NEGATIVE BINOMIAL RATIONALE

BY

THOMAS O. CARLSON

"If thou art wise thou knowest thine own ignorance"—Luther

To one who did his statistical teething in what I would term the era of the 'Lexis Theory', the arrival of more searching tools upon the actuarial scene through the notable contributions of Harwayne, Dropkin, Simon, Hewitt, Bailey, Roberts and others of a younger generation, which has been likened by others to an invasion from Mars, is more accurately analogous to the keg-ling that awoke Rip Van Winkle. When, at our meeting last spring, I felt called upon to defend vigorously this 'New Look' of our Proceedings, a closer survey of these frontiers appeared to be in order.

Because I still have very strongly the childhood instinct to look into the "why" of everything in my experience, my attention has focused principally upon the rationale underlying the utilization of the negative binomial distribution in actuarial analysis. We are all interested in finding tools that work. But we should not be satisfied as actuaries without probing into any unfamiliar mathematical model until we know why it works, because thus only do we learn whether it is the best model for the purpose or whether it can be improved upon, and also what extensions of its utility may be available. The arbitrary use of the Pearson Type III distribution in the derivations of the negative binomial presented to us in detail in this body raises questions in many minds not answered by the excellent, yet to me too brief, paper presented last spring by Mr. Simon ("An Introduction to the Negative Binomial Distribution and its Applications"). I should like therefore to take the time to review that paper and the pioneering introduction to this valuable tool presented by Mr. Dropkin ("Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," PCAS XLVI, p. 165) so as to extract therefrom and interpret the material relating to rationale; and then I shall present new material not yet considered in our Proceedings which I think casts important light on the investigation.

The most frequent derivation, and the one with which we have here become familiar, stems from the assumption of a variability in the accident-expectancy from individual to individual in a statistical population, with such variability following the Pearson Type III distribution, but with the distribution for each value of the accident expectancy so determined following the Poisson. It is natural to ask: "Why Pearson Type III?" The answer, as given by Dropkin, is threefold:

a) it is a skew distribution,
b) it leads to conveniently "simple" (sic!) equations, and
c) there are tables (of the Incomplete Gamma Function) available for use in the resulting evaluation.

To this rationale must be added a fourth consideration: d) it works better than its predecessor formula, the Poisson distribution.

The skewness of the Pearson Type III formula satisfies our intuitive knowledge of accident proneness in individuals. As respects the simplicity of the resulting equations, this is due in large measure to the fact that the number of parameters is only two, which is not the case in the non-degenerate Pearson forms; thus we might add a fifth reason to our rationale which can be considered a subreason partially explanatory of b) above: e) the total number of parameters is held to two.

But all this still leaves the rationale of the Pearson Type III utilization upon an empirical basis. It leaves our curiosity unsatisfied. Let us turn therefore to other derivations of the final formula, so well summarized in Simon's paper already referred to. He mentions a number of derivations not utilizing the Pearson Type III assumption. The historically original derivation by Yule in 1910 develops, to quote Simon, "the distribution of the number of fatalities that would occur during the n^{th} exposure to a disease in excess of r exposures where r is the minimum number which will be fatal and the effects of repeated attacks of the disease act cumulatively." In terms of accidents this could read: the distribution of the number of accidents occurring during the n^{th} exposure in excess of r exposures where r is the minimum number to produce an accident and the successive exposures to accident act cumulatively; i.e., the probability that $r + n$ exposures to accident are required for the occurrence of n accidents.

Let us examine this derivation briefly. (Wilks: Mathematical Statistics has perhaps the clearest presentation.)

Let P (success) $= \frac{1}{q}$, P (failure) $= \frac{p}{q}$, i.e., $q - p = 1$, since $\frac{1}{q} + \frac{p}{q} = 1$.

The probability of $n - 1$ successes and r failures in $r + n - 1$ trials is

$$\frac{(r + n - 1)!}{(n - 1)!\, r!} \cdot \left(\frac{1}{q}\right)^{n-1} \cdot \left(\frac{p}{q}\right)^{r}$$

and if we multiply by $\frac{1}{q}$ we have the probability that $r + n$ trials are required for n successes, or

$$\frac{(r + n - 1)!}{(n - 1)!\, r!} \cdot \left(\frac{1}{q}\right)^{n} \cdot \left(\frac{p}{q}\right)^{r} \qquad (1)$$

which is the general term in the expansion of

$$(q - p)^{-n}$$

This form incidentally indicates most clearly why the terminology "negative binomial" is commonly attached to this distribution. If we substitute $P = \frac{1}{a}$ in (1) we have

$$\frac{a^{n}}{r!\, \Gamma(n)} \cdot \frac{\Gamma(r + n)}{(1 + a)^{r+n}} \qquad (1a)$$

which is the form presented in Dropkin's original paper on the negative binomial (his formula (7)).

It is not unreasonable to conjecture that Yule in 1920, in his collaboration with Greenwood, recognized the Pearson Type III expression upon differentiation of the formula developed by him in 1910, and backtracked thence to the now familiar and most common derivation by assuming a Pearson Type III distribution for the Poisson parameter, and deduced therefrom the interpretation as respects the variation of accident expectation (or accident proneness) among the individuals exposed to accident hazard.

Simon's first model (in the paper already referred to) illustrates the Yule 1910 approach, his second model the 1920 approach of Yule and Greenwood.

The analogy between the Yule 1910 development and the Markov-chain approach of Bartlett, under which the chance of an additional accident depends upon the number which have already ocurred, is apparent, explaining perhaps more clearly, however, the use of the negative binomial (e.g., by Polya and Eggenberger in Zeitschrift f. agen. Math. u. Mech. III, p. 279, 1933) for determining the probability of x cases occurring in an epidemic, the so-called "contagion function" use. It may be noted that the two Kendall derivations referred to but not elaborated by Simon are those of 1) Yule and Greenwood in 1920 and 2) Yule in 1910.

This brings us to a derivation and an interpretation not yet discussed in these Proceedings. Simon has noted that Arthur L. Bailey used the negative binomial in the Proceedings in 1950; Mr. Bailey's source was the "Theory of Probabilities" by Jeffreys which we had both picked up the year before and discussed together, and which is still a most informative reference on the subject. Jeffreys develops a distribution of the number of claims on the basis of certain assumptions connecting the distributions of multiple-claim accidents, and produces a negative binomial distribution for the number of claims starting with a Poisson distribution for each group of n-claim accidents. A more general development along these lines, of which Jeffreys' is only a particular case, is to be found in an article by R. Lüders, in German, in Biometrika ("Die Statistik der Seltenen Ereignisse," Biometrika, Volume 26, p. 180, 1934). Both of these references could well be added to Simon's bibliography presented to us last spring.

The rationale of Lüders' development, which should be of particular interest to actuaries, is predicated upon the assumption that single-claim accidents as a group follow the Poisson distribution, as does the group of two-claim accidents, the group of three-claim accidents, and so on. The development initially assumes that these respective Poisson distributions are independent; but this complex multiple Poisson distribution of the number of claims reduces to the negative binomial distribution when the parameters of the independent distributions are reduced to two by making them interdependent through the assumed relationship

$$a_k = \frac{a_1}{k} \cdot b^{k-1} \tag{2}$$

a_1 and a_k being the parameters of the accident distributions involving respectively a single claim or k claims in an accident. In other words, the negative binomial here provides a distribution of claims corresponding to a Poisson distribution of accidents with the expectations of an accident involving 1, 2, 3, ... claims inter-connected by the modified power-series relation (2).

Since the development involves some interesting by-products on the way I shall indicate it as briefly as practicable.

$$\text{Let } P_{x_m} = \frac{e^{-a_m} \cdot a_m^{x_m}}{x_m!} \qquad (m = 1,2,3,\ldots) \tag{3}$$

represent the probability that exactly x_m accidents with m claims associated with each will occur.

Let $r = x_1 + 2x_2 + 3x_3 + \ldots.$

represent the number of claims.

Then the probability of exactly r claims occurring, assuming that the respective simple Poisson distributions (3) are independent is

$$P(r) = e^{-a_1 - a_2 - \ldots\ldots} \cdot \sum_{r = x_1 + 2x_2 + \ldots} \frac{a_1^{x_1} \cdot a_2^{x_2} \ldots.}{x_1!\, x_2! \ldots.} \tag{4}$$

Since this is a general formula that assumes that the occurrence of single-claim accidents is independent of the occurrence of two claim accidents, and so on, there is developed below an evaluation of the first three moments, which will be of use later. The factorial-moment generating function is

$$f(z) = \sum_r z^r\, P\,(r)$$

$$= e^{-a_1 - a_2 - \ldots} \cdot e^{a_1 z + a_2 z^2 + \ldots\ldots} \tag{5}$$

It is immediately obvious that

$$\sum_r P(r) = f(1) = 1$$

If we set

$$A(z) = \log f(z) = -a_1 - a_2 - \ldots. + a_1 z + a_2 z^2 + \ldots \tag{6}$$

then $f(z) = e^{A(z)}$

and $f'(1) = \sum_r r \cdot P(r) = a_1 + 2a_2 + 3a_3 + \ldots. = \text{mean}$ (7a)

By further differentiation and the use of formulas relating factorial moments with ordinary moments (see, for example, Korn and Korn: Mathematical Handbook, 18.3-10.), we find that

$$\mu_2 = a_1 + 2^2 a_2 + 3^2 a_3 + \ldots\ldots = \text{variance} \tag{7b}$$

$$\mu_3 = a_1 + 2^3 a_2 + 3^3 a_3 + \ldots\ldots = \text{3rd moment about mean} \tag{7c}$$

Now let us reduce the number of parameters to two by use of the relation (2), setting $a = a_1$:

$$A(z) = -a\left(1 + \frac{b}{2} + \frac{b^2}{3} + \ldots\right) + a\left(z + \frac{bz^2}{2} + \frac{b^2 z^3}{3} + \ldots\right)$$

$$= -\frac{a}{b}\left[-\log(1-b)\right] + \frac{a}{b}\left[-\log(1-bz)\right]$$

Therefore $f(z) = (1-b)^{\frac{a}{b}} \cdot (1-bz)^{\frac{-a}{b}}$

But $P(r)$ is the coefficient of z^r in the expansion of

$$f(z) = \sum_r z^r \cdot P(r),$$

or $$P(r) = \frac{f^{(r)}(o)}{r!}$$

Now $$f^{(r)}(z) = (1 - b)^{\frac{a}{b}} (-b)^r \left[(-1)^r . \frac{a}{b} \cdot \left(\frac{a}{b} + 1\right) \cdot \cdots \left(\frac{a}{b} + r - 1\right)\right] \cdot (1 - bz)^{-\left(\frac{a}{b} + r\right)}$$

so that $$P(r) = (1 - b)^{\frac{a}{b}} \cdot \binom{-a/b}{r} \cdot (-b)^r \tag{8a}$$

which is the exact form obtained by Dropkin in his formula (7) referred to above, if we substitute

$$a = \frac{n}{1 + d}, \quad b = \frac{1}{1 + d}$$

Dropkin's form being

$$P(r) = \left(\frac{d}{1 + d}\right)^n \cdot \binom{-n}{r} \cdot \left(\frac{-1}{1 + d}\right)^r \tag{8b}$$

(8a) is the general term in the expansion of

$$\left(\frac{1}{1 - b} - \frac{b}{1-b}\right)^{\frac{-a}{b}} \text{ and}$$

(8b) is the general term in the expansion of

$$\left(\frac{1 + d}{d} - \frac{1}{d}\right)^{-n}$$

To make connection with the form (1) shown above, substitute

$$p = \frac{1}{d} \quad q = 1 + p, \text{ so that } \frac{1}{q} = \frac{d}{1 + d}, \quad \frac{p}{q} = \frac{1}{1 + d}$$

Then $$P(r) = \left(\frac{1}{q}\right)^n \cdot \binom{-n}{r} \cdot \left(\frac{-p}{q}\right)^r$$

$$= \frac{(r + n - 1)!}{r!\ (n - 1)!} \cdot \left(\frac{1}{q}\right)^n \cdot \left(\frac{p}{q}\right)^r \tag{8c}$$

which is the general term in the expansion of $(q - p)^{-n}$, being identical with (1).

The moments are most neatly derived from this form by use of the moment-generating function, as demonstrated by Simon ("The Negative Binomial and Poisson Distributions Compared", *PCAS XLVII*, p. 20.)

$$\varphi(\theta) = \sum_r P(r) \cdot e^{\theta r} = (q - pe^{\theta})^{-n}$$

$$\text{whence } E(r) = \frac{\partial \varphi(o)}{\partial \theta} = n\,p \qquad \text{(for 8c)}$$

$$= \frac{n}{d} \qquad \text{(for 8b)}$$

$$= \frac{a}{1 - b} \qquad \text{(for 8a)}$$

$$\text{Similarly } E(r^2) = \frac{\partial^2 \varphi(o)}{\partial \theta^2} = n\,p + n(n+1)\,p^2$$

$$\text{Whence } \mu_2 = n\,p\,q = n\,p + n\,p^2 \qquad \text{(for 8c)}$$

$$= \frac{n}{d} + \frac{n}{d^2} \qquad \text{(for 8b)}$$

$$= \frac{a}{(1 - b)^2} \qquad \text{(for 8a)}$$

By a similar process,

$$\mu_2 = np + 3\,np^2 + 2\,np^3 \qquad \text{(for 8c)}$$

$$= \frac{n}{d} + 3\,\frac{n}{d^2} + 2\,\frac{n}{d^3} \qquad \text{(for 8b)}$$

$$= \frac{a\,(1 + b)}{(1 - b)^3} \qquad \text{(for 8a)}$$

These may be cross-checked by applying the same process to $\varphi(\theta) = \left(\frac{1 - be^{\theta}}{1 - b}\right)^{-\frac{a}{b}}$ for the moments of (8a) directly, or to $\varphi(\theta) = \left(1 - \frac{e^{\theta} - 1}{d}\right)^{-n}$ for the moments of (8b) directly.

The clarity of the significance of the parameters in the (8a) form should be noted: a is the expectancy of single-claim accidents, b is the factor which links this expectancy with those of two-claim accidents, of three-claim accidents, and so on, through formula (2).

The number of parameters in the general formula (4) can be reduced by a variety of assumptions, producing a number of related formulas. For example, if we let $a_2 = a_3 = a_4 = \ldots\ldots = 0$, we have the one-parameter Poisson distribution for which m (= mean), μ_2 and μ_3 are the first terms in the three expressions (7a)—(7c) above; if we let $a_3 = a_4 = \ldots\ldots = 0$, we have a two-parameter distribution in which m, μ_2 and μ_3 are the first two terms in (7a)—(7c); similarly, the three-parameter distribution derived by letting $a_4 = a_5 = \ldots\ldots = 0$, has m, μ_2 and μ_3 equal to the first three terms of (7a)—(7c) respectively.

It is interesting to note that this particular three-parameter distribution provides a closer fit than the negative binomial distribution for data on the number of railway accident fatalities in the Saar in a test made by Lüders; in other words, the assumption that such fatalities occur only singly or in pairs or in three's but with these three expectancies unrelated each to each, accords a closer fit in this case than the assumption that they occur in groupings of 1, 2, 3, . . . at a time with the frequencies of these occurrences linked as in relation (2) herein; or, if you will, which is also significant, closer than the assumption that the probability of a fatal accident varies by individual in accordance with a Pearson Type III distribution. This reminds us that the modified power-series relationship assumed in (2) is of course essentially as arbitrary as the Pearson Type III assumption; yet the underlying idea and the results open a fertile area for further investigation, which should include the associated formulas developed herein.

The final justification of any of these formulas lies in the results of tests. I have not had the facilities or the time to test the ideas suggested by these various developments and hope that this will in due course be done by others having both. In particular the possibility of utilizing the negative binomial formula for fitting a distribution of the number of claims is worthy of more study, since what actuaries have at hand usually is a claim count rather than an accident count. We should determine whether its fit is closer with claim distributions than with accident distributions, or more exactly, whether its fit is closer with multiple occurrences in a single accident counted separately than with a strict accident count.

As Simon has remarked, a study of the negative binomial opens up a rather amazing variety of applications and interpretations, many of them of interest to us as actuaries. These observations on rationale by no means exhaust the subject, but should really serve to whet our curiosity, and they merely bear out the quotation that prefaced this paper. In closing, let me say that once again Pope's dictum has been fulfilled: "There is no study that is not capable of delighting us after a little application to it."