $w_i$  are used, our equation for  $\hat{k}$  becomes

$$\sum_{i=1}^{\nu} w_i t_i(\hat{k}) = 0. \tag{5.3}$$

Treating  $\hat{k} - k$  as infinitesimal and working to the first order for  $N_i$  large, we find easily

$$\text{var}(\hat{k}) \sim \frac{\sum w_i^2 \text{var}(t_i)}{(\sum w_i \partial t_i / \partial k)^2}, \tag{5.4}$$

$$N \text{var}(t_i) \sim 2k(k+1)p_i^2(1+p_i)^2, \tag{5.5}$$

$$\partial t_i / \partial k \sim p_i^2. \tag{5.6}$$

$\text{var}(\hat{k})$  is a minimum when the weights  $w_i$  are taken to be proportional to

$$\frac{1}{\text{var}(t_i)} \frac{\partial t_i}{\partial k}, \quad \text{i.e. to} \quad \frac{N_i}{(1+p_i)^2}.$$

We therefore choose

$$w_i = \frac{N_i - 1}{(\bar{r}_i + \hat{k})^2}. \tag{5.7}$$

The numerator  $N_i - 1$  is more convenient than  $N_i$ , since  $(N_i - 1) s_i^2$  is found in the course of calculating  $s_i^2$ . Thus the method is to choose  $\hat{k}$  to satisfy  $\Sigma T_i(\hat{k}) = 0$ , where

$$T_i(\hat{k}) = \frac{(N_i - 1) s_i^2 - (N_i - 1 - \hat{k}^{-1}) (\bar{r}_i + \bar{r}_i^2/\hat{k})}{(\bar{r}_i + \hat{k})^2} \tag{5.8}$$

It is easy to verify that  $T_i(k)$  has expected value  $O(N_i^{-1})$ . The method is equivalent to taking an appropriately weighted average of the estimates (5.1); and, to the first order for  $N_i$  large, the variance of  $\hat{k}$  is the reciprocal of the sum of reciprocals of the variances (3.11) for each sample. In samples from a single population, the correlation between  $T(k)$  and  $\bar{r}$  is  $O(N^{-1})$ .

We can deal similarly with estimates of  $k$  based on Method 2. Corresponding to (5.2) we may consider

$$u_i(\hat{k}) = n_{0i} - \left(1 + \frac{\bar{r}_i}{\hat{k}}\right)^{-\hat{k}} \left\{N_i - \frac{\bar{r}_i(\hat{k} + 1)}{2(\bar{r}_i + \hat{k})}\right\}, \tag{5.9}$$

the expected value of which is  $O(N_i^{-1})$ . The optimum weight factor  $w_i$  (for  $N_i$  large) is

$$w_i = \frac{-\ln(1 - X_i) - X_i}{(1 - X_i)^k [(1 - X_i)^{-k} - 1 - kX_i]}, \tag{5.10}$$

or rather an estimate of this. Since there is little to be gained by exactitude in weight factors, we may prefer to take instead

$$w_i = \ln(1 + \bar{r}_i/\hat{k}), \tag{5.11}$$

which has roughly the same effect and is easier to calculate. If (5.10) is used, we find that, to the first order for  $N_i$  large, the variance of  $\hat{k}$  is the reciprocal of the sum of reciprocals of the variances (3.13) for each sample. To test whether  $k$  changes progressively with  $m$ , we may plot  $U_i = w_i u_i$  against  $\bar{r}_i$  and look for a correlation. In samples from a single population, the correlation between  $U$  and  $\bar{r}$  is  $O(N^{-1})$ .\*

### 6. FITTING THE LOGARITHMIC SERIES DISTRIBUTION

Fisher's logarithmic series distribution was proposed as a model for the relative abundance of different species found in trap catches or other methods of sampling; in particular, to describe the relation between the numbers of moths of different species caught in a light-trap over a period of time. Suppose there are  $N$  species that might be observed, and that their abundances (numbers expected to be caught in a unit period of time) are distributed as if they were a sample of size  $N$  from a Type III distribution, proportional to a  $\chi^2$  with  $2k$  degrees of freedom. We suppose further that the individuals of each species move independently, so that the number of individuals of any species caught has a Poisson distribution with mean value equal to the abundance multiplied by the length of time of observation. The numbers of individuals caught per species then form a random sample of size  $N$  from the negative binomial distribution (1.1). The results of the observation can be expressed by the numbers  $n_i$  of species represented by  $i$  individuals for all  $i \geq 0$ . We define  $S$  and  $I$  as in § 1, so that  $S$  is the number of species represented by at least one individual, and  $I$  is the total

\* In my 1949 paper there is an oversight on this point. Having noticed a correlation between  $U_i$  and  $\bar{r}_i$  in some data, I added (p. 172): 'The effect is too marked to be attributed to the negative correlation between  $n_0$  and  $\bar{r}$  that occurs in repeated sampling of the same population.' This is literally true, but misleading, since the relevant correlation is that between  $U$  and  $\bar{r}$ , not  $n_0$  and  $\bar{r}$ , and that is  $O(N^{-1})$ .

number of individuals of all species observed. The probability distribution of  $n_i$  for  $i \geq 1$ , given  $N$ ,  $k$  and  $X$ , is

$$\frac{N!}{(N-S)! \prod_{i=1}^{\infty} n_i!} (1-X)^{kN} k^{n_1} \left(\frac{k(k+1)}{2!}\right)^{n_2} \left(\frac{k(k+1)(k+2)}{3!}\right)^{n_3} \dots X^I. \tag{6.1}$$

If we set  $Nk = \alpha$ , and consider the limit  $k \rightarrow 0$ ,  $N \rightarrow \infty$ , with  $\alpha$  constant, we find easily that the above breaks up into a product of Poisson frequency functions for  $n_i$  ( $i \geq 1$ ), as indicated at (1.8). If the time of exposure or attractive power of the trap were multiplied by a factor  $c$ , without the abundances of the species being affected, it follows from the above assumptions that  $p$  would be changed to  $cp$ , and therefore  $X$  to  $cp/(cp+1)$ , while  $\alpha$  would be unaltered.  $\alpha$  is thus a property of the biological association that is being examined, and has been termed by C. B. Williams the ‘index of diversity’ of the association.

The log-likelihood function of the observations is

$$L = \alpha \ln(1-X) + S \ln \alpha + I \ln X - \sum_{r=1}^{\infty} \{n_r \ln r + \ln n_r!\}. \tag{6.2}$$

Thus  $S$  and  $I$  are jointly sufficient for estimating  $\alpha$  and  $X$ , and the maximum-likelihood equations are

$$\left. \begin{aligned} I &= \hat{\alpha} \hat{X} / (1 - \hat{X}) = \hat{\alpha} \hat{p}, \\ S &= -\hat{\alpha} \ln(1 - \hat{X}) = \hat{\alpha} \ln(1 + \hat{p}). \end{aligned} \right\} \tag{6.3}$$

On inverting the matrix of expectations of second derivatives of  $L$ , to find the variances of these estimates in the usual way, we get

$$\left. \begin{aligned} [-\ln(1-X) - X] \text{var}(\hat{X}) &\sim -\frac{1}{\alpha} X(1-X)^2 \ln(1-X), \\ [-\ln(1-X) - X] \text{cov}(\hat{X}, \hat{\alpha}) &\sim -X(1-X), \\ [-\ln(1-X) - X] \text{var}(\hat{\alpha}) &\sim \alpha. \end{aligned} \right\} \tag{6.4}$$

These formulae would certainly be correct asymptotically if the right-hand sides were divided by  $\nu$  and we were considering pooled estimates from  $\nu$  completely independent samples, with  $\nu$  tending to infinity. But we are actually concerned with one sample. Let us see in what sense, if any, (6.4) can still be regarded as correct. If  $\alpha$  and  $p$  are both large, differentiation of any derivative of  $L$  by  $p$  changes its order of magnitude by a factor  $1/p$ , and differentiation by  $\alpha$  changes its order by a factor  $1/\alpha$ . The second derivatives are effectively constant if in probability

$$\frac{\delta\alpha}{\alpha} + \frac{\delta p}{p} = o(1),$$

where  $\delta\alpha = \hat{\alpha} - \alpha$ ,  $\delta p = \hat{p} - p$ . Assuming (6.4) to be true,

$$\delta\alpha = O\left(\sqrt{\frac{\alpha}{\ln p}}\right), \quad \delta p = O\left(\frac{p}{\sqrt{\alpha}}\right),$$

in probability. Hence formulae (6.4) are correct asymptotically if  $\alpha \rightarrow \infty$  while  $p$  is constant or increases (or, more generally, has a positive lower bound).

But this result is not entirely satisfactory, since  $\alpha$  is a constant of the association being observed, and although in many of the examples cited by Fisher *et al.* (1943) and Williams

(1944)  $\alpha$  is fairly large, there is no logical necessity for it to be so.\* The only adjustable feature of the sample is the time of exposure (or the attractive power) of the trap, by increasing which  $p$  may be increased. It is therefore of interest to develop asymptotic formulae valid as  $p \rightarrow \infty$  with  $\alpha$  constant. To the first order when  $p$  is large, the variance of  $\hat{\alpha}$  given at (6.4) is

$$\text{var}(\hat{\alpha}) \sim \frac{\alpha}{\ln p}, \tag{6.5}$$

while to the next order of approximation

$$\text{var}(\hat{\alpha}) \sim \frac{\alpha}{\ln p - 1}. \tag{6.6}$$

Let us see whether in fact one or both of these is correct, as  $p \rightarrow \infty$ .

From (6.3),  $\hat{\alpha}$  is determined by the equation

$$S = \hat{\alpha}\{\ln(I + \hat{\alpha}) - \ln \hat{\alpha}\}. \tag{6.7}$$

The distribution of  $S$  is Poisson, with mean  $\alpha \ln(1 + p)$ ; the distribution of  $I$  is negative binomial with mean  $\alpha p$  and exponent  $\alpha$ . Thus while the former approaches normality as  $p \rightarrow \infty$ , and has coefficient of variation tending to zero, the latter does not. The distribution of  $\ln(I + \alpha)$  also does not approach normality as  $p \rightarrow \infty$ , but the asymptotic distribution is known (Anscombe, 1948), and we have

$$E\{\ln(I + \alpha)\} = \ln p + \psi(\alpha) + O(p^{-\beta}), \quad \text{var}\{\ln(I + \alpha)\} = \psi'(\alpha) + O(p^{-\beta}), \tag{6.8}$$

where  $\beta$  is any quantity such that  $0 < \beta \leq 1$ ,  $\beta < \alpha$ . We write now  $\delta S$ ,  $\delta I$ ,  $\delta \ln(I + \alpha)$  for the differences between  $S$ ,  $I$ ,  $\ln(I + \alpha)$ , and their respective mean values. Then (6.7) can easily be shown to give, in probability as  $p \rightarrow \infty$ ,

$$\delta \alpha = \frac{\delta S + \alpha\{\ln \alpha - \psi(\alpha) - \delta \ln(I + \alpha)\}}{\ln p - \{1 + \ln \alpha - \psi(\alpha) - \delta \ln(I + \alpha)\} + o(1)}. \tag{6.9}$$

Considering the first terms of numerator and denominator, we see at once that the distribution of  $\hat{\alpha}$  is asymptotically normal and that (6.5) is correct. To see whether (6.6) is also correct, we need to evaluate  $E\{\delta S \delta \ln(I + \alpha)\}$  and  $E\{(\delta S)^2 \delta \ln(I + \alpha)\}$ . The joint distribution of  $S$  and  $I$  has probability-generating function

$$E(t^S u^I) = \exp\{\alpha(t - 1) \ln(1 + p) - \alpha t \ln(1 + p - pu)\}. \tag{6.10}$$

Writing  $x = \{S - \alpha \ln(1 + p)\} \{\alpha \ln(1 + p)\}^{-1}$  and  $y = Ip^{-1}$ ,

we find, on expanding the characteristic function of  $x$  and  $y$  for  $p$  large and applying the Fourier inversion formula, the following asymptotic continuous distribution of  $x$  and  $y$ :

$$\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} \left[ 1 + \frac{1}{\sqrt{(\alpha \ln p)}} \left\{ \frac{1}{6}x(x^2 - 3) + \alpha x(\ln y - \psi(\alpha)) \right\} + O\left(\frac{1}{\ln p}\right) \right] dx dy. \tag{6.11}$$

\* The fact that  $\alpha$  is not adjustable does not in itself bar the use of an asymptotic formula valid as  $\alpha \rightarrow \infty$ , since in any case such asymptotic formulae are used as approximations. Even if the parameter concerned is adjustable, only one value is usually available for consideration, and not an infinite sequence of values. An example of confusion on this point is the criticism by Kendall (1948) of a limit situation considered by Jones (1948) in the theory of systematic sampling. Jones gives a formula for the error variance which is asymptotically correct as the population extent tends to infinity, with constant spacing between sample points. Kendall, remarking that the population extent is not in general adjustable, gives a formula asymptotically correct as the spacing between sample points tends to zero, with the population extent constant. As an approximation to the actual situation, Jones's formula is the better (see Kendall, 1948, equations (20) and (25)).

$$\begin{aligned} \text{Hence easily} \quad & \left. \begin{aligned} \mathbf{E}\{\delta S \delta \ln(I + \alpha)\} &= \alpha \psi'(\alpha) + O\{(\ln p)^{-\frac{1}{2}}\}, \\ \mathbf{E}\{(\delta S)^2 \delta \ln(I + \alpha)\} &= O(1). \end{aligned} \right\} \end{aligned} \quad (6.12)$$

$$\text{Finally, from (6.9),} \quad \mathbf{E}(\hat{\alpha}) = \alpha \left\{ 1 + \frac{\ln \alpha - \psi(\alpha) + o(1)}{\ln p} \right\}, \quad (6.13)$$

$$\text{var}(\hat{\alpha}) = \frac{\alpha}{\ln p + \alpha \psi'(\alpha) - 2(1 + \ln \alpha - \psi(\alpha)) + o(1)}. \quad (6.14)$$

On inserting the asymptotic expansions of  $\psi(\alpha)$  and  $\psi'(\alpha)$  in powers of  $\alpha^{-1}$ , we obtain the right-hand side of (6.6) with remainder term in the denominator which is  $o(1)$  for both  $p$  and  $\alpha$  large. Thus (6.6) is not correct, to the order suggested, if  $p$  is large, unless  $\alpha$  is also large.

The formulae for  $\text{var}(\hat{\alpha})$  that we have just discussed differ from one another only in accuracy; they are all approximations to the same true value, which has not been found. An essentially different formula has been given by Fisher (Fisher *et al.* 1943). When expressed in a form similar to (6.6), it is

$$\text{var}(\hat{\alpha}) \sim \frac{\alpha \ln 2}{(\ln p - 1)^2}. \quad (6.15)$$

This formula is appropriate to a special type of comparison, namely, between estimates of  $\alpha$  for the same biological association derived from similar nearby traps, where it may be assumed that the individual species have exactly the same abundances (or at least the same relative abundances), and the difference between the catches at any two traps arises solely from Poisson variation in the numbers caught of each species. In such a sampling process, the estimate of  $\alpha$  given by (6.7) is substantially biased, since only one set of relative abundances is involved. If we consider the variation in the bias for all possible sets of relative abundances following the limiting Type III distribution assumed, the overall variance of  $\hat{\alpha}$  is increased from Fisher's value to that already considered. The larger variance is appropriate to comparing the values of  $\hat{\alpha}$  from observations on different sorts of biological association, involving perhaps entirely different families of species, and also to comparing values of  $\hat{\alpha}$  from observations on the same sort of biological association observed at different seasons of the year or in different years, when, even if the families of species are the same, the relative abundances of the species are different and may be supposed in aggregate to constitute independent samples from the limiting Type III distribution. Fisher's formula, in fact, is not likely to be often useful, since if we desire to test whether the abundances of the individual species are the same (or in proportion) at a number of traps it will be correct in the first place to make direct comparisons of counts of individual species, by a  $\chi^2$  contingency-table test, or by analysis of variance after making a square-root transformation. If it has been established that the relative abundances of the species are not the same at different traps, it is likely that they will differ sufficiently to appear in aggregate to be independent samples from the hypothetical parent Type III distribution. They may indeed be so different as to suggest quite different parent distributions, and a test of this point would be based on the total variance of  $\hat{\alpha}$ .

For example, Williams gives figures for captures of Noctuidae during a period of three months in 1933 at two traps (Fisher *et al.* 1943). One trap, on a roof-top, gave  $S = 58$ ,  $I = 1856$ ,  $\hat{\alpha} = 11.37$ ,  $\hat{p} = 163$ ; the other, in a field a quarter of a mile away, gave  $S = 40$ ,  $I = 929$ ,  $\hat{\alpha} = 8.51$ ,  $\hat{p} = 109$ . Fisher's formula (6.15) for the standard error of either estimate of  $\alpha$  gives approximately 0.67, indicating a significant difference between them. Whether or not the relative abundances of the species differed at the two traps would be more efficiently

tested by comparing counts of individual species. Formula (6.6) gives for the standard error of either estimate of  $\alpha$  approximately 1.59, against which the observed difference is not significant. A significant difference in the richness of the associations observed at the two traps might be demonstrated, perhaps, by showing that the forty species caught in the field trap had the same relative abundances (within the limits of Poisson variation in numbers caught) as in the roof trap, while the remaining species caught in the roof trap were significantly more abundant, relatively to the others, than in the field trap (where the catches were zero). Apart from some such argument based on comparing catches of individual species, we cannot conclude from the figures for  $S$  and  $I$  alone that the biological associations observed at the two traps differed in diversity index  $\alpha$ .

7. TESTS FOR DEPARTURE FROM THE LOGARITHMIC SERIES FORM OF DISTRIBUTION

Numerous alternatives to the logarithmic series distribution suggest themselves. Fisher has considered the negative binomial form (6.1) with  $k > 0$  and  $X$ ,  $k$  and  $N$  unknown. We can obtain other three-parameter distributions by replacing the Type III distribution of species abundances by any other distribution of a non-negative random variable. The situation is the same as for the heterogeneous Poisson sampling considered in § 2, except that  $n_0$  is not observed and  $N$  is an unknown parameter requiring estimation.

Fisher gives a test for departure from the logarithmic form of distribution towards that at (6.1) with  $k > 0$ . The appropriate statistic, in addition to  $S$  and  $I$ , is

$$J = \sum_{r=2}^{\infty} n_r \left( 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{r-1} \right), \tag{7.1}$$

of which the expected value is  $\frac{1}{2}\alpha[\ln(1+p)]^2$ . Thus  $J - S^2/(2\hat{\alpha})$  may be taken as a test criterion, and its sampling variance can be investigated by the methods already indicated. If we are content with a first-order asymptotic result when  $\alpha \rightarrow \infty$ , we may consider the matrix of expectations of second derivatives of the logarithm of the likelihood function (6.1), the differentiations being with respect to  $X$ ,  $\alpha$ , and  $k$ . Setting  $k = 0$ , we obtain

$$\left\{ \text{var} \left( J - \frac{S^2}{2\hat{\alpha}} \right) \right\} \sim \frac{\alpha \left[ \frac{1}{12} \{ -\ln(1-X) \}^3 \{ -\ln(1-X) - 4X \} + \phi(X) \{ -\ln(1-X) - X \} \right]}{-\ln(1-X) - X}, \tag{7.2}$$

where 
$$\phi(X) = \sum_{r=2}^{\infty} \left( 1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{(r-1)^2} \right) \frac{X^r}{r}. \tag{7.3}$$

When  $p$  is large 
$$\phi(X) \sim A \ln p - B, \tag{7.4}$$

where, in terms of the Riemann  $\zeta$ -function,

$$A = \zeta(2) = 1.6449, \quad B = 2\zeta(3) = 2.4041; \tag{7.5}$$

and hence, ignoring a factor  $1 + O(p^{-1})$ , as in (6.6),

$$\text{var} \left( J - \frac{S^2}{2\hat{\alpha}} \right) \sim \frac{\alpha \left[ \frac{1}{12} (\ln p)^3 (\ln p - 4) + (A \ln p - B) (\ln p - 1) \right]}{\ln p - 1}. \tag{7.6}$$

Fisher denotes the right-hand side of (7.2) by  $i$  and has tabulated it.

Applications of the distribution made by Williams suggest that another sort of departure from the logarithmic series form may be worth investigating. In a complex association it might happen that while certain components of the association exhibit logarithmic series

distributions the whole does not, since the component distributions have different parameters. Let  $\alpha_i, X_i, p_i$  relate to the  $i$ th component association ( $i = 1, 2, \dots, \nu$ ), and consider a logarithmic series distribution with parameters  $\alpha_0, X_0, p_0$ , chosen to give the same expected numbers of species and individuals as in the whole association. Then

$$\left. \begin{aligned} \alpha_0 \ln(1+p_0) &= \sum_{i=1}^{\nu} \alpha_i \ln(1+p_i), \\ \alpha_0 p_0 &= \sum_{i=1}^{\nu} \alpha_i p_i. \end{aligned} \right\} \quad (7.7)$$

Hence 
$$\frac{\ln(1+p_0)}{p_0} = \sum_{i=1}^{\nu} w_i \frac{\ln(1+p_i)}{p_i}, \quad (7.8)$$

where 
$$w_i = \alpha_i p_i / \sum_{j=1}^{\nu} \alpha_j p_j,$$

so that  $\ln(1+p_0)/p_0$  is a weighted mean of  $\ln(1+p_i)/p_i$ . If  $E$  and  $E^*$  denote respectively expectations for the actual distribution and for the fitted logarithmic series distribution, we have

$$\left. \begin{aligned} E(n_r) &= \sum_{i=1}^{\nu} \alpha_i p_i \frac{p_i^{r-1}}{r(1+p_i)^r}, \\ E^*(n_r) &= \left\{ \sum_{i=1}^{\nu} \alpha_i p_i \right\} \frac{p_0^{r-1}}{r(1+p_0)^r}. \end{aligned} \right\} \quad (7.9)$$

We therefore consider  $Y \equiv p^{r-1}(1+p)^{-r}$  as a function of  $Z \equiv p^{-1} \ln(1+p)$ . When  $r = 1$ , we find  $d^2 Y/dZ^2 > 0$  for all  $p$ , so that  $Y$  is a convex function of  $Z$ , and therefore  $E^*(n_1) \leq E(n_1)$ , with equality only if the  $p_i$  are all equal. For  $r \geq 2$ , the sign of  $d^2 Y/dZ^2$  depends on  $p$  and can be studied in detail for each  $r$ . It is not difficult to show that  $E^*(n_r) < E(n_r)$  for a range of small values of  $r$  and also for large values of  $r$ , while the inequality is reversed in a middle range (assuming the  $p_i$  not all equal). If all the  $p_i$  are large and not very unequal, the values of  $r$  where the inequality changes are roughly

$$\frac{p_0}{2 \ln p_0} \quad \text{and} \quad 2p_0.$$

The appropriate statistic for detecting a small degree of inequality in the  $p_i$  is easily seen from the likelihood function to be  $\sum_{r=2}^{\infty} n_r r(r-1)$ , of which the expected value is  $\alpha p^2$  when there is no heterogeneity, i.e. when  $p_i = p$  for all  $i$  and  $\sum_{i=1}^{\nu} \alpha_i = \alpha$ . Hence we may take

$$W = \frac{\hat{\alpha} \sum_{r=2}^{\infty} n_r r(r-1)}{I^2} \quad (7.10)$$

as test criterion. It is easy to show that, asymptotically for large  $\alpha$ ,

$$E(W) \sim 1, \quad \text{var}(W) \sim \frac{1}{\alpha} \left[ \frac{2}{X^2} - \frac{1}{-\ln(1-X) - X} \right], \quad (7.11)$$

or, when  $p$  is large also, 
$$\text{var}(W) \sim \frac{1}{\alpha} \left[ 2 - \frac{1}{\ln p - 1} \right]. \quad (7.12)$$

The test is a limiting form of Test 2 of § 4.



Several distributions of classification of species into genera given by Williams (1944) seem to show this kind of heterogeneity. Thus for Coccidae of the world classified by MacGillivray (Williams's Table 11),  $S = 352$ ,  $I = 1763$ ,  $\hat{\alpha} = 132.2$ ,  $\hat{p} = 13.34$ ,  $W = 2.76$ ,  $s.e.(W) = 0.11$ . The observed number of monotypic genera  $n_1$  is 181, which is much above its expected value, 123.0, on the basis of the logarithmic series distribution. From  $r = 2$  to  $r = 25$ , roughly, the  $n_r$  are on the whole less than expected, while for higher  $r$  they appear again to be greater than expected.

$W$  is closely related to the characteristic  $K$  of Yule (1944). Thus

$$K = 10,000 \frac{W}{\hat{\alpha}} = 10,000 \frac{\sum n_r r(r-1)}{I^2}. \quad (7.13)$$

It is easy to show that, asymptotically for large  $\alpha$ ,

$$E(K) = 10,000\alpha^{-1}, \quad \text{var}(K) = 2 \times 10^8[\alpha^3 X^2]^{-1}, \quad (7.14)$$

if the observations are drawn from a logarithmic series distribution. Simpson (1949) has shown that, asymptotically for large  $p$ , but with  $\alpha$  not assumed large,

$$E(K) = 10,000(\alpha + 1)^{-1}. \quad (7.15)$$

As a statistic for estimating  $\alpha$ ,  $K$  is of low efficiency. Its proper function is to test distribution shape.

I owe my interest in this subject to stimulating conversations with Dr C. B. Williams. Several persons have offered helpful suggestions and criticisms, in particular Mr D. G. Kendall, Mr J. G. Skellam and my colleagues at Cambridge, Dr J. Wishart, Dr H. E. Daniels and Mr D. V. Lindley. The work was begun at Rothamsted Experimental Station, and some help with the computations on which the figure is based was given by Mr B. M. Church.

#### REFERENCES

- AITKEN, A. C. (1939). *Statistical Mathematics*. Edinburgh: Oliver and Boyd.  
 ANSCOMBE, F. J. (1948). *Biometrika*, **35**, 246.  
 ANSCOMBE, F. J. (1949). *Biometrics*, **5**, 165.  
 ANSCOMBE, F. J. (1950). *Ann. Appl. Biol.* **37**, 286.  
 BEALL, G. (1940). *Ecology*, **21**, 460.  
 BUXTON, P. A. (1940). *Parasitology*, **32**, 296.  
 CERNUSCHI, F. & CASTAGNETTO, L. (1946). *Ann. Math. Statist.* **17**, 53.  
 FELLER, W. (1943). *Ann. Math. Statist.* **14**, 389.  
 FINNEY, D. J. (1941). *J. R. Statist. Soc. Suppl.* **7**, 155.  
 FISHER, R. A. (1931). *British Association Mathematical Tables*, **1**, xxvi.  
 FISHER, R. A. (1941). *Ann. Eugen., Lond.*, **11**, 182.  
 FISHER, R. A., CORBET, A. S. & WILLIAMS, C. B. (1943). *J. Anim. Ecol.* **12**, 42.  
 GADDUM, J. H. (1945). *Nature, Lond.*, **156**, 463.  
 GREENWOOD, M. & YULE, G. U. (1920). *J. R. Statist. Soc.* **83**, 255.  
 HALDANE, J. B. S. (1941). *Ann. Eugen., Lond.*, **11**, 179.  
 HALDANE, J. B. S. (1945). *Biometrika*, **33**, 222.  
 HALDANE, J. B. S. (1949). *J. R. Statist. Soc. B*, **11**, 1.  
 IRWIN, J. O. (1941). *J. R. Statist. Soc. Suppl.* **7**, 101.  
 JONES, A. E. (1948). *Biometrika*, **35**, 283.  
 KENDALL, D. G. (1949). *J. R. Statist. Soc. B*, **11**, 230.  
 KENDALL, M. G. (1943). *The Advanced Theory of Statistics*, **1**. London: Griffin.  
 KENDALL, M. G. (1948). *Biometrika*, **35**, 291.

- LÜDERS, R. (1934). *Biometrika*, **26**, 108.  
MCKENDRICK, A. G. (1914). *Proc. Lond. Math. Soc.* **13**, 401.  
NEWBOLD, E. M. (1927). *J. R. Statist. Soc.* **90**, 487.  
NEYMAN, J. (1939). *Ann. Math. Statist.* **10**, 35.  
PÓLYA, G. (1930). *Ann. Inst. Poincaré*, **1**, 117.  
PRESTON, F. W. (1948). *Ecology*, **29**, 254.  
QUENOUILLE, M. H. (1949). *Biometrics*, **5**, 162.  
SHENTON, L. R. (1949). *Biometrika*, **36**, 450.  
SIMPSON, E. H. (1949). *Nature, Lond.*, **163**, 688.  
THOMAS, M. (1949). *Biometrika*, **36**, 18.  
WILLIAMS, C. B. (1937). *Ann. Appl. Biol.* **24**, 404.  
WILLIAMS, C. B. (1944). *J. Ecol.* **32**, 1.  
WISHART, J. (1947). *J. Inst. Actu. Stud. Soc.* **6**, 140.  
YULE, G. U. (1910). *J. R. Statist. Soc.* **73**, 26.  
YULE, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge University Press.