FITTING NEGATIVE BINOMIAL DISTRIBUTIONS BY THE METHOD OF MAXIMUM LIKELIHOOD

BY

LeROY J. SIMON

I—INTRODUCTION

Maximum likelihood solutions for negative binomial distributions have been worked out by a number of authors. The purpose of this paper will be to develop the solutions in an insurance context, and to investigate one phase that has not been touched upon in the literature. The formulas developed will be applied to some actual data.

Dropkin[1] has considered the process of fitting the negative binomial distribution by the method of moments, to a set of complete data. In this paper, the same problem will be treated using the method of maximum likelihood; but first, two problems will be solved where the number of observations in the zero case (claim-free insureds in insurance applications) is subject to some special condition.

One type of a special condition would be if the zero cases were suspected of having been censored in some manner. This censoring might arise because claim-free policy files were destroyed if they did not renew, while all other files were retained. Another example with a similar distortion in the zero case was considered by Harwayne[2] when it seemed likely that a number of zero cases would appear in the records of the California Motor Vehicle Department, for persons who did not actually drive at all in the state during the period covered by the study.

We will also consider the special condition of a truncated negative binomial. This will often arise in insurance applications because it is usually much easier to locate and study those policies which had one or more claims during the experience period, rather than checking the entire policy file. If a study is made of only the policies which had claims, we get a truncated distribution, where the zero case has been entirely eliminated.

Finally, it may be in order to comment on how the method of maximum likelihood compares with the method of moments. The method of moments uses as many moments of the distribution as are necessary to obtain a solution. Many of the mathematical models that we use are described by one or two parameters. Hence, one or two moments are sufficient for a solution. In an occasional problem, we may find that the third moment must be utilized. When the third moment is introduced, large sampling errors result, and the fits are not too satisfactory.

The method of maximum likelihood is based on the principle that the best estimate of the population parameters is that estimate which maximizes the

[1] Lester B. Dropkin, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," CAS, XLVI, p. 166 (1959).

[2] Frank Harwayne, "Merit Rating in Private Passenger Automobile Liability Insurance and the California Driver Record Study," CAS, XLVI, p. 192 (1959).

probability of obtaining the observed sample. This method avoids the use of moments, but often requires difficult and extensive calculations. It will usually produce answers which are very similar to the method of moments, if second-order moments are the highest needed for a solution by this latter method. However, when higher-order moments are introduced, the reduced sampling errors of parameters estimated by the maximum likelihood method easily off-set the calculating difficulty.[3]

II — MATHEMATICAL DEVELOPMENT

The form for dealing with the negative binomial which has been used in recent volumes of our Proceedings,[4] is cumbersome to manipulate in the manner desired in this paper; so, instead of using

$$f(x) = \left(\frac{a}{a+1}\right)^r \left(\frac{-1}{a+1}\right)^x \binom{-r}{x} \tag{1}$$

let's define

$$p = \frac{1}{a} \tag{2}$$

$$q = 1+p = \frac{a+1}{a} \tag{3}$$

and rewrite the binomial coefficient to produce

$$f(x) = \binom{r+x-1}{x} p^x q^{-r-x} \tag{4}$$

First we will treat the special condition where the zero case is subject to some distortion. Let $(1 + \theta)$ be the measure of distortion in the zero case. Then the probability of observing x claims on a policy will be given by

$$\left.\begin{aligned} f(x) &= q^{-r}(1+\theta)/(1+\theta q^{-r}) & x &= 0 \\ f(x) &= \binom{r+x-1}{x} p^x q^{-r-x}/(1+\theta q^{-r}) & x &= 1, 2, \ldots \end{aligned}\right\} \tag{5}$$

The denominators on the right side of (5) are necessary so that $\sum f(x) = 1$, thus making the total probability equal unity. Let n_0 be the number of sample cases in which $x = 0$ and let N be the total of all cases in the sample. The likelihood function for such a sample is:

$$P(x_1, \cdots, x_N;\, q, \theta, r) =$$
$$q^{-n_0 r}(1+\theta)^{n_0}(1+\theta q^{-r})^{-n_0} \prod_{x>0} \binom{r+x-1}{x} p^x q^{-r-x} (1+\theta q^{-r})^{-1} \tag{6}$$

The maximum likelihood solution is obtained by taking the logarithm of (6),

[3] F. N. David and N. L. Johnson, "The Truncated Poisson," *Biometrics*, VIII, pp. 275-85 (1952). On page 284 they illustrate that the standard error of the parameter p determined by the method of moments is eight times the standard error of p determined by the method of maximum likelihood.

[4] CAS, XLVI, p. 166 (1959) and CAS, XLVII, p. 1, p. 20, p. 37 and p. 55 (1960).

differentiating partially with respect to each parameter and setting the results equal to zero:

$$L = n_0 \log (1+\theta) + \sum_{x>0} \log \binom{r+x-1}{x} + \log p \sum x - Nr \log q$$
$$- \log q \sum x - N \log (1 + \theta q^{-r}) \tag{7}$$

$$\frac{\partial L}{\partial \theta} = \frac{n_0}{1+\theta} - \frac{N q^{-r}}{1+\theta q^{-r}} = 0 \tag{8}$$

$$\frac{\partial L}{\partial q} = \frac{\sum x}{p} - \frac{Nr}{q} - \frac{\sum x}{q} + \frac{N \theta r q^{-r-1}}{1+\theta q^{-r}}$$
$$= \frac{\sum x}{pq} - \frac{Nr}{q(1+\theta q^{-r})} = 0 \tag{9}$$

The partial derivative of L with respect to r is less easily obtained,[5] but eventually leads to

$$\frac{\partial L}{\partial r} = \sum_{x>0} \left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1} \right) - N \log q + \frac{N \theta q^{-r} \log q}{1+\theta q^{-r}}$$
$$= \sum_{x>0} \left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1} \right) - \frac{N \log q}{1+\theta q^{-r}} = 0 \tag{10}$$

Solving (8) for θ we have

$$\theta = (n_0 q^r - N)/(N - n_0) \tag{11}$$

Substituting (11) in (9) and (10):

$$\frac{\sum x}{pq} - \frac{(N-n_0) r}{q(1-q^{-r})} = 0 \tag{12}$$

$$\sum_{x>0} \left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1} \right) - \frac{(N-n_0) \log q}{1-q^{-r}} = 0 \tag{13}$$

---

[5] It is helpful in the development of this partial derivative to know that:

$$\frac{\partial}{\partial r} \left[\log \binom{r+x-1}{x} \right] = \frac{\partial}{\partial r} \left[\log \frac{(r+x-1)!}{x!(r-1)!} \right]$$
$$= \frac{\partial}{\partial r} \left[\sum_{i=0} \log (r+x-1-i) - \sum_{i=0} \log (x-i) - \sum_{i=0} \log (r-1-i) \right]$$
$$= \sum_{i=0} \frac{1}{r+x-1-i} - \sum_{i=0} \frac{1}{r-1-i}$$
$$= \frac{1}{r+x-1} + \frac{1}{r+x-2} + \cdots + \frac{1}{r+1} + \frac{1}{r}$$

These equations cannot be solved directly and the values of q and r must be found by some iterative process. The difficulty of solving the maximum likelihood equations is the principal deterrent to their wide-spread use.

The second special condition that we will study is the case where $\theta = -1$ in (5); i.e., we have no measurement of the zero case. This is the situation in insurance where a study is made only of those policies which have claims on them. In this case no measurement is made of n_0, so let N' represent the total number of cases in the truncated sample in which n_0 is missing. Then,

$$f(x) = \binom{r+x-1}{x} p^x q^{-r-x}/(1-q^{-r}) \qquad x = 1,2\ldots \tag{14}$$

$$P(x_1, \cdots, x_N; q, r) = \prod_{x>0} \binom{r+x-1}{x} p^x q^{-r-x} (1-q^{-r})^{-1} \tag{15}$$

$$L = \sum_{x>0} \log \binom{r+x-1}{x} + \log p \left(\sum x\right) - N' r \log q - \log q \sum x$$
$$- N' \log (1 - q^{-r}) \tag{16}$$

$$\frac{\partial L}{\partial q} = \frac{\sum x}{pq} - \frac{N'r}{q(1-q^{-r})} = 0 \tag{17}$$

$$\frac{\partial L}{\partial r} = \sum_{x>0}\left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1}\right) - \frac{N' \log q}{1-q^{-r}} = 0 \tag{18}$$

Comparing (17) with (12) and (18) with (13), the two are identical when we recall that N' and $(N - n_0)$ are the same thing; viz., the number of cases in the sample with the zero case excluded. This means that the two cases we have considered thus far are identical and if we have evidence that the zero case has been subject to distortion, we might as well throw it out and fit the curve as though our sample had the zero case censored out completely. This is not a surprising result since we must lose one degree of freedom in either event, so it makes little difference if we do it by adjusting an imperfect measure of n_0 or by entirely manufacturing an n_0.

For use in Section III—Application, we will now develop the formulas for fitting a truncated negative binomial by the method of moments. It has been shown[6] that the first three moments of the negative binomial are, in our notation:

$$\mu'_1 = rp \tag{19}$$

$$\mu'_2 = rp\,(q + rp) \tag{20}$$

$$\mu'_3 = rp\,(r^2p^2 + 3rp^2 + 3rp + 2p^2 + 3p + 1) \tag{21}$$

Notice that

$$\sum_{x>0} x^i = \sum_{x\geq 0} x^i \text{ for any } i > 0.$$

Since, $\sum x^2 / \sum x$ from the truncated sample will be an estimate of μ'_2/μ'_1, and

[6] LeRoy J. Simon, "The Negative Binomial and Poisson Distributions Compared," *CAS*, XLVII, p. 20 (1960).

since $\sum x^3/\sum x$ from the truncated sample will be an estimate of μ'_3/μ'_1 we can set $\sum x^2/\sum x = rp\ (q + rp)/rp$ and (22)

$$\sum x^3/\sum x = rp\ (r^2p^2 + 3rp^2 + 3rp + 2p^2 + 3p + 1)/rp \tag{23}$$

Solve these two equations for r and p and get

$$1 + p = q = \frac{(\sum x)\ (\sum x^3) - (\sum x^2)^2}{(\sum x)\ (\sum x^2 - \sum x)} \tag{24}$$

$$r = \frac{2\ (\sum x^2)^2 - (\sum x^2)\ (\sum x) - (\sum x^3)\ (\sum x)}{(\sum x)^2 + (\sum x)\ (\sum x^3) - (\sum x^2)^2 - (\sum x^2)\ (\sum x)} \tag{25}$$

With q and r determined from either a maximum likelihood solution or by the method of moments, N can be calculated from

$$N = N' + n_0 = N' + Nq^{-r}$$

$$N = N'/(1 - q^{-r}) \tag{26}$$

The fitted curve is thus determined.

Let us now consider the problem of fitting a complete negative binomial distribution by the method of maximum likelihood. Equation (4) gives the probability distribution of x claims where $x = 0, 1, 2 \ldots$ The likelihood function gives $P\ (x_1, x_2, \cdots, x_N; q, r) = \prod_x \binom{r+x-1}{x} p^x q^{-r-x}$ (27)

Then,

$$L = \sum_x \log \binom{r+x-1}{x} + \log p \sum x - Nr \log q - \log q \sum x \tag{28}$$

$$\frac{\partial L}{\partial q} = \frac{\sum x}{pq} - \frac{Nr}{q} = 0 \tag{29}$$

$$\frac{\partial L}{\partial r} = \sum_{x>0}\left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1}\right) - N \log q = 0 \tag{30}$$

Solve (29) for q and substitute in (30):

$$q = \frac{\sum x}{Nr} + 1 \tag{31}$$

$$\sum_{x>0}\left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1}\right) - N \log\left(1 + \frac{\sum x}{Nr}\right) = 0 \tag{32}$$

Here again we are unable to get a direct solution for p. However Equation (32) has only one unknown and, therefore, can be solved by a trial and error process.

III — APPLICATION

To illustrate the application of the various formulas, refer to Table 1. The data in the first two columns is taken from a study by Blensley and Head.[7] Column (3) utilizes Dropkin's formulas[8] which are, in our notation and slightly rewritten:

$$q = \frac{N \sum x^2 - (\sum x)^2}{N \sum x} \tag{33}$$

$$r = \frac{(\sum x)^2}{N \sum x^2 - (\sum x)^2 - N \sum x} \tag{34}$$

Column (4) results from a solution of equation (32) by trial and error and substitution of that value in (31). The χ^2 tests on columns (3) and (4) indicate that the negative binomial is a good fit and illustrate that the two methods produce very similar answers when only the first two moments are used.

Column (5) comes from a solution of equations (24) and (25). It fits closely as indicated by the χ^2 value.

Column (6) was difficult to obtain because equations (17) and (18) must be solved simultaneously. Various methods were attempted, but the method suggested by David and Johnson[9] was modified slightly, and used. From (17) and (18) we have

$$\frac{\sum x}{r} = \frac{N' (q-1)}{(1-q^{-r})} \tag{35}$$

$$\sum_{x>0} \left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1} \right) = \frac{N' \log q}{1-q^{-r}} \tag{36}$$

Then (36)/(35) gives

$$\frac{r \sum\limits_{x>0} \left(\frac{1}{r} + \frac{1}{r+1} + \cdots + \frac{1}{r+x-1} \right)}{\sum x} = \frac{\log q}{q-1} \tag{37}$$

The procedure then is to select a starting value of r and evaluate the left hand side of (37). Enter the center of Table 2 with this value and read out the corresponding value of q. Substitute this value together with the estimated r in

$$(\sum x)\ (1-q^{-r})/N'\ (q-1) \tag{38}$$

thus producing an improved estimate of r. Repeat the process until q and r become stable. Experience indicates that convergence is quite slow and much

[7] R. C. Blensley and J. A. Head, "Statistical Determination of Effect of Paved Shoulder Width on Traffic Accident Frequency," *Highway Research Board Bulletin*, CCXL, p. 4 (1959). A sample element is defined in the study to be a one mile section of level and tangent primary rural two-lane Oregon highway with lane width of 10′ or more, which had paved shoulders and 30% or less sight restriction taken for a one-year period.

[8] Dropkin, p. 166.

[9] David and Johnson, p. 284.

can be gained by "leap frogging" ahead in the direction indicated by the improved estimate of r. Care must also be exercised in carrying sufficient significant digits, especially when getting the value of q, otherwise false indications will be given on where the final solution lies. In the example used, it was found advisable to get q to three, or even four, decimal places.

Column (6) fits more closely than column (5) as indicated by the sharply reduced value of χ^2. Despite the calculating complexity, it would be the method to use in practice. Just in case anyone is tempted to compare columns (3) and (4) with column (5) and conclude that this disproves my contention that the maximum likelihood method is better, I suggest the reader change n_0 in column (2) from 99 to 68 and re-calculate column (3). The exercise will be revealing.

IV — CONCLUSION

If some distortion is known or suspected in the zero case, we conclude that we might as well discard the observations in the zero case and deal only with the remaining data.

A second major conclusion is that the method of moments and the maximum likelihood method produce essentially the same result when used for fitting complete curves with a negative binomial distribution. Therefore, the method of moments would be used in practice because of the ease in calculation; and equations (33) and (34) would be used to determine q and r.

For a truncated distribution the method of maximum likelihood gives substantially improved results, and therefore it is recommended despite the calculating difficulty. Equations (37) and (38) would be used to obtain successively improved approximations to q and r.

After dealing with a variety of sample distributions, it seems to the author that the negative binomial has a great deal of plasticity and will conform well to a great variety of empirical data.

TABLE 1

Distribution of Sample Elements by Number of Accidents*
and Negative Binomial Curves Fitted by Various Methods

(1)	(2)	(3)	(4)	(5)	(6)
		Calculated Frequencies of Negative Binomials			
Number of Accidents	*Number of Sample Elements*	*Method of Moments Regular*	*Maximum Likelihood Regular*	*Method of Moments Truncated*	*Maximum Likelihood Truncated*
x	f_0	f_t	f_t	f_t	f_t
0	99	95.3	95.8	—	—
1	65	76.1	75.9	74.7	69.0
2	57	50.6	50.4	49.7	51.1
3	35	31.4	31.3	30.9	33.2
4	20	18.8	18.8	18.5	20.1
5	10	11.0	11.0	10.8	11.6
6	4	6.4	6.4	6.3	6.5
7	0				
8	3				
9	4 } 8	8.4	8.4	8.1	7.5
10	0				
11	1				
Total	298	298.0	298.0	199.0	199.0
χ^2		4.07	4.06	3.98	2.21
d.f.		5	5	4	4
Probability		.55	.55	.40	.70
r		1.4974	1.476	1.4983	2.1610
q		2.1407	2.157	2.1402	1.8817
Equations used		(33) & (34)	(31) & (32)	(24) & (25)	(37) & (38)

Summary Statistics

$N = 298$	$\sum x^2 = 1959$	$N' = 199$
$\sum x = 509$	$\sum x^3 = 10643$	$x = 1.70805$
		$\sigma^2 = 3.65638$

* See Footnote 7 in text.

TABLE 2

Value of log q/(q-1)

q	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	Proportional Parts 1	2	3	4	5
1.0	—	.9950	.9900	.9853	.9805	.9758	.9712	.9666	.9620	.9576	5	10	14	19	24
1.1	.9531	.9487	.9444	.9402	.9359	.9317	.9276	.9235	.9195	.9155	4	8	12	17	21
1.2	.9116	.9077	.9039	.9000	.8963	.8926	.8889	.8853	.8816	.8781	4	8	11	15	19
1.3	.8745	.8711	.8676	.8642	.8608	.8574	.8541	.8508	.8476	.8444	3	7	10	13	17
1.4	.8412	.8380	.8349	.8318	.8287	.8257	.8227	.8197	.8168	.8138	3	6	9	12	15
1.5	.8109	.8081	.8052	.8024	.7996	.7968	.7941	.7914	.7887	.7860	3	6	8	11	14
1.6	.7833	.7807	.7781	.7755	.7730	.7704	.7679	.7654	.7629	.7605	3	5	8	10	13
1.7	.7580	.7556	.7532	.7508	.7485	.7462	.7438	.7415	.7392	.7370	2	5	7	9	12
1.8	.7347	.7325	.7303	.7281	.7259	.7238	.7216	.7195	.7174	.7153	2	4	6	9	11
1.9	.7132	.7111	.7091	.7070	.7050	.7030	.7010	.6990	.6970	.6951	2	4	6	8	10
	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9					
2.	.6932	.6745	.6571	.6407	.6253	.6109	.5972	.5843	.5720	.5604					
3.	.5493	.5388	.5287	.5191	.5099	.5011	.4927	.4846	.4768	.4693					
4.	.4621	.4552	.4485	.4420	.4358	.4297	.4239	.4183	.4128	.4075					
5.	.4024	.3974	.3925	.3878	.3833	.3788	.3745	.3703	.3662	.3622					
6.	.3584	.3546	.3509	.3473	.3438	.3403	.3370	.3337	.3305	.3274					
7.	.3243	.3213	.3184	.3155	.3127	.3100	.3073	.3047	.3021	.2995					

PROCEEDINGS

NOVEMBER 15-17, 1961

PRESIDENTIAL ADDRESS BY WILLIAM LESLIE, JR.

Once again I have the honor to address you as has been the custom for each President to do. This occasion marks the completion of my second term and I would not leave office without again thanking you for the honor. I can report to my successor that in several senses the Society seems more healthy today than it did two years ago. The financial position of the Society has been greatly enhanced by the introduction of the Invitational Program and the adoption of a schedule of registration fees in connection with the semi-annual meetings. Largely due to our improved financial position the Council of the Society has authorized the employment of a Secretary-Treasurer on what most of us hope will be a long term basis.

I once again call your attention to the increase in not only the number of papers presented to us but in the very substantial upgrading of their content. The work now being done by many of our members in the area of risk theory, for example, a matter to which I will refer later in this address, will, I am sure, bring considerable added lustre to the Society.

There seems little doubt that in some measure due to the cohesion produced by the Society's standards and status the influence of many of our members in the insurance business is steadily increasing.

One of the more intriguing but as yet unfinished projects which will surely absorb some of the time of your next President will be the furthering of the already substantial efforts to bring about closer coordination and cooperation among the several actuarial bodies and quite specifically between the Casualty Actuarial Society and the Society of Actuaries.

In today's address I wish to touch on a subject about which one reads many comments being made by insurance industry spokesmen with great frequency. The word "chaos" or its equivalent is being used over and over again to describe one or more problems today facing the insurance industry. Responsible executives with several decades of experience behind them are reporting that today's conditions represent the need to solve problems the like of which they have not seen previously in their careers.

There seem many aspects to this report of chaotic conditions. We hear it in discussions of the problems of independent companies viz a viz rating