THE NEGATIVE BINOMIAL DISTRIBUTION

by

aroslav
John J. Bartko, B.A.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

STATISTICS

August 1960

Blacksburg, Virginia

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION AND SUMMARY	4
II.	FORMS OF THE NEGATIVE BINOMIAL DISTRIBUTION	6
III.	ANALYTIC PROPERTIES	9
	3.1 The Pascal Form	9
	3.2 Characteristic Function	9
	3.3 Cumulant Generating Function	10
	3.4 Factorial Moments	10
	3.5 Factorial Cumulants	12
	3.6 Skewness and Kurtosis	12
	3.7 Cumulative Distribution Function	12
IV.	COMPOUND, LIMITING, AND TRUNCATED FORMS	15
	4.1 Compound Form	15
	4.2 Limiting Forms	16
	4.2.1 The Poisson Distribution	16
	4.2.2 The Gamma Distribution	17
	4.2.3 The Logarithmic Distribution	18
	4.3 Truncated NBD	20
V.	OCCURRENCE	28
	5.1 Inverse Binomial Sampling	28
	5.2 Poisson Distribution with Varying Mean	28
	5.3 Birth - Death Process	29
	5.4 Contagious Distributions	30

CHAPTER		PAGE
VI.	COMPUTATIONAL PROCEDURES ON THE PASCAL FORM	35
	6.1 Estimating p	35
	6.2 Unbiased Estimate of Variance of $\bar{p}$	36
	6.3 Confidence Limits on p and *Biometrika Table 41*	38
	6.4 An Example of Inverse Binomial Sampling	41
VII.	ESTIMATION OF PARAMETERS	46
	7.1 Efficiency of Estimation by Moments	46
	7.2 Calculation by Maximum Likelihood	51
VIII.	HYPOTHESIS TESTING	56
	8.1 Inverse Binomial Sampling	56
	8.2 χ^2-tests for Contingency Tables of Negative Multinomial Types	61
IX.	ACKNOWLEDGMENTS	67
X.	REFERENCES	68
XI.	RELATED REFERENCES	71
XII.	VITA	75

I. INTRODUCTION AND SUMMARY

The negative binomial distribution, hereafter referred to simply as NBD, was discussed by Pascal and Fermat (Todhunter, 1865). However, the earliest known derivation and publication is due to Montmort in 1714. A particular form of the negative binomial, the inverse binomial sampling form, is sometimes referred to as the Pascal distribution.

The NBD has the same sample space (0,1,2,...) as the Poisson. It is a two parameter distribution. In particular, the Pascal form is a two parameter distribution over the non-negative integers. The Pascal form is

$$f(x) = \binom{x+r-1}{x} p^r q^x \quad , \qquad x = 0,1,2,\ldots$$

with parameters r, p and $p + q = 1$. We can observe clearly from the above how the term negative binomial arose. For $\binom{x+r-1}{x}q^x$ is generated by a binomial with a negative exponent, viz. $(1 - q)^{-r}$.

It is to be noted that for the negative binomial, $\mu_2 > \mu$ whereas for the ordinary binomial $\mu_2 < \mu$ and for the Poisson $\mu_2 = \mu$. A preliminary calculation of $s^2/\overline{x}$ is often a good guide to which of the three is likely to fit the data best.

In more recent years NBD applications have increased, particularly in health and accident statistics, binomial sampling, population counts, data of contagious distributions and communications.

The author has had an interest in the NBD for some time. This thesis, a result of that interest, is an extensive review of the NBD as found in the major literature on the subject. The various forms of the distribution will be presented with concentration on the Pascal or inverse binomial sampling form with many illustrative examples.

II. FORMS OF THE NEGATIVE BINOMIAL DISTRIBUTION

Throughout the literature various forms of the NBD are encountered. These forms are presented here. For convenience, the substitutions for transformation of the Pascal form into the other forms are given.

The Pascal Form,

$$\binom{r + x - 1}{x} q^x p^r \; ; \quad \text{parameters } r, p \qquad (2.1)$$

$$x = 0, 1, 2, \ldots$$

$$p + q = 1,$$

also

$$\binom{n - 1}{r - 1} p^r q^{n-r} \; ; \quad \text{parameters } r, p, \qquad (2.2)$$

where $r + x = n$.

$$\binom{r + x - 1}{x} \frac{p_o^r}{(1 + p_o)^{x+r}} \; ; \quad \text{parameters } r, p_o \qquad (2.3)$$

in (2.1) let

$$p = \frac{p_o}{1 + p_o} , \qquad q = \frac{1}{1 + p_o} .$$

$$\binom{k + x - 1}{x} p_1^x q_1^{-k-x} \; ; \quad \text{parameters } k, p_1 \qquad (2.4)$$

in (2.1) let

$$p = \frac{1}{q_1} , \qquad q = \frac{p_1}{q_1} , \quad r = k .$$

$$\frac{\Gamma(r_1 + p_2)}{\Gamma(r_1 + 1)\Gamma(p_2)} \left(\frac{m}{p_2 + m}\right)^{r_1}\left(\frac{p_2}{p_2 + m}\right)^{p_2}; \qquad (2.5)$$

parameters m, p_2

in (2.1) let

$$x = r_1, \quad r = p_2, \quad p = \frac{p_2}{p_2+m}, \quad q = \frac{m}{p_2+m} .$$

$$\frac{\Gamma(k + n)}{\Gamma(k)\; n!} \left(1 + \frac{\mu}{k}\right)^{-k}\left(\frac{\mu}{\mu+k}\right)^{n}; \quad \text{parameters } \mu,\ k \qquad (2.6)$$

in (2.1) let

$$x = n,\ r = k,\ p = \frac{k}{k+\mu}\ , \quad q = \frac{\mu}{\mu+k} .$$

$$\frac{\Gamma(k + r_2)}{r_2!\,\Gamma(k)} \left(\frac{m}{m+k}\right)^{r_2}\left(\frac{k}{m+k}\right)^{k}; \quad \text{parameters } m,\ k \qquad (2.7)$$

in (2.1) let

$$x = r_2, \quad r = k, \quad p = \frac{k}{k+m}, \quad q = \frac{m}{m+k} .$$

$$\left(\frac{c}{c+1}\right)^{p_3} \frac{1}{\Gamma(p_3)} \frac{\Gamma(r_3+p_3)}{\Gamma(r_3+1)} \frac{1}{(c+1)^{r_3}}\ ; \quad \text{parameters } p_3,\ c \qquad (2.8)$$

in (2.1) let

$$x = r_3, \quad r = p_3, \quad p = \frac{c}{c+1}\ , \quad q = \frac{1}{c+1} .$$

$$\binom{-r}{k} p^r(-q)^k ; \qquad \text{parameters } r,\ p \tag{2.9}$$

in (2.1) let $x = k$ and note that

$$\binom{r+k-1}{k} = (-1)^{-k}\binom{-r}{k} .$$

III. ANALYTIC PROPERTIES

Although many forms of the NBD are utilized, the Pascal form is of major importance. Its analytic properties will be presented in detail below.

3.1 The Pascal Form

$$f(x) = \binom{x+r-1}{x} q^x p^r , \qquad x = 0,1,2,\ldots$$

3.2 Characteristic Function

$$\begin{aligned}\phi_x(t) &= \mathcal{E}(e^{ixt}) \\ &= \sum_{x=0}^{\infty} e^{ixt} p^r \binom{x+r-1}{x} q^x \\ &= p^r[(1-qe^{it})^{-r}] .\end{aligned}$$

Accordingly, the mean of the Pascal form is rqp^{-1} and the variance is rqp^{-2}. Note that the sum of two independent negative binomial variates $x_1 + x_2$ with the same p is also a negative binomial variate. For

$$\begin{aligned}\phi_{x_1+x_2}(t) &= \phi_{x_1}(t) \cdot \phi_{x_2}(t) \\ &= p^{r_1}(1-qe^{it})^{-r_1} \cdot p^{r_2}(1-qe^{it})^{-r_2} \\ &= p^r(1-qe^{it})^{-r}, \quad r = r_1 + r_2 .\end{aligned}$$

3.3 Cumulant Generating Function

$$\psi_x(t) = \ln \phi_x(t)$$

$$= -r \ln [1 - \frac{q}{p} (e^{it} - 1)] .$$

Expanding and identifying coefficients we have

$$\kappa_1 = rqp^{-1} \qquad \kappa_2 = rqp^{-2}$$

$$\kappa_3 = rq(1+q)p^{-3} \qquad \kappa_4 = rq(1 + 4q + q^2)p^{-4} .$$

Finally from the relations between cumulants and moments

$$\mu_1' = rqp^{-1}$$

$$\mu_2 = rqp^{-2}$$

$$\mu_3 = rq(1 + q)p^{-3}$$

$$\mu_4 = rq[1 + q(3r + q + 4)]p^{-4} .$$

3.4 Factorial Moments

The factorial moments can be obtained from the definition or from the factorial moment generating function, (f.m.g.f.).

From the definition

$$\mu_{[k]}' = \mathcal{E}[x(x-1)\ldots(x-k+1)] ,$$

we have

$$\mu'_{[k]} = r(r+1)\ldots(r+k-1)q^k p^{-k} \ .$$

In particular,

$$\mu'_{[1]} = rqp^{-1}$$

$$\mu'_{[2]} = r(r + 1)q^2 p^{-2}$$

$$\mu'_{[3]} = r(r+1)(r+2)q^3 p^{-3}$$

$$\mu'_{[4]} = r(r+1)(r+2)(r+3)q^4 p^{-4} \ .$$

The f.m.g.f. is given by

$$P(1 + t) = \sum_{i=0}^{\infty} \left(\frac{\mu'_{[i]} t^i}{i!}\right) \ .$$

Consequently $\omega(t)$, the f.m.g.f. for the Pascal form is

$$\omega(t) = p^r[1 - q(1 + t)]^{-r}$$

$$= (1 - q\,\frac{t}{p})^{-r} \ .$$

The factorial moments about the origin $\mu'_{[k]}$ are given by the coefficient of $\frac{t^k}{k!}$ in the above expression $\omega(t)$.

Hence $\mu'_{[k]} = \frac{\Gamma(k+r)}{\Gamma(r)} \left(\frac{q}{p}\right)^k$,

or as previously shown

$$\mu'_{[k]} = r(r+1)\ldots(r+k-1)q^k p^{-k} \ .$$

3.5 Factorial Cumulants

The factorial cumulants are given by the logarithm of the f.m.g.f., i.e., $\ln[\omega(t)]$ so that

$$K_{[k]} = (k-1)!\left(\frac{q}{p}\right)^k r .$$

3.6 Skewness and Kurtosis

The coefficients of skewness and kurtosis are:

$$\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \frac{1+q}{(rq)^{\frac{1}{2}}}$$

$$\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \frac{(1+4q+q^2)}{rq} .$$

3.7 Cumulative Distribution Function

A very interesting and convenient property of the Pascal form is that its cumulative distribution function

$$F(y) = \sum_{x=0}^{y} \binom{x+r-1}{x} q^x p^r$$

can be computed by entering tables of the incomplete Beta function. This result due to Pearson and Fieller (1933) is shown below.

The remainder after y terms of the Maclaurin series in the integral form for a function g(x) is

$$R = \frac{1}{\Gamma(y)} \int_0^x g^{(y)}(x-t)t^{y-1}dt.$$

Writing the above NBD as $p^r[(1-q)^{-r}]$, and taking $g(q) = (1-q)^{-r}$ we have

$$g^{(y)}(q) = (-1)^y\ [-r(-r-1)(-r-2)\ldots(-r-y+1)]\ (1-q)^{-r-y}$$

$$= \frac{\Gamma(r+y)(1-q)^{-r-y}}{\Gamma(r)} .$$

Writing this expression with the p^r factor R is therefore,

$$R = p^r \frac{\Gamma(r+y)}{\Gamma(r)\Gamma(y)} \int_0^q (1-q+t)^{-r-y} t^{y-1} dt .$$

Let $1-u = \frac{1-q}{1-q+t}$ then when $t = 0$, $u = 0$ and $t = q$, $u = q$;

$dt = \frac{1-q}{(1-u)^2} du$ and now R becomes

$$p^r \frac{\Gamma(r+y)}{\Gamma(r)\Gamma(y)} \int_0^q (1-u)^{r-1} u^{y-1} du\ (1-q)^{-r} .$$

$$R = \frac{\Gamma(r+y)}{\Gamma(r)\Gamma(y)} B_q(y,r)$$

$= I_q(y,r)$; ~~when~~ where $I_x(a,b)$ is the standard notation for the incomplete Beta function:

$$I_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1} dt}{B(a,b)} .$$

Therefore,

$$\sum_{x=y}^{\infty} \binom{x+r-1}{x} p^r q^x = I_q(y,r). \qquad (3.7.1)$$

Now $F(y)$, the sum of the terms of the Pascal form from $x = 0, 1, \ldots y$ is given by

$$1 - I_q(y + 1, r) = I_p(r, y + 1) .$$

Note that the probability of y failures, i.e., $P(y)$, before the desired number of successes r is given by

$$P(y) = F(y) - F(y-1) ,$$

which is analogous to the situation in the positive binomial.

It is interesting to note that the use of incomplete Beta tables for the evaluation of this cumulative distribution function has been rediscovered by Patil (1960).

IV. COMPOUND, LIMITING, AND TRUNCATED FORMS

4.1 Compound Form

In this section, using the results of Kemp and Kemp (1956), it will be shown that forms of the NBD can be combined with the Beta distribution giving rise to a hypergeometric distribution.

In particular, consider the NBD

$$P(k) = \binom{-r}{k}(-q)^k(1-q)^r$$

combined with the Beta distribution of q

$$\frac{q^{a-1}(1-q)^{-a+b}}{B(a,-a+b+1)} ,$$

from which we have

$$P(k) = \int_0^1 \frac{\binom{-r}{k}(-q)^k(1-q)^r q^{a-1}(1-q)^{-a+b}}{B(a, -a+b+1)}\, dq .$$

$$P(k) = \frac{(-1)^k\binom{-r}{k}\, B(a+k, -a+b+r+1)}{B(a, -a+b+1)}$$

$$= \frac{\binom{-a}{k}\binom{b}{-r-k}}{\binom{-a+b}{-r}}$$

which is a hypergeometric distribution where

$$a < 0, \quad r < 0, \quad r-b-1 < 0, \quad 0 < a+b+1, \quad k = 0,1,2,\ldots .$$

$P(k)$ is then the probability function of the number of failures k, before the desired number of successes r is obtained, where the parameter q is known to have a Beta distribution.

4.2 Limiting Forms

Three limiting forms of the NBD will be discussed in this section.

4.2.1 The Poisson Distribution

Assume the form of the NBD

$$f(x) = \binom{x+r-1}{x} p^r q^x \quad ,$$

whose moment generating function is

$$\left(\frac{p}{1 - qe^{\theta}}\right)^r .$$

Let λ be fixed such that $q = \frac{\lambda}{r}$; let $p \to 1$, $q \to 0$, then

$$\lim_{r \to \infty} \left(\frac{p}{1 - qe^{\theta}}\right)^r$$

$$= \lim_{r \to \infty} \left(\frac{1 - \frac{\lambda}{r}}{1 - \frac{\lambda}{r} e^{\theta}}\right)^r$$

$$= e^{\lambda(e^{\theta}-1)} \quad ,$$

the moment generating function for the Poisson distribution,

$$\frac{e^{-\lambda}\lambda^x}{x!} \quad ,$$

where x is the number of failures before the desired number of successes, r.

4.2.2 The Gamma Distribution

The distribution of "time" before the rth occurrence of some event is given by the Gamma distribution. This distribution arises if trials are made at regular time intervals, Δt. Obviously $r + x$ trials require time $(r + x)\Delta t$. Then $f(x)$ is the probability that $r + x$ time intervals are required in order to obtain the rth success.

Write

$$f(x) = \binom{r + x - 1}{x} q^x p^r$$

as

$$\frac{\Gamma(x+r)\, q^x p^r}{\Gamma(r)\Gamma(x+1)} .$$

Let $p = \lambda\Delta t$, $t = (r+x)\Delta t$, so that the probability p, of a success is proportional to the observed time increment Δt, with constant of proportionality λ. Then

$$\frac{f(x)}{\Delta t} = \frac{\Gamma(\frac{t}{\Delta t})(1-\lambda\Delta t)^{\frac{t}{\Delta t}-r}(\lambda\Delta t)^r}{\Delta t\,\Gamma(r)\Gamma(\frac{t}{\Delta t} -r+1)}$$

$$= \frac{(\frac{t}{\Delta t}-1)\cdots(\frac{t}{\Delta t}-r+1)(1-\lambda\Delta t)^{\frac{t}{\Delta t}}(\lambda\Delta t)^r}{\Delta t\,\Gamma(r)\;(1 - \lambda\Delta t)^r} .$$

Approximating the product of the $r - 1$ factors containing $\frac{t}{\Delta t}$ by the leading term in the polynomial expansion we obtain,

$$\frac{f(x)}{\Delta t} \approx \frac{1}{\Delta t} \left(\frac{t}{\Delta t}\right)^{r-1} \frac{(1 - \lambda\Delta t)^{\frac{t}{\Delta t}}(\lambda\Delta t)^r}{\Gamma(r)(1 - \lambda\Delta t)^r}$$

$$= \frac{t^{r-1}(1 - \lambda\Delta t)^{\frac{t}{\Delta t}}\lambda^r}{\Gamma(r)(1-\lambda\Delta t)^r} \quad .$$

And finally,

$$\lim_{\Delta t \to 0} f(x) = \frac{t^{r-1}\lambda^r e^{-\lambda t}}{\Gamma(r)} ,$$

the Gamma distribution.

4.2.3 The Logarithmic Distribution

Suppose q represents the probability of the presence of an attribute and p of its absence. The NBD for the case considered below will be

$$\binom{x + r - 1}{x} p^r q^x ,$$

where x is the number of attributes present, and r is the number of attributes absent. This strange use of r will become clearer in the following discussion.

Suppose we do not observe, (or may not be capable of observing), the situation in which there is complete absence of the attribute, i.e., $x = 0$. For example consider this derivation of the logarithmic distribution, (Kendall, 1958) as a model for the number, x, of different species found in trap catches. Clearly, if the trap is at all effective, there will be at least one species present, i.e., $x \geq 1$ and obviously the zero class is absent. We are then led to consider a negative binomial with the zero class missing, i.e., the distribution with frequencies proportional to

$$p^r[rq, \frac{r(r+1)q^2}{2!}, \ldots] .$$

The total frequency is $1 - p^r$ and we may write the distribution as

$$\frac{rp^r}{1 - p^r}[q, \frac{(r+1)q^2}{2!}, \frac{(r+1)(r+2)q^3}{3!}, \ldots] .$$

If the experiment is extensive enough so that we can be virtually certain that every species present in the locality is present in at least one of the traps we have the case where $r = 0$ and thus are led to consider the limiting form, with $r = 0$, of this truncated NBD.

Now let $r \to 0$; for the first factor we have,

$$\lim_{r \to 0} \frac{r}{p^{-r}-1} = \lim_{r \to 0} \frac{r}{e^{-r\ln p}-1} = \frac{1}{-\ln p} .$$

The distribution then tends to the form,(considering the brackets with $r = 0$)

$$\frac{-1}{\ln(1-q)} (q, \frac{q^2}{2}, \frac{q^3}{3}, \ldots) .$$

That is to say, the frequency at integral values $x \geq 1$ is the coefficient of t^x in the frequency generating function,

$$P(t) = \frac{1}{\ln(1-q)} \ln(1-qt) \ .$$

This distribution is the logarithmic distribution.

In some instances it may be the case that the truncated NBD itself will give more satisfactory results than the logarithmic distribution. It is discussed in some detail in the next section.

4.3 Truncated NBD

In fitting data where the zero class is not observed it may be the case that a truncated NBD gives a better fit than the logarithmic distribution. Sampford (1955) considers the truncated NBD and gives procedures and tables to obtain iteratively either the moment or maximum likelihood estimates of the parameters. Included also is a discussion of the efficiency of these procedures.

Simplified methods for estimation of the parameters of the truncated distribution, by a modified method of moments and by a modification of maximum likelihood procedures due to Brass (1958), are considered below.

Consider the NBD of the form (2.4)

$$P(i) = \frac{(k+i-1)!p^i}{(k-1)!i!(1+p)^{k+i}} \ ; \quad i = 0,1,2,\ldots;p,\ k>0$$

so that

$$P(0) = \frac{1}{(1+p)^k} \ .$$

To obtain the corresponding probabilities for the truncated distribution, P(i) must be divided by [1 - P(0)]; writing

$$\omega = \frac{1}{1 + p}$$

and

$$\eta = 1 - \omega$$

it follows that

$$P_T(i) = \frac{\omega^k (k+i-1)!\eta^i}{(1-\omega^k)(k-1)!i!} ; \qquad i = 1,2,\ldots,$$

where the subscript T denotes the truncated distribution.

The factorial moments for the truncated distribution are

$$\mu'_{[j]} = \frac{(k+j-1)!\eta^j}{(k-1)!\omega^j(1-\omega^k)}$$

$$\mu'_{[1]} = \mu = \frac{k\eta}{\omega(1 - \omega^k)}$$

$$\mu'_{[2]} = \frac{k(k+1)\eta^2}{\omega^2(1 - \omega^k)}$$

and

$$\mu'_2 = \frac{k\eta(1 + k\eta)}{\omega^2(1 - \omega^k)} .$$

Hence

$$\sigma^2 = \frac{k\eta[1 - \omega^k(1 + k\eta)]}{\omega^2(1 - \omega^k)^2} .$$

We shall also write for the proportion in the first class of the truncated distribution

$$P = P_T(1) = \frac{k\eta\omega^k}{1-\omega^k}$$

and solve this for ω^k. By substituting this, i.e.

$$\omega^k = \frac{P}{k\eta + P}$$

and the expression for μ obtained earlier into the expression for σ^2 we obtain

$$\omega = \frac{\mu(1 - P)}{\sigma^2}$$

and from the expression for μ

$$k = \frac{\omega\mu - P}{1 - \omega} \quad .$$

Replacement of the moments and P by sample values leads to very simple estimates of ω and k,

$$\bar{\omega} = \frac{m}{s^2}\left(1 - \frac{n_1}{n}\right)$$

and $$\bar{k} = \frac{\bar{\omega}m - \frac{n_1}{n}}{1 - \bar{\omega}} \quad , \qquad (4.3.1)$$

where the bars denote the estimates. n_i is the number of sample observations with measurement i, and n the total sample number; m and s^2 are the unbiased sample estimates of μ and σ^2, that is,

$$m = \sum_{i=1}^{\infty} \frac{in_i}{n}$$

and

$$s^2 = \sum_{i=1}^{\infty} \frac{n_i(i-m)^2}{n-1} \quad .$$

$\bar{\omega}$ and $\bar{k}$ are consistent estimates of ω and k but are not unbiased. Of course, when n is large the effect of bias will be slight.

The mean M of the complete NBD is

$$M = \frac{k\eta}{\omega}$$

and its estimate from the above is given as

$$\bar{M} = m - \frac{s^2 n_1}{m(n-n_1)} \quad . \tag{4.3.2}$$

Although fitting by the first two moments and the proportion in the first class has many desirable properties, its efficiency for low values of ω is not sufficiently high for it to be preferred to the maximum likelihood method very often. It appears worth-while then to examine how the maximum likelihood equations may be modified to simplify their solution.

Considering the maximum likelihood estimation procedure the log likelihood is

$$\ell n\ L = \sum_{i=1}^{\infty} n_i \ell n \left[\frac{\omega^k (k+i-1)!(1-\omega)^i}{(1-\omega^k)(k-1)!i!}\right]$$

$$= nk \ln \omega - n \ln(1-\omega^k) + \sum_{i=1}^{\infty} n_i i \ln(1-\omega)$$

$$- \sum_{i=1}^{\infty} n_i \ln i! + \sum_{i=1}^{\infty} n_i \sum_{s=0}^{i-1} \ln(k+s),$$

which gives the maximum likelihood equations

$$\frac{nk}{\omega(1-\omega^k)} - \frac{nm}{1-\omega} = 0, \tag{4.3.3}$$

and

$$\frac{n \ln \omega}{1-\omega^k} + \sum_{i=1}^{\infty} n_i \sum_{s=0}^{i-1} \frac{1}{k+s} = 0,$$

which following Haldane (1941) is conveniently rewritten as

$$\frac{n \ln \omega}{1-\omega^k} + \sum_{j=1}^{R} \frac{1}{k+j-1} \sum_{i=j}^{R} n_i = 0, \tag{4.3.4}$$

where R is the highest observed value of i. Eliminating ω^k by $\omega^k = \frac{P}{k\eta + P}$ and replacing P by its sample value $\frac{n_1}{n}$ we obtain from (4.3.3)

$$\bar{\omega} = \frac{\bar{k} + \frac{n_1}{n}}{\bar{k} + m} \tag{4.3.5}$$

and from (4.3.4)

$$\frac{m(\bar{k} + \frac{n_1}{n})}{\bar{k}(m - \frac{n_1}{n})} \ln\left(\frac{\bar{k} + \frac{n_1}{n}}{\bar{k} + m}\right) + \frac{1}{n} \sum_{j=1}^{R} \frac{1}{\bar{k}+j-1} \sum_{i=j}^{R} n_i = 0.$$

$\bar{k}$ can be found from the second equation where the estimate from the method of moments gives a convenient first entry for

$\bar{k}$ and $\bar{\omega}$ follows immediately from the first after the maximum likelihood value for $\bar{k}$ is determined.

The above methods of fitting are illustrated on the data below taken from Brass (1958). The observations are the number of children ever born to a sample of mothers over forty years of age.

	No. of Children Per Mother												
i	1	2	3	4	5	6	7	8	9	10	11	12	Total
No. of Mothers n_i	49	56	73	41	43	23	18	18	7	7	3	2	340

$$m = 3.9912, \quad s^2 = 5.9734, \quad \frac{n_1}{n} = \frac{49}{340} = 0.1441 \quad .$$

From (4.3.1) we have

$$\bar{\omega} = \frac{3.9912(1 - 0.1441)}{5.9734}$$

$$\bar{\omega} = 0.572$$

and

$$\bar{k} = \frac{0.572 \cdot 3.9912 - 0.1441}{1 - 0.572}$$

$$\bar{k} = 5.00 \tag{4.3.6}$$

also from (4.3.2)

$$\bar{M} = 3.9912 - \frac{5.9734 \cdot 49}{3.9912(340 - 49)}$$

$$\bar{M} = 3.7392 \quad .$$

Now considering the modified maximum likelihood solutions, write

$$\phi = \frac{m}{m - \frac{n_1}{n}}\left(1 + \frac{\frac{n_1}{n}}{k}\right)\ell n\left(1 - \frac{m - \frac{n_1}{n}}{k + m}\right) + \frac{1}{n}\sum_{j=1}^{R}\frac{1}{k+j-1}\sum_{i=j}^{R} n_i \, .$$

Then $\frac{d\phi}{dk} = \phi'$,

$$\phi' = \frac{m}{k(k+m)} - \frac{m\frac{n_1}{n}}{k^2(m - \frac{n_1}{n})}\ \ell n\left(1 - \frac{m - \frac{n_1}{n}}{k + m}\right)$$

$$- \frac{1}{n}\sum_{j=1}^{R}\left(\frac{1}{k+j-1}\right)^2 \sum_{i=j}^{R} n_i \, .$$

The iterative procedure is to calculate ϕ and ϕ' for some value of k; a second value of k which will make ϕ closer to zero is then found by the aid of ϕ', on the assumption that the relation between ϕ and k is roughly linear over a short range. Further improved approximations are obtained, by linear interpolation, from the values of ϕ for the preceding two estimates of k.

k	ϕ	ϕ'
5.00	-0.00064	-0.00258
4.75	+0.00008	-
4.78	0.00000	-

The first approximation to k is taken as 5.00 taken from (4.3.6). Normally this estimate can be rounded to a convenient value. ϕ and ϕ' can then be calculated quite readily. The second approximation to k is

$$5.00 - \frac{-0.00064}{-0.00258} = 4.75.$$

One linear interpolation between 5.0 and 4.75 gives 4.78 for which ϕ is zero to the accuracy which is justified by the number of digits calculated for m and $\frac{n_1}{n}$.

From this value of $\bar{k}$, $\bar{\omega}$ can be found directly from (4.3.5).

$$\bar{\omega} = \frac{4.78 + 0.1441}{4.78 + 3.9912} = 0.561 .$$

These estimates are compared in the table below.

	ω	k
Modified Method of Moments	0.572	5.00
Modified Maximum Likelihood	0.561	4.78

The estimates by the two procedures differ little and the efficiency of both methods is about 95%.

V. OCCURRENCE

A number of ways are known in which the NBD can arise. Several such situations are discussed below.

5.1 Inverse Binomial Sampling

If a proportion p of individuals in a population possesses a certain characteristic, the number of observations in excess of r which must be taken to obtain exactly r individuals with the characteristic has a NBD, i.e.,

$$f(x) = \binom{x+r-1}{x} p^r q^x, \qquad x = 0,1,2,\ldots \quad .$$

This is also called the Pascal distribution and was first used in this context by Montmort and Pascal.

As an example of inverse binomial sampling consider the following situation. Suppose in the mating of two pure strains of mice the probability that any offspring is of a certain type A is p, a constant. The experimenter is satisfied as soon as he has r A's. The distribution of the number of offspring in excess of r that he has to examine to achieve this clearly has the Pascal form.

5.2 Poisson Distribution with Varying Mean

If the mean λ of a Poisson distribution varies randomly from occasion to occasion a compound Poisson distribution results. In particular if λ is distributed as a gamma distribution, i.e.,

$$\frac{c^{\alpha}e^{-c\lambda}\lambda^{\alpha-1}}{\Gamma(\alpha)},$$

then multiplying this distribution by $\frac{e^{-\lambda}\lambda^{r}}{r!}$ and integrating over λ, i.e.,

$$\frac{\int_0^\infty c^{\alpha}e^{-c\lambda}\,\lambda^{\alpha-1}e^{-\lambda}\,\lambda^{r}\,d\lambda}{\Gamma(\alpha)\,r!}$$

gives the NBD

$$\left(\frac{c}{c+1}\right)^{\alpha}\frac{\Gamma(r+\alpha)}{\Gamma(\alpha)\Gamma(r+1)(c+1)^{r}},$$

which is of the form (2.8) with parameters α and c. This fact was first realized by Greenwood and Yule (1920) and has been used in studying accident and sickness statistics where the varying Poisson mean takes account of the varying risk throughout the population.

5.3 Birth-Death Process

Consider a population of members which can by splitting or otherwise give birth to new members but cannot die. Assume that during any short time interval of length Δt each member has probability $\lambda\Delta t + o(\Delta t)$ to create a new one; the constant λ determines the rate of increase of the population. If there is no interaction among the members and at time t the population size is n, then the probability of an increase during $(t, t + \Delta t)$ is $n\lambda\Delta t + o(\Delta t)$. The probability $Pn(t)$ that the population numbers exactly n elements satisfies

$$P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t)$$

$$P_0'(t) = -\lambda_0 P_0(t)$$

with $\lambda_n = n\lambda$, that is

$$P_n'(t) = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t), \quad (n \geq 1).$$

If i is the population size at time $t = 0$, then the conditions $P_i(0) = 1$, $P_n(0) = 0$ for $n \neq i$ apply. The solution for $n \geq i$ is given by

$$P_n(t) = \binom{n-1}{n-i} e^{-i\lambda t}(1-e^{-\lambda t})^{n-i} .$$

This distribution is a special case of the NBD. This type of process was first studied by Yule (1924) in connection with the mathematical theory of evolution. Furry (1939) found a similar result in work with cosmic radiation. Kendall (1949) considers a population of individuals who independently reproduce individuals at random intervals.

The model has been applied to the growth of some living populations, eg, populations of bacteria, and to spread of an infectious disease in a community.

5.4 Contagious Distributions

The use of the NBD in "contagion" was first derived by McKendrick (1914). Student in 1907 noted the applicability of the distribution when he wrote, "If the presence of one individual in a division increases the chance of other individuals falling into that division, a negative binomial distribution will fit best, but if it decreases the chance, a positive binomial."

Quenouille (1949) considers the following situation. Suppose that the number of groups observed on any one occasion is distributed in the Poisson form, so that the probability of observing n groups is

$$P(n \text{ groups}) = \frac{e^{-m}m^n}{n!},$$

where m is the mean of the Poisson distribution. Then the probability of observing s individuals in any group is

$$P(s \text{ individuals}) = \sum_{n=0}^{\infty} P(n \text{ groups}) P\binom{s \text{ individuals}}{\text{in } n \text{ groups}}$$

where the probability of observing s individuals in any one group is $\alpha x^s/s$ or the coefficient of t^s in $-\alpha \ln(1 - xt)$, where $\alpha = -1/\ln(1 - x)$. It will be recognized that this is the logarithmic distribution considered earlier. This meets the requirement that the presence of an individual in a group increases the chance of other individuals falling in the group. Likewise, as Quenouille shows, the probability of observing s individuals in n groups is the coefficient of t^s in $[-\alpha \ln(1 - xt)]^n$. Thus we have

$$P(\text{s individuals}) = \text{coefficient of } t^s \text{ in}$$

$$\sum_{n=0}^{\infty} \frac{e^{-m}m^n}{n!}[-\alpha \ln(1 - xt)]^n$$

$$= \text{coefficient of } t^s \text{ in}$$

$$\exp[-m - \alpha m \ln(1 - xt)]$$

$$= \text{coefficient of } t^s \text{ in}$$

$$(1 - xt)^{-\alpha m} e^{-m}$$

$$= \frac{(1-x)^{\alpha m}(\alpha m + s - 1)!x^s}{(\alpha m - 1)!s!}, \text{ since } (1-x)^{-\alpha} = e.$$

This is the same as the $(s + 1)$th term in a negative binomial series with parameter $1 - x$ and index αm. Consequently, the probability distribution of the number of individuals in random samples is a negative binomial.

If colonies or groups of individuals are distributed randomly over an area (or in time) in such a way that the number of groups observed in samples of fixed area (or duration) has a Poisson distribution and the number of individuals within each group has a logarithmic distribution, then the total number of individuals has a NBD.

Irwin (1941), considers a derivation concerning numbers of individual accidents when the number of multiple accidents is Poisson distributed and the number of individuals within a multiple accident is logarithmically distributed.

The following situation considered by Lüders (1934) is one in which the NBD arises in accident theory in connection with multiple accidents.

Let

$$P{v_n} = \frac{e^{-h_n} h_n^{v_n}}{v_n!}$$

be the distribution of a time interval containing v_n accidents each involving n people. If an event such as an accident occurs at random, and if the events may be classified according as they give rise to 1,2,3,... individual cases (eg, individual accidents), and if the events are independent, then the probability of equal time intervals containing r individual accidents is

$$P_r = e^{-h_1-h_2\cdots-h_r} \sum_{v_1+2v_2+\cdots rv_r=r} \frac{h_1^{v_1} h_2^{v_2}\cdots h_r^{v_r}}{v_1!v_2!\cdots v_r!} .$$

Alternatively, P_r is the coefficient of z^r in

$$f(z) = e^{-h_1-h_2-h_3-\cdots} e^{h_1z+h_2z^2+h_3z^3+\cdots} .$$

If, in particular,

$$h_n = \frac{hd^{n-1}}{n(1+d)^n} ,$$

then

$$h_1z+h_2z^2+\cdots = \frac{h}{d}\left[\frac{dz}{1+d} + \frac{d^2z^2}{2(1+d)^2} + \frac{d^3z^3}{3(1+d)^3} + \cdots\right]$$

$$= - \frac{h}{d} \ln(1 - \frac{zd}{1+d}) ,$$

consequently

$$f(z) = e^{\frac{h}{d} \ln(1 - \frac{d}{1+d})} e^{-\frac{h}{d} \ln(1 - \frac{zd}{1+d})}$$

$$= (1 + d)^{-\frac{h}{d}} (1 - \frac{zd}{1+d})^{-\frac{h}{d}} ,$$

a form of the NBD.

This is the distribution of the number of accidents to individual people in the chosen time interval, not the distribution of persons having 0,1,2,... accidents in a given time interval.

The first of these situations, the Pascal form or inverse binomial sampling, is the simplest mathematically, and is the only one where the mathematical model is likely to hold exactly in practice.

VI. COMPUTATIONAL PROCEDURES ON THE PASCAL FORM

6.1 Estimating p

Occasions may arise in which the frequency of an attribute p is unknown. For the inverse binomial form the estimation of p from the experimental data is a simple matter.

Consider this case, from Haldane (1943), in haematology in which it is desired to know the frequency p, of an abnormal type of red blood cell. It is important to note here that we are estimating the frequency of an attribute taken from a population so much larger than the sample that it may be regarded as infinite.

Let p be the frequency of abnormal cells and $q = 1 - p$,
r be the number of abnormal cells counted,
$n = r+x$ be the total number of cells counted,
$\bar{p} = \frac{r-1}{n-1}$ be the estimate of p.

Clearly n may have any positive value exceeding r-1. Let P(n) be the probability that exactly n cells are counted before r abnormal cells are observed. This formulation yields

$$P(n) = \binom{n-1}{r-1} p^r q^{n-r} ,$$

the inverse binomial form of the NBD.

Note that $\bar{p}$ is an unbiased estimate of p for,

$$\mathcal{E}(\overline{p}) = \sum_{n=r}^{\infty} \frac{r-1}{n-1} P(n)$$

$$= \sum_{n=r}^{\infty} \binom{n-2}{r-2} p^r q^{n-r} .$$

Let $n - r = x$, then

$$\sum_{n=r}^{\infty} \binom{n-2}{r-2} p^r q^{n-r} = p^r \sum_{x=0}^{\infty} \binom{r+x-2}{x} q^x$$

$$= p^r (1-q)^{-r+1}$$

$$= p .$$

Thus $\overline{p}$ is an unbiased estimate of p. It should be noted that $\overline{p}$ is of the same general form as the biased maximum likelihood estimate of p, i.e., r/n.

6.2 Unbiased Estimate of Variance of $\overline{p}$

Haldane (1943) in giving the unbiased estimate of p also gave an expression for the variance of $\overline{p}$. The variance, however, was approximate and biased. Finney (1949) found an exact expression for the variance of $\overline{p}$ and an unbiased estimate of it.

By definition,

$$\mathrm{Var}(\overline{p}) = \mathcal{E}(\overline{p}^2) - [\mathcal{E}(\overline{p})]^2$$

$$= \mathcal{E}(\overline{p}^2) - p^2 .$$

In order to investigate $\mathcal{E}(\overline{p}^2)$, Finney introduced the auxiliary function, $U(\alpha, \beta)$ defined by

$$U(\alpha, \beta) = \binom{n - \alpha - \beta - 1}{r - \alpha - 1} \Big/ \binom{n-1}{r-1} .$$

The average value of $U(\alpha, \beta)$ i.e., $\mathcal{E}[U(\alpha,\beta)]$ is

$$\mathcal{E}[U(\alpha,\beta)] = \sum_{n=r}^{\infty} P(n)\, U(\alpha,\beta)$$

$$= p^{\alpha} q^{\beta} ,$$

where $\quad P(n) = \binom{n-1}{r-1} p^r q^{n-r}$.

Expanding $\overline{p}^2$ in a series of $U(2,\beta)$ namely

$$\overline{p}^2 = U(2,0) + \frac{U(2,1)}{r} + \frac{2!U(2,2)}{r(r+1)} + \frac{3!U(2,3)}{r(r+1)(r+2)} + \cdots$$

we have

$$\mathcal{E}(\overline{p})^2 - p^2$$
$$= \frac{p^2 q}{r}\left(1 + \frac{2!q}{r+1} + \frac{3!q^2}{(r+1)(r+2)} + \cdots\right)$$

$$= \mathrm{Var}(\overline{p}) \quad .$$

The above suggests the estimate $s^2 = \overline{p}^2 - U(2,0)$, which is obviously unbiased for,

$$\mathcal{E}(s^2) = \mathcal{E}(\overline{p}^2) - \mathcal{E}[U(2,0)]$$

$$= \mathcal{E}(\overline{p}^2) - p^2$$

$$= \mathrm{Var}(\overline{p}).$$

s^2 can be expressed very simply in terms of the sample since

$$s^2 = \overline{p}^2 - U(2,0)$$

$$= \left(\frac{r-1}{n-1}\right)^2 - \frac{(r-1)(r-2)}{(n-1)(n-2)}$$

$$= \frac{(r-1)(n-r)}{(n-1)^2(n-2)}$$

$$= \frac{\bar{p}(1-\bar{p})}{n-2} .$$

This form is very suitable for computation. For the planning of sampling investigations, however, an expression in terms of r is more suitable, i.e.,

$$s^2 = \frac{\bar{p}^2(1-\bar{p})}{r-1-\bar{p}} .$$

The standard error is a satisfactory indicator of the error of estimation of p only when r is large. For small r a clever method also due to Finney (1949) for reading confidence limits on p directly from Biometrika Table 41 will be shown in the next section.

6.3 Confidence Limits on p and Biometrika Table 41

It may not have been generally realized that methods and tables for determining exact confidence limits for direct or binomial sampling may be adapted very easily to inverse binomial sampling. For if p_u is the upper $\frac{1}{2}\alpha$ confidence limit for a value of p estimated from inverse sampling, by definition it is the value of p for which the probability of obtaining the quota r, in n or more trials is exactly $\frac{1}{2}\alpha$. This value p_u is the same value of p which would give (r-1) or less with the attribute in a sample of predetermined size (n-1) in direct sampling. Likewise p_L, the corresponding lower limit, is the value of p which would give the quota in n trials or less with a probability of $\frac{1}{2}\alpha$, and is, therefore, the same as the value of p which in direct sampling, would give r or more with the attribute in a sample of n. Hence the rule may be stated:

(i) The upper limit on p is the upper limit for a direct binomial sample which has r-1 successes in n-1 trials; i.e., enter Biometrika Table 41 with c = r-1, n = n-1.

(ii) The lower confidence limit on p in inverse binomial sampling is the lower limit for a direct binomial sample which has r successes in n trials; i.e., enter Biometrika Table 41 with c = r, n = n.

A more explicit proof of the relationships on which the above rules are founded can be given as follows.

The upper limit p_u for the symmetric 1-α interval on p in inverse binomial sampling satisfies

$$\sum_{x=y}^{\infty} \binom{x+r-1}{x} p_u^r (1-p_u)^x = \tfrac{1}{2}\alpha$$

by definition. Now let y+r = n, then

$$\sum_{x=y}^{\infty} \binom{x+r-1}{x} p_u^r (1-p_u)^x = I_{(1-p_u)}(y, r)$$

by (3.7.1). Using the identities of the Beta distribution,

$$\begin{aligned} I_{(1-p_u)}(y,r) &= 1 - I_{p_u}(r,y) \\ &= 1 - I_{p_u}(r, n-r) \ . \end{aligned}$$

For the analogous binomial situation the upper limit p_u' satisfies

$$\sum_{x=0}^{r-1} \binom{n-1}{x} p_u'^{x} (1-p_u')^{n-1-x} = \tfrac{1}{2}\alpha \ .$$

Since

$$\sum_{x=0}^{n-1} \binom{n-1}{x} p_u'^{x} (1-p_u')^{n-1-x} = 1 - I_{p_u'}(r, n-r),$$

$$p_u = p_u' \ .$$

The lower limit p_L for the symmetric $1-\alpha$ interval on p in inverse binomial sampling satisfies

$$\sum_{x=0}^{y} \binom{x+r-1}{x} p_L^{r} (1 - p_L)^{x} = \tfrac{1}{2}\alpha$$

by definition. Let $y+r = n$, then

$$\sum_{x=0}^{y} \binom{x+r-1}{x} p_L^{r} (1 - p_L)^{x} = I_{p_L}(r, y+1)$$

$$= I_{p_L}(r, n - r + 1) \ .$$

For the analogous binomial situation the lower limit p_L' satisfies

$$\sum_{x=r}^{n} \binom{n}{x} p_L'^{x} (1 - p_L')^{n-x} = \tfrac{1}{2}\alpha$$

$$= I_{p_L'}(r, n - r + 1).$$

Hence $\quad p_L = p_L' \ .$

As an example of the use of this procedure consider the following situation.

An experimenter has concluded an inverse sampling plan in which he obtained four desired attributes in ten trials. Here $r = 4$, $n = 10$, the unbiased estimate of p is $\frac{r-1}{n-1} = 1/3$. In order to find the upper confidence limit on p i.e., p_u enter Biometrika Table 41 with $c = 3$ and $n = 9$ and find the value of about 0.70. The lower confidence limit is found by entering the table with $c = 4$, $n = 10$, to which corresponds the value 0.11. Consequently, the two sided 95% confidence interval for p is $0.11 \leq p \leq 0.70$.

6.4 An Example of Inverse Binomial Sampling

The estimation of the total size of plant or animal populations is of great importance in a variety of biological problems, such as studies of population growth, ecological adaptation, genetic constitution, and natural selection. Practical consequences of these studies are the maintenance of human food supplies and the control of insect pests.

A basic technique from Chapman (1952) for the estimation of a population size is as follows. One catches, tags and releases a certain number of animals taken at random from the population. A further random sample is caught and the proportion of marked animals noted. Then the total number of marked animals released divided by the proportion of marked animals in the second sample can be used as an estimate of the total population size.

Consider an inverse binomial sampling plan in which the number of tagged animals to be recovered by sampling is fixed.

The following notation will be used:

N : the population size being estimated,

t : the number of marked animals placed in the population,

r : the number of tagged animals recovered in the sample,

n : the total number of animals taken to recover r tagged animals.

If sampling is with replacement, $P(n)$, the probability of having to sample n individuals to obtain r marked ones, is given by

$$P(n) = \binom{n-1}{r-1}\left(\frac{t}{N}\right)^r\left(1 - \frac{t}{N}\right)^{n-r} ,$$

with r the fixed parameter and n the random variable.

The moment generating function of the above is

$$\left(\frac{t}{N}\right)^r e^{r\theta}\left[1 - e^{\theta}\left(1 - \frac{t}{N}\right)\right]^{-r} .$$

Consequently, $\mathcal{E}(n) = rN/t$ and nt/r is an unbiased estimate of N. The variance of nt/r is $(N^2 - Nt)/r$.

By considering the moment generating function of z where z is given by

$$z = \frac{nt/r - N}{\sqrt{N^2/r - Nt/r}} , \qquad (6.4.1)$$

Chapman shows by routine algebra and the usual manipulations that $\ell n\, M_z(\theta)$ tends to $\theta^2/2$ as r tends to infinity. Therefore, z is approximately distributed N(0,1) for large r. This fact may be used to set up confidence intervals and tests for N. Chapman did not give the confidence interval for N, but it is found by solving (6.4.1) for N. In particular the 95% confidence limits for N are found from

$$N = \frac{r(2nt-1.96^2t) \pm \sqrt{r^2(2nt-1.96^2t)^2 - 4n^2t^2r(r-1.96^2)}}{2r(r - 1.96^2)} \; .$$

However, Chapman seems to have overlooked that a more satisfactory approach to confidence limits on N is to use the procedure of section 6.3, i.e., the use of Biometrika Table 41. Let t/N = p, replacing N by nt/r, p becomes r/n, then by the rules of section 6.3 p_L and p_u can be determined such that

$$P(p_L \leq p \leq p_u) \geq 1 - \alpha \; .$$

Then on replacing p by t/N the confidence statement

$$P(t/p_u \leq N \leq t/p_L) \geq 1 - \alpha$$

is obtained.

For r and t/N both small, as will more frequently be the case, Chapman (1952) further considers the moment generating function of the random variable 2np, where p = t/ N for convenience.

$$M_{2np}(\theta) = p^r e^{2pr\theta}[1 - (1-p)e^{2p\theta}]^{-r}$$

and

$$\ell n\, M_{2np}(\theta) = 2pr\theta - r\ell n(1 - 2\theta + R(p))$$

where $R(p) \rightarrow 0$ as $p \rightarrow 0$, therefore,

$$\lim_{p \rightarrow 0} M_{2np}(\theta) = (1 - 2\theta)^{-r}$$

which is the moment generating function of the χ^2 distribution with 2r degrees of freedom. Hence $2nt/N$ has approximately a χ^2 distribution with 2r degrees of freedom. The confidence interval for N in this case is

$$\frac{2tn}{\chi^2_{2r(1-\alpha/2)}} < N < \frac{2tn}{\chi^2_{2r(\alpha/2)}} ,$$

where $\chi^2_{2r(1-\alpha/2)}$ and $\chi^2_{2r(\alpha/2)}$ are critical values such that areas $1-\alpha/2$ and $\alpha/2$ are to the left of these points.

Chapman made no comparison of the confidence intervals which he discussed. Table I compares the confidence intervals found for N by the normal approximation and by use of Biometrika Table 41. For this comparison $t = 100$, $r = 50$, $n = 100$ and $\mathcal{E}(N) = 200$.

Table I

	Two Sided 95% C.I. for N
Normal	$170 \leq N \leq 255$
Table 41	$166 \leq N \leq 250$

Table II gives a comparison of the confidence intervals found for N by the χ^2 approximation and by use of Biometrika Table 41. For this comparison t = 100, r = 4, n = 40, and $\mathcal{E}(N) = 1000$.

Table II

	Two Sided 95% C.I. for N
χ^2	$456 \leq N \leq 3670$
Table 41	$416 \leq N \leq 3333$

Table I shows that Biometrika Table 41 should be used for confidence limits on N in place of the normal approximation since results are exact and the labor saved is considerable. For both cases in Table II, the labor is slight, however, the exact method of Table 41 should be used.

Further, it is worth-while to note that the method of Biometrika Table 41 is a general, exact method and applicable in many situations, and should be used over approximations whose conditions may not always be satisfied.

VII. ESTIMATION OF PARAMETERS

7.1 Efficiency of Estimation by Moments

The estimation of parameters of other forms of the NBD is not as straightforward as it is in the case of the Pascal form. Fisher's (1941) work is presented below on this topic.

In this section the determinants of the covariance matrices for large-sample estimates found by moments and by maximum likelihood will be compared yielding the expression 1/E. Also a simple guide will be presented which will indicate when moment estimates are satisfactory.

Assume the form (2.4) of the NBD

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

with parameters k, p; q = 1+p. The moments for this expression are as follows

$$\mu_1' = pk$$

$$\mu_2 = pqk$$

$$\mu_3 = pq(q+p)k$$

$$\mu_4 - 3\mu_2^2 = pq(1+6pq)k \, .$$

For large samples and the method of moments, estimates of p and k are

$$p = \frac{s^2 - \bar{x}}{\bar{x}} \quad , \qquad k = \frac{\bar{x}^2}{s^2 - \bar{x}} \, , \qquad (7.1.1)$$

where $\bar{x}$ is the mean and s^2 is the variance calculated from the sample data.

The covariance matrix of $\bar{x}$ and s^2 for large sample size 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

$$\frac{\partial(p,\, k)}{\partial(\bar{x},\, s^2)} \ .$$

The Jacobian is

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

by substitution of $\bar{x} = kp$, $s^2 = kp(1+p)$ the Jacobian becomes $\frac{-1}{p^2k}$.

The determinant of the covariance matrix of p and k estimated by the first two moments is then

$$\frac{2}{N^2}\, p^3q^3k^2(k+1)\left(\frac{-1}{p^2k}\right)^2 = \frac{2q^3(k+1)}{pN^2}\ .$$

We may compare this with the corresponding determinant for any method of efficient estimation. The most convenient method of comparison is to calculate the information matrix, which will be the reciprocal of the covariance matrix for efficient estimation. The expression 1/E, then will be given by the product of the determinant of the covariance matrix of the moment estimates multiplied by the determinant of the information matrix of efficient estimates.

Take the general term of the NBD

$$f(x) = \binom{k+x-1}{x}\frac{p^x}{(1+p)^{k+x}}\ , \quad 1+p = q,$$

$$\frac{-\partial^2 \ln f(x)}{\partial p^2} = \frac{x}{p^2} - \frac{k+x}{(1+p)^2}$$

substituting $\mu = x = pk$ we have $i_{pp} = \frac{k}{pq}$. Where i_{pp}, i_{pk}, and i_{kk} are elements of the information matrix. Next,

$$\frac{-\partial^2 \ln f(x)}{\partial p\, \partial k} = \frac{1}{q}\ ,$$

or $$i_{pk} = \frac{1}{q}\ .$$

Finally,

$$\frac{-\partial^2 \ell n\ f(x)}{\partial k^2} = \mathrm{Trig}(k-1) - \mathrm{Trig}(k+x-1)$$

where

$$\mathrm{Trig}(x), \text{ the trigamma function}, = \frac{d^2 \ell n\ x!}{dx^2}$$

$$= \frac{1}{(x+1)^2} + \frac{1}{(x+2)^2} + \ldots + \frac{1}{(x+n)^2} + \ldots.$$

Hence,

$$\frac{-\partial^2 \ell n\ f(x)}{\partial k^2} = \frac{1}{k^2} + \frac{1}{(k+1)^2} + \ldots + \frac{1}{(k+x-1)^2} \quad .$$

This expansion averaged for varying x gives i_{kk} in the form

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

Now let $r = \frac{p}{q}$, $\frac{1}{q} = 1 - r$, then

$$i_{kk} = \sum_{x=1}^{\infty} \frac{(k+x-1)!r^x(1-r)^k}{(k-1)!x!} \left(\frac{1}{k^2} + \ldots + \frac{1}{(k+x-1)^2}\right)$$

$$= (1-r)^k[kr(\frac{1}{k^2}) + \frac{k(k+1)r^2}{2}\left(\frac{1}{k^2} + \frac{1}{(k+1)^2}\right) + \ldots]$$

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

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

Finally the determinant of the information matrix

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

is

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

Multiplying this expression by the determinant of the covariance matrix corresponding with the method of moments we obtain the reciprocal of the efficiency

$$1/E = 2 \sum_{x=2}^{\infty} \frac{p^{x-2}(x-1)!(k+1)!}{xq^{x-2}(k+x-1)!}$$

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

E is always less than unity. The expression shows that it is near unity when $\frac{p}{q(k+2)}$ is small.

A good guide to follow as to whether moment estimates are satisfactory is to observe if

$$(1 + \frac{1}{p})(k + 2) \qquad (7.1.2)$$

exceeds 20, where k and p are numerical values obtained by the method of moments from (7.1.1). If this product exceeds 20 then moment estimates are satisfactorily efficient, i.e., $> 90\%$.

7.2 Calculation by Maximum Likelihood

If the guide, i.e., (7.1.2) indicates that moment estimates are not satisfactory then another method of estimation must be employed.

Estimation by maximum likelihood is efficient. The following by Haldane (1941) is a derivation of the estimates of p and k by maximum likelihood. An iterative scheme for the estimation of k is presented, which is easily performed on a desk calculator. Furthermore, this method does not require the use of tables of digamma functions.

Assume the form of the NBD

$$f(x) = \frac{(k+x-1)!p^x}{x!(k-1)!(1+p)^{k+x}}, \quad q = 1+p.$$

Let n_i be the observed frequency of x_i, R the maximum value of x_i such that $i = 0,1,\ldots,R$. Let $N = \sum_{i=0}^{R} n_i$, the total number of observations and

$$\bar{x} = \frac{1}{N}\sum_{i=0}^{R} x_i n_i$$

the mean value of x_i. Then the likelihood is

$$L = f(x_0, x_1, \ldots, x_R) = \left[\frac{(k+x_0-1)\ldots(k+1)kp^{x_0}}{x_0!(1+p)^{k+x_0}}\right]^{n_0}.$$

$$\left[\frac{(k+x_1-1)\ldots(k+1)kp^{x_1}}{x_1!(1+p)^{k+x_1}}\right]^{n_1}\ldots\left[\frac{(k+x_R-1)\ldots(k+1)kp^{x_R}}{x_R!(1+p)^{k+x_R}}\right]^{n_R} .$$

Now

$$\ln L = \sum_{i=0}^{R} n_i \ln f(x_i)$$

$$= \sum_{i=0}^{R} n_i[x_i \ln p - (k+x_i)\ln(1+p) + \sum_{s=0}^{x_i-1} \ln(k+s) - \ln x_i!] .$$

$$\frac{\partial \ln L}{\partial p} = \sum_{i=0}^{R} n_i\left[\frac{x_i}{p} - \frac{(k+x_i)}{1+p}\right] .$$

Accordingly, the estimate of p is given by $p = \frac{\bar{x}}{k}$ (7.2.1)

where the estimate of k is obtained from the following.

$$\frac{\partial \ln L}{\partial k} = \sum_{i=0}^{R} n_i\left[-\ln(1+p) + \sum_{s=0}^{x_i-1} \frac{1}{k+s}\right] ,$$

and

$$\frac{\partial \ln L}{\partial k} = 0$$

when

$$N\ln(1+p) = \sum_{i=0}^{R} n_i \sum_{s=0}^{x_i-1} \frac{1}{k+s}$$

or

$$N\ln(1+p) = \sum_{i=0}^{R-1} \frac{1}{k+i} \sum_{s=i+1}^{R} n_s .$$

$$N[\ln(k+\bar{x}) - \ln k] = \frac{n_1+n_2+\ldots+n_R}{k} + \frac{n_2+n_3+\ldots+n_R}{k+1} + \ldots + \frac{n_R}{k+R-1} . \qquad (7.2.2)$$

As an illustration of the two methods of estimation i.e., moments and maximum likelihood, consider the following data from Fisher (1941). The table gives a sample of sheep classified according to the number of ticks found on each.

i	0	1	2	3	4	5	6	7	8	9	10	
number of ticks x_i	0	1	2	3	4	5	6	7	8	9	10	
number of sheep n_i	7	9	8	13	8	5	4	3	0	1	2	N = 60

$\bar{x} = 3.25$

$s^2 = 5.9195$

For the method of moments, estimates of p and k are found by (7.1.1) i.e.,

$$p = \frac{s^2 - \bar{x}}{\bar{x}} \quad ; \qquad k = \frac{\bar{x}^2}{s^2 - \bar{x}}$$

$$p = 0.821 \quad ; \qquad k = 3.95 \; .$$

Note that $(1 + \frac{1}{p})(k+2) = 13.2$, so that by the criterion (7.1.2) the maximum likelihood estimates are the desirable ones.

It is worthwhile to note that using the moment estimates i.e., $p = 0.821$, $k = 3.95$ and

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

the efficiency of estimation by moments is about 89%. Therefore, when the expression (7.1.2) exceeds 20 the experimenter can be certain of efficient estimates.

For the method of maximum likelihood is it first necessary to determine the value of k which satisfies (7.2.2). The value of k found by the method of moments, i.e., 3.95 gives a convenient starting entry for the use of (7.2.2). The value of k for which the right hand side equals the left hand side of the expression is the value for which we are seeking and will be the maximum likelihood value for k. The value so found for k was 3.75. As a sample calculation the numerical work is presented below.

$$N[\ell n(k+\bar{x})-\ell n\ k] = \frac{n_1+n_2+\dots+n_R}{k} + \dots + \frac{n_R}{k+R-1} .$$

Hence, from our data

$$60[\ell n(3.75+3.25)-\ell n\ 3.75] = \frac{53}{3.75} + \frac{44}{4.75} + \frac{36}{5.75} + \dots + \frac{2}{12.75}$$

$$37.44900 = 37.44975 ,$$

which for our purposes is very close agreement. Therefore, the maximum likelihood estimate of k is 3.75 and the maximum likelihood estimate of p is found by (7.2.1), i.e.,

$$p = \frac{\bar{x}}{k} , \qquad p = 0.867 .$$

Hence the results are

Parameter	Estimation by 1st and 2nd Moments	Estimation by Maximum Likelihood
p	0.821	0.867
k	3.95	3.75

VIII. HYPOTHESIS TESTING

8.1 Inverse Binomial Sampling

Barnard's (1946) method for testing the hypothesis H_o: $p_1 = p_2$ in inverse binomial sampling is considered below.

This procedure is especially useful in situations where the probabilities p_1, p_2, are small, and where we have some idea beforehand about the true value of the ratio p_1/p_2. For example, in development work on an instrument we may isolate at least two causes of failure. A proposed modification may be hoped to remove one of these causes, while it leaves the other unaffected. From past experience we can often say that if the modification is successful, our failures will be reduced by a certain percentage; and we may then try to determine whether our hopes are justified by an appeal to experiment.

The procedure then consists in arranging to try the modified one, under similar conditions, and to continue to try each until a predetermined number r_1 of failures of the modified instrument and a predetermined number r_2 of failures of the unmodified instrument are observed (inverse binomial sampling). Then if p_1 refers to the modified instrument and p_2 to the unmodified one, and n_1, n_2 are the number of observations made in the two cases, the probability of the pair (n_1, n_2) is

$$\binom{n_1 - 1}{r_1 - 1} p_1^{r_1} q_1^{n_1 - r_1} \binom{n_2 - 1}{r_2 - 1} p_2^{r_2} q_2^{n_2 - r_2}$$

while if $p_1 = p_2 = p$ the probability of obtaining some pair of results having the same total $n_1 + n_2 = N$ is

$$\binom{N - 1}{r_1 + r_2 - 1} p^{r_1 + r_2} q^{N-(r_1+r_2)} .$$

The relative probability on the null hypothesis $p_1 = p_2$, of obtaining the pair (n_1, n_2) out of all results with the same total N, is

$$\frac{[N-(r_1+r_2)]!(r_1+r_2-1)!(n_1-1)!(n_2-1)!}{(N-1)!(r_1-1)!(r_2-1)!(n_1-r_1)!(n_2-r_2)!}$$

which is independent of p. Hence, if we sum over more extreme pairs having the same N, r_1 and r_2, we shall obtain the significance level of our result.

For example, with $r_1 = 1$, $r_2 = 2$, if we find $n_1 = 10$, $n_2 = 4$ we have the significance level

$$= \frac{2!\ 11!}{13!\ 0!\ 1!} \left(\frac{9!\ 3!}{9!\ 2!} + \frac{10!\ 2!}{10!\ 1!} + \frac{11!\ 1!}{11!\ 0!} \right)$$

$$= \frac{1}{13} .$$

If p_1 and p_2 are small, n_1 and n_2 will be large, and an approximation becomes desirable. Noticing that if X is the number of trials required for one failure,

$$P(X \geq x) = q_1^x = \exp(-k_1 x) = 1 - \int_0^{k_1 x} e^{-t}\, dt ,$$

where $k_1 = -\ln(1-p_1)$, we see that $2X/k_1$ is distributed approximately as χ^2 with two degrees of freedom, and the approximation is improved by taking $Y = X - \frac{1}{2}$ in place of X.

Now $n_1 - r_1/2$ is the sum of r_1 variables, independently distributed like $k_1 Y$, and so $2(n_1 - r_1/2)/k_1$ is distributed approximately as χ^2 with $2r_1$ degrees of freedom. Hence

$$R = \frac{(n_1 - r_1/2)k_2 r_2}{(n_2 - r_2/2)k_1 r_1}$$

is distributed approximately as F with $(2r_1, 2r_2)$ degrees of freedom. On the null hypothesis the k's divide out, and so, for n_1, n_2 large we obtain the rule, to calculate

$$R = \frac{(n_1 - r_1/2)r_2}{(n_2 - r_2/2)r_1}$$

and to enter the F tables with this value of R for $(2r_1, 2r_2)$ degrees of freedom.

The test is a valid one, whatever the values of r_1 and r_2. But it is best to fix r_1 and r_2 (a) according to how much experimentation we are prepared to do, on the average, and (b) so as to maximize the chance of getting a significant result when p_1 and p_2 have the values we hope they are going to have.

The expected values of n_1 and n_2 are

$$\mathcal{E}(n_1) = \frac{r_1}{p_1} \quad \text{and} \quad \mathcal{E}(n_2) = \frac{r_2}{p_2}$$

so that guessing p_1 and p_2 we can guess how much experimentation we are likely to do, on the average. If, for example, we expect 2% failures and 5% failures respectively and we take (r_1, r_2) to be (1, 2) then we should expect about 50 trials of the modified design and 40 of the unmodified design.

By entering the F tables with R/θ instead of R, we can of course test whether $p_1 = \theta p_2$, instead of $p_1 = p_2$, if we so wish. Another generalization is to the case of a $2 \times \ell$ trial, where, provided $p_1, p_2, \ldots, p_\ell$ are all small, we can use Bartlett's test or Hartley's test when the degrees of freedom are equal. Barnard did not elaborate this point. It seems worth-while, therefore, to examine how this situation can be referred to Bartlett's test.

Note that

$$\frac{2(n_1 - r_1/2)}{k_1} \frown \chi^2_{2r_1} ,$$

$$\frac{(n_1 - r_1/2)}{r_1} \frown \frac{(\chi^2_{2r_1})k_1}{2r_1} .$$

Identify $k_1 = \sigma_1^2$ and $\dfrac{(n_1 - r_1/2)}{r_1} = s^2_{2r_1}$ i.e., s^2 with $2r_1$ degrees of freedom. Hence, from Bartlett's test

$$M = (N' - \ell)\ell n\ s_p^2 - \Sigma[(n_i' - 1)\ell n\ s_i^2] ,$$

where N' = total number of observations, n_i' = number of observations in the ith sample, ℓ = number of samples, becomes in our notation

$$M = 2\Sigma r_i \ell n\ s_p^2 - 2\Sigma r_i \ell n\ s_i^2 ,$$

where $$s_i^2 = \frac{n_i - r_i/2}{r_i} , \quad i = 1, 2, \ldots, \ell$$

and

$$s_p^2 = \frac{2\Sigma\left[\frac{r_i(n_i - r_i/2)}{r_i}\right]}{2\Sigma r_i}$$

$$= \frac{\Sigma(n_i - r_i/2)}{\Sigma r_i} .$$

Refer M to $\chi^2_{\ell-1}$ where ℓ is the number of groups examined as a first approximation.

If $r_1 = r_2 = \ldots = r_\ell$, as may well be the case since the experimenter is controling the experiment, perform Hartley's test

$$\frac{s^2_{max}}{s^2_{min}}$$

and refer to <u>Biometrika Table 31</u>.

8.2 χ^2-tests for Contingency Tables of Negative Multinomial Types

From a finite population, where sampling takes place with replacement, consisting of categories FE, $\overline{F}E$, $F\overline{E}$, and $\overline{FE}$ in proportions p_{11}, p_{01}, p_{10} and p_{00} respectively, individuals are drawn at random until r $\overline{FE}$'s are found. Obviously the sample size is a random variable. Steyn's (1959) work on this problem is presented below.

The following tables show the observed frequencies and the cell probabilities for the negative fourfold table. These tables are obtained when two events E and F can occur or fail either alone or together and sampling is continued until the rth joint failure, has occurred.

Frequencies

	$\overline{E}$	E	
$\overline{F}$	r	f_{01}	$r + f_{01}$
F	f_{10}	f_{11}	$f_{10} + f_{11} = y$
$r + f_{10}$		$f_{01} + f_{11} = x$	

Probabilities

	$\overline{E}$	E	
$\overline{F}$	p_{00}	p_{01}	$q = 1-p$
F	p_{10}	p_{11}	p
	q'	p'	1

For the probability function in k variables consider, for example, the case of k mutually exclusive events E_i with probabilities α_i, $i = 1,2,\ldots,k$; $\alpha_0 = 1 - \sum_{i=1}^{k} \alpha_i$ being the probability for failure of all the E_i. When sampling is stopped on obtaining the rth failure the probability for having x_i successes of E_i is given by the negative multinomial function

$$\frac{(r + \sum_{i=1}^{k} x_i - 1)!}{(r-1)!x_1!x_2!\ldots x_k!} \alpha_1^{x_1} \alpha_2^{x_2} \ldots \alpha_k^{x_k} \alpha_0^{r} .$$

It was proved by Steyn (1955) that this function tends to the multivariate normal probability function for large r, and that

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - r\alpha_i/\alpha_0)^2}{r\alpha_i/\alpha_0} - \frac{(r + \sum_{i=1}^{k} x_i - r/\alpha_0)^2}{r/\alpha_0}$$

is asymptotically (for $r \to \infty$) distributed as χ^2 with k degrees of freedom.

Consequently for the above case it follows that

$$\chi^2 = \frac{(f_{01}-rp_{01}/p_{00})^2}{rp_{01}/p_{00}} + \frac{(f_{10}-rp_{10}/p_{00})^2}{rp_{10}/p_{00}} + \frac{(f_{11}-rp_{11}/p_{00})^2}{rp_{11}/p_{00}}$$

$$- \frac{(r + \Sigma f_{ij} - r/p_{00})^2}{r/p_{00}} , \qquad (8.2.1)$$

with three degrees of freedom.

The problem now lies in the estimation of the probabilities p_{00}, p_{01}, p_{10} and p_{11} from the observed frequencies and its influence on the degrees of freedom. This problem will be considered under the hypothesis that the events E and F are independent i.e., $p_{00} = qq'$, $p_{01} = qp'$, $p_{10} = pq'$, $p_{11} = pp'$, so that only the determination of p and p' from the observed frequencies is necessary.

Under the hypothesis that rows and columns are independent

(8.2.1) will be of the form:

$$\chi^2 = \frac{(f_{01}-rp'/q')^2}{rp'/q'} + \frac{(f_{10}-rp/q)^2}{rp/q} + \frac{(f_{11}-rpp'/qq')^2}{rpp'/qq'}$$

$$- \frac{(r + \Sigma f_{ij} - r/qq')^2}{r/qq'} . \qquad (8.2.2)$$

Now consider the marginal distribution of x_1 given by

$$\sum_{x_2} \cdots \sum_{x_k} \frac{(r + x_1 + \sum_{i=2}^{k} x_i - 1)!}{(r-1)!\, x_1! \ldots x_k!} \alpha_1^{x_1} \alpha_2^{x_2} \ldots \alpha_k^{x_k} \alpha_0^{r}$$

$$= \frac{(r + x_1 - 1)!}{(r-1)!\, x_1!} \left(\frac{\alpha_1}{\alpha_0 + \alpha_1}\right)^{x_1} \left(\frac{\alpha_0}{\alpha_0 + \alpha_1}\right)^{r}$$

so that the distribution of x_1 for all values of the other variables is a NBD with parameters r and $\alpha_0/(\alpha_0 + \alpha_1)$.

Using the above fourfold table with its frequencies f_{01}, f_{10}, f_{11}, and parameters r, qp', pq', pp', the marginal distribution of E (i.e., the distribution of x which is obtained when sampling is stopped on obtaining the rth joint failure $\overline{FE}$), is distributed as a NBD with parameters r and $qq'/(qq' + p')$.

Similarly, for the marginal distribution of F, y is distributed as a NBD with parameters r and $qq'/(qq' + p)$.

Using the general property of the negative binomial, i.e., $\mathcal{E}(x/r) = q/p$, it follows that x/r is an unbiased estimate of p'/qq' or $p'p/q'q + p'/q'$ since $p + q = 1$. Also y/r is an unbiased estimate of p/qq' or $p'p/q'q + p/q$. Thus $(\hat{p}'/q')$ and $(\hat{p}/q)$, the estimates of p'/q' and p/q, can be obtained by solving the equations

$$(\hat{p}'/q')(\hat{p}/q) + (\hat{p}'/q') = x/r$$

and

$$(\hat{p}'/q')(\hat{p}/q) + (\hat{p}/q) = y/r$$

which give

$$\begin{aligned}(\hat{p}'/q') &= \frac{1}{2r}\left[-(r-x+y) + \sqrt{(r+x+y)^2 - 4xy}\right] \\ (\hat{p}/q) &= \frac{1}{2r}\left[-(r+x-y) + \sqrt{(r+x+y)^2 - 4xy}\right].\end{aligned} \qquad (8.2.3)$$

Using these estimates (8.2.2) reduces to

$$\chi^2 = \left[f_{01} - r\left(\frac{\hat{p}'}{q'}\right)\right]^2 \left[\frac{1}{r\left(\frac{\hat{p}'}{q'}\right)} + \frac{1}{r\left(\frac{\hat{p}}{q}\right)} + \frac{1}{r\left(\frac{\hat{p}}{q}\right)\left(\frac{\hat{p}'}{q'}\right)} - \frac{1}{r\left(\frac{\hat{1}}{qq'}\right)}\right]$$

where

$$\left(\frac{\hat{1}}{qq'}\right) = 1 + \left(\frac{\hat{p}'}{q'}\right) + \left(\frac{\hat{p}}{q}\right) + \left(\frac{\hat{p}}{q}\right)\left(\frac{\hat{p}'}{q'}\right). \qquad (8.2.4)$$

χ^2 is now distributed with one degree of freedom.

Now consider an example. The following table given by Kendall (1943, page 302), shows data for inoculation against cholera.

	Not Attacked	Attacked	Totals
Inoculated	5	431	436
Not-Inoculated	9	291	300 (y)
Totals	14	722 (x)	736

Suppose the experimenter had had reason to expect that the combination inoculated but not attacked would occur very seldom and had planned his experiment to terminate as soon as a frequency of 5 for this combination was obtained, thereby fixing a priori the frequency in one cell of the fourfold table but not the total of the sample.

From (8.2.3)

$$\left(\frac{p'}{q'}\right) = 85.0969$$

and

$$\left(\frac{p}{q}\right) = 0.6969 \ .$$

χ^2 calculated from (8.2.4) is about 9.1 which is significant at the 5% level and one degree of freedom.

If the experiment had been planned so that the positive binomial applied and testing performed in the usual manner, i.e., testing the hypothesis of no association based on two positive binomial marginal distributions the χ^2 would have been 3.3, which is non-significant at the 5% level. In this case the sums of the observed and expected frequencies are equal. In the negative binomial situation, the sums are not equal for one of the cell frequencies is fixed a priori. However, as shown previously the expected frequencies may still be derived from the marginal totals.

The extension to the case where instead of two there are h possibilities E_0, $E_1, \ldots, E_{h-1}$ for the event E, and ℓ possibilities F_0, $F_1, \ldots, F_{\ell-1}$ for the event F gives rise to $h\ell$ joint possibilities. From a population where the individuals fall in these $h\ell$ classes a sample is taken until the frequency in the cell common to the first row and first column is r.

Again χ^2 is

$$\chi^2 = \frac{\sum\limits_{i,j} \left(f_{ij} - \frac{rp_{ij}}{p_{00}}\right)^2}{\frac{rp_{ij}}{p_{00}}} - \frac{\left(r + \Sigma f_{ij} - \frac{r}{p_{00}}\right)^2}{\frac{r}{p_{00}}}$$

with $h\ell - 1$ degrees of freedom under the hypothesis of independence between E and F. If estimates of the parameters from the marginal totals are obtained, the resulting χ^2 is distributed with $(h - 1)(\ell - 1)$ degrees of freedom.

IX. ACKNOWLEDGMENTS

The author wishes to express his sincerest appreciation to for his invaluable guidance and advice in preparing this thesis.

Gratitude is extended to for his many suggestions and constructive criticisms. Thanks are also due for referring the author to a little known reference. Appreciation is expressed to for his advice, assistance, and encouragement to the graduate student.

To , the author is indebted for the preparation of the final typewritten copies.

This study was supported in part by a research grant from the National Institutes of Health, under contract number 2G.

X. REFERENCES

Anscombe, F. J., "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions," Biometrika, 37 (1950), pp. 358-382.

Anscombe, F. J., "Sequential Estimation," Jour. Roy. Stat. Soc., B, 15 (1953), pp. 1-29.

Bailey, T. J., "On Estimating the Size of Mobile Populations from Recapture Data," Biometrika, 38 (1951), pp. 293-306.

Barnard, B. A., "Sequential Tests in Industrial Statistics," Jour. Roy. Stat. Soc., B, 8 (1946), pp. 1-26.

Brass, W., "Simplified Methods of Fitting the Truncated Negative Binomial Distribution," Biometrika, 45 (1958) pp. 59-68.

Chapman, D. G., "Inverse Multiple and Sequential Sample Censuses," Biometrics, 8 (1952), pp 286-306.

Feller, W., An Introduction to Probability Theory and Its Applications. Wiley, New York, (1957).

Feller, W., "On the Theory of Stochastic Processes, with Particular Reference to Applications," Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, (1945-46), pp. 403-432.

Finney, D. J., "Errors of Estimation in Inverse Sampling," Nature, 160 (1947), pp. 195-196.

Finney, D. J., "On a Method of Estimating Frequencies," Biometrika, 36 (1949), pp. 233-235.

Fisher, R. A., "Negative Binomial Distribution," Annals Eugenics, 11 (1941), pp. 182-187.

Fisher, R. A., "Note on the Efficient Fitting of the Negative Binomial," Biometrics, 9 (1954), pp. 197-199.

Fraser, D. A. S., Statistics an Introduction. Wiley, New York, (1958).

Furry, W. H., "On Fluctuation Phenomena in the Passage of High Energy Electrons Through Lead," Phys. Rev., 52 (1939) pp. 569-581.

Greenwood, M., and Yule, G. U., "An Inquiry into the Nature of Frequency Distributions of Multiple Happenings," Jour. Roy. Stat. Soc., 83 (1920), pp. 255-279.

Haldane, J. B. S., "A Labour Saving Method of Sampling," Nature, 155 (1945), pp. 49-50.

Haldane, J. B. S., "On a Method of Estimating Frequencies," Biometrika, 33, (1943) pp. 222-225.

Haldane, J. B. S., "The Fitting of Binomial Distributions," Annals Eugenics, 11 (1941), pp. 179-181.

Hall, H.S. and Knight, S. R., Higher Algebra. MacMillan and Co., Ltd., 4th Edition, (1936), p. 388.

Irwin, J. O., "Discussion on Chambers and Yule's Paper," Jour. Roy. Stat. Soc. B, 7 (1940-41), pp. 101-109.

Kemp, C. D., and Kemp, A. W., "Generalized Hypergeometric Distributions," Jour. Roy. Stat. Soc., B 18 (1956) pp. 202-211.

Kendall, M. G., "Stochastic Processes and Population Growth," Jour. Roy. Stat. Soc., B, 11 (1949), pp. 230-264.

Kendall, M. G. and Stuart, A., The Advanced Theory of Statistics, Vol. I, Hafner, New York, (1958).

Kendall, M. G., The Advanced Theory of Statistics, Vol. I, London, (1943), p. 302.

Lüders, R., "Die Statistik der seltenen Ereignisse," Biometrika, 26 (1934), pp. 108-128.

McKendrick, A. G., "Studies on the Theory of Continuous Probabilities with Special Reference to its Bearing on Natural Phenomena of a Progressive Nature," Proc. London Math. Soc. (2), 13 (1914), pp. 401-416.

Patil, G. P., "On Evaluation of Negative Binomial Distribution Functions," Ann. Math. Stat., 31 (1960), p. 527.

Pearson, K., and Fieller, E. C., "On the Application of the Double Bessel Function to Statistical Problems," Biometrika, 25 (1933), pp. 158-178.

Pearson, Karl, Tables of the Incomplete Beta-Function, Cambridge University Press, (1934).

Quenouille, M. H., "A Relation Between the Logarithmic, Poisson, and Negative Binomial Series," Biometrics, 5 (1949), pp. 162-164.

Sampford, M. R., "The Truncated Negative Binomial Distribution," Biometrika, 42 (1955), pp. 58-69.

Sickel, H. S., "The Estimation of the Parameters of a Negative Binomial Distribution with Special Reference to Psychological Data," Psychometrika, 16 (1951), pp. 107-127.

Steyn, H. S., "On X^2-Tests for Contingency Tables of Negative Multinomial Types," Statistica Neerlandica, 13 (1959), pp. 433-444.

Steyn, H. S., "On Discrete Multivariate Probability Functions of Hypergeometric Type," Koninkl. Nederl. Akademie van Watenschappen, A, 58 (1955), pp. 588-595.

Terry, Milton E., Personal Communication.

Todhunter, I., A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace. Cambridge and London, MacMillan and Co., (1865).

Yule, G. U., "A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis," Phil. Trans. Roy. Soc., London, B, 213 (1924), pp. 21-87.

XI. RELATED REFERENCES

Aitken, A. C., and Gonin, H. T., "On Fourfold Sampling with and Without Replacement," Proc. Roy. Soc. Edin., 55 (1935), pp. 114-125.

Anscombe, F. J., "The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution," Biometrics, 5 (1949), pp. 165-173.

Anscombe, F. J., "The Transformation of Poisson, Binomial and Negative-Binomial Data," Biometrika, 35 (1948), pp. 246-254.

Arbous, A. G., and Kerrick, J. E., "Accident Statistics and the Concept of Accident-Proness," Biometrics, 7 (1951) pp. 340-432.

Bartlett, M. S., "The Square Root Transformation in Analysis of Variance," Jour. Roy. Stat. Soc., B, 3 (1936), pp. 68-78.

Bartlett, M. S., "The Use of Transformations," Biometrics, 3 (1947), pp. 39-52.

Beall, G., and Rescia, R. R., "A Generalization of Neyman's Contagious Distributions," Biometrics, 9 (1953) pp. 354-386.

Beall, G., "The Transformation of Data from Entomological Field Experiments so that the Analysis of Variance Becomes Applicable," Biometrika, 32 (1941-42), pp. 243-262.

Bliss, C. I., "Fitting the Negative Binomial Distribution to Biological Data," Biometrics, 9 (1953), pp. 176-196.

Bliss, C. I., and Owen, A. R. G., "Negative Binomial Distributions with a Common k," Biometrika, 45 (1958), pp. 37-58.

Blom, G., "Transformations of the Binomial, Negative Binomial, Poisson and χ^2 Distributions," Biometrika, 41 (1954), pp. 302-316.

Blom, G., Corrigenda to, "Transformations of the Binomial, Negative Binomial, Poisson and χ^2 Distributions," Biometrika, 43 (1956), p. 235.

Brockmeyer, E., Halstrøm, H. L., and Jensen, A., "The Life and Works of A. K. Erlang," Trans. Dan. Acad. Techni. Sci., No. 2, (1948).

Cernuschi, F. and Castagnetto, L., "Chaines of Rare Events," Ann. Math. Stat., 17 (1946), pp. 53-61.

Chapman, D. G., "The Estimation of Biological Populations," Ann. Math. Stat., 25 (1954), pp. 1-15.

Cochran, W. G., "Some Methods for Strengthening the Common χ^2 Tests," Biometrics, 10 (1954), pp. 417-451.

Cramér, H., Mathematical Methods of Statistics. Princeton University Press, (1946).

David, F. N., and Johnson, N. L. "The Truncated Poisson," Biometrics 8 (1952), pp. 275-285.

DeGroot, M. H., "Unbiased Sequential Estimation for Binomial Populations," Ann. Math. Stat. 30 (1959), pp. 80-101.

Feller, W., "On a General Class of Contagious Distributions," Ann. Math. Stat., 14 (1943), pp. 389-400.

Fitzpatrick, R., "The Detection of Individual Differences in Accident Susceptibility," Biometrics, 14 (1958), pp. 50-68.

Goodman, L. A., "Sequential Sampling Tagging for Population Size Problems," Ann. Math. Stat., 24 (1953), pp. 56-69.

Greenwood, M., "Accident Proneness," Biometrika, 37 (1950), pp. 24-29.

Gurland, J., "Some Interrelations Among Compound and Generalized Distributions," Biometrika, 44 (1957), pp. 265-268.

Hartley, H. O., "Maximum Likelihood Estimation from Incomplete Data," Biometrics, 14 (1958), pp. 174-194.

Irwin, J. O., "A Distribution Arising in the Study of Infectious Diseases," Biometrika, 41 (1954), pp. 266-268.

Kendall, D. G., "On Some Modes of Population Growth Leading to R.A. Fisher's Logarithmic Series Distribution," Biometrika, 35 (1948), pp. 6-15.

Lindley, D. V., "Binomial Sampling Schemes and the Concept of Information," Biometrika, 44 (1957), pp. 179-186.

McGuire, J. U., Brindley, T. A., and Bancroft, T. A., "The Distribution of European Corn Borer Larvae Pyrausta Nubilalis (HBN.), in Field Corn," Biometrics, 13 (1957), pp. 65-78.

Mood, A. M., Introduction to the Theory of Statistics. McGraw-Hill, New York, (1950).

Murty, V. N., "A Note on Bhattacharyya Bounds for the Negative Binomial Distribution," Ann. Math. Stat., 27 (1956), pp. 1182-1183.

Newbold, E. M., "Practical Applications of the Statistics of Repeated Events, Particularly to Industrial Accidents," Jour. Roy. Stat. Soc., 90 (1927), pp. 487-547.

Oakland, G. B., "An Application of Sequential Analysis to Whitefish Sampling," Biometrics, 6 (1950), pp. 59-67.

Olmstead, P. S., "Note on Theoretical and Observed Distributions of Repetitive Occurrences," Ann. Math. Stat., 11 (1940), pp. 363-366.

Rao, C. R., Advanced Statistical Methods in Biometric Research. Wiley, New York, (1952).

Rider, P. R., "Truncated Binomial and Negative Binomial Distributions," J. Amer. Stat. Assn., 50 (1955), pp. 877-883.

Rutherford, R. S. G., "On a Contagious Distribution," Ann. Math. Stat., 25 (1954), pp. 703-713.

Sandelius, M., "An Inverse Sampling Procedure for Bacterial Plate Counts," Biometrics, 6 (1950), pp. 291-292.

Shellam, J. G., "A Probability Distribution Derived from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials," Jour. Roy. Stat. Soc., B, 10 (1948), pp. 257-261.

Steyn, H. S., "On Discrete Multivariate Probability Functions," Koninkl. Nederl. Akademie van Wetenschappen, A, 54 (1951) pp. 23-30.

Tweedie, M. C. K., "Inverse Statistical Variates," Nature, 155 (1945), p. 453.

Tweedie, M. C. K., "The Estimation of Parameters from Sequentially Sampled Data on a Discrete Distribution," Jour. Roy. Stat. Soc., B 14 (1952), pp. 238-245.

Walsh, J. E., "Approximate Probability Values for Observed Number of Successes from Statistically Independent Binomial Events with Unequal Probabilities," Sankhyā, 15 (1955), pp. 281-290.

Wiid, A. J. B., "On the Moments and Regression Equations of the Fourfold Negative and Fourfold Negative Factorial Binomial Distributions," Proc. Roy. Soc. Edin., A. 65 (1957-58), pp. 29-34.

Wise, M. E., "The Use of the Negative Binomial Distribution in an Industrial Sampling Problem," Jour. Roy. Stat. Soc, B, 8 (1948), pp. 202-211.

Wishart, J., "Cumulants of Multivariate Multinomial Distributions," Biometrika, 36 (1949), pp. 47-58.

The vita has been removed from the scanned document

ABSTRACT

This thesis is an extensive review of the major literature dealing with the negative binomial distribution. An account of the forms of the distribution is presented. This includes compound, limiting and truncated forms and examples of their occurrence. Parameter estimation is treated in a few specific cases.

In particular, the inverse binomial sampling form, its analytic properties, occurrence in biological as well as industrial situations, estimation of the parameter p, and hypothesis testing are discussed in detail. Many examples are included.

In the discussion of the inverse binomial sampling form several extremely useful, little known relations are presented. Among these is a discussion of the use of the *Tables of the Incomplete Beta Function* for the evaluation of the cumulative distribution function. Also discussed is the method of adapting *Biometrika Table 41*, which gives confidence limits on p in the positive binomial case, to the inverse binomial sampling case.