DISCUSSION BY SHOLOM FELDBLUM

Actuaries have generally used the Poisson distribution to model accident frequencies for "repeatable" risks—that is, where more than one accident is possible in an exposure period. The reasons for this are both theoretical and practical. Theoretically, if the following conditions are true, then the Poisson distribution is indicated:

1. The probability of exactly one accident in an infinitesimal unit of time dt is approximately equal to $k\ dt$, where k is a constant, during any interval of time in the exposure period.
2. The probability of more than one accident in an infinitesimal unit of time is negligible compared to the probability of exactly one accident.
3. The distribution is "memoryless"; that is, the numbers of accidents in distinct intervals of time are independent.

Practically, the Poisson distribution is mathematically convenient in numerous ways:

1. The mean and variance of the Poisson distribution are equal, so the variance may be estimated along with the mean from a simple averaging of raw results.
2. Since the mean and variance are equal, their ratio is unity, a known constant. This makes "classical pure premium credibility" easier to calculate, as discussed by Mayerson, Jones, and Bowers [1].
3. The Poisson distribution is conjugate to the gamma, a distribution both convenient and realistic for modeling the mean accident frequency among individuals in a population. This makes Bayesian estimation of future mean accident frequency distributions particularly convenient.
4. The Poisson claim frequency distribution can be combined with a claim size distribution to form a "compound Poisson" aggregate claim distribution. The compound Poisson distribution has advantages over other compound distributions. For instance, if the claim size distribution is discrete (or can be realistically modeled by a discrete distribution), the aggregate claim distribution can be determined by a recursive procedure, which facilitates the mathematics of determining this distribution [2].

Lester Dropkin's paper complicates this simplified Poisson world [3]. He points out that if the accident frequency is Poisson distributed for each individual in a population, but the mean accident frequencies vary by individual, then the accident frequency for the population as a whole is no longer Poisson distributed.

In particular, if the mean accident frequencies among individuals are gamma distributed, and the accident frequency for each individual is Poisson distributed, then the accident frequency for the population as a whole has a negative binomial distribution.

For the Poisson frequency, the mean and variance are equal. For the negative binomial, the variance is always greater than the mean. Using Dropkin's notation, the mean of the negative binomial is r/a, while the variance is $r(a + 1)/a^2$. (In this notation, r is the scale parameter and a is the shape parameter of the underlying gamma distribution.)

This analysis shows that the wider the dispersion of mean accident frequencies among individuals, the greater the variance of the total population accident frequency. For instance, suppose that the mean accident frequency for the population as a whole is 1; that is, $r = a$. For the underlying gamma distribution, the mean is a/r and the variance is a/r^2; thus, the ratio of the variance to the square of the mean is $1/a$. That is, as a decreases, the values are more widely dispersed relative to the mean, and as a increases, the values are more closely situated to each other. By examining the variance of the negative binomial distribution, we note that as a decreases, the variance of the population accident frequency increases, and vice versa as a increases.

Surprisingly, this result is not generally true. In life insurance, the accident frequency is generally modeled as a Bernoulli random variable, since at most one claim is possible per individual per exposure period. The mean death rate of the population, that is, the parameter of the Bernoulli random variable, may be determined from actual data by dividing the number who die at a given age by the number exposed in the population at that age. For example, if there are 1,000 individuals at age 50 in the population, and 20 individuals die at age 50, then the mean death rate at age 50 is 2%.[1]

Using the actual data, we hypothesize that the mean death rate is 2%. Assuming also that the death rate is 2% for each individual, the variance of the death rate for each individual is (0.02) (0.98) = 0.0196. The variance of the estimate of the mean death rate is (0.02) (0.98)/1000 = 0.0000196.

After reading Dropkin's analysis, one may question this: since the individual death rates vary about 2%, and only average to 2% for the population as a whole, should not the population variance differ from 0.0196? Should it not be

[1] To be exact, we should assume that there is no migration; i.e., there are no new entrants or withdrawals at age 50. Thus, there were 1,000 individuals who attained age 50; 20 of these died during the course of the year; and 980 individuals attained age 51.

similar to the Poisson case, where if the population mean accident rate is 2%, but the individual mean accident rates vary about 2%, the population variance is greater than 2%?

The answer is no, as can be seen by a simple example, as well as by a more formal mathematical proof. Suppose there are two individuals, with a mean population death rate of 50%. Assume two cases:

1. Each individual has a death rate of 50%.
2. One individual has a death rate of 75%, and the other individual has a death rate of 25%.

For each case we determine the first two moments for each individual, the moments of the "mixture" distribution, and the variance of the "mixture" distribution.

1. Each individual has a death rate of 50%. For each individual, both the first and second moments are 0.50, and so the first and second moments of the mixture distribution are also 0.50. Therefore the variance of the mixture distribution is 0.50 − (0.50) (0.50) = 0.25.
2. One individual has a death rate of 75%, and the other individual has a death rate of 25%. For the first individual, the first and second moments are 0.75, while for the second individual, they are 0.25. Therefore, the first and second moments of the mixture distribution are 0.50, and the variance is 0.25.

The general proof follows the same reasoning. Suppose each individual has a mean death rate of p_i, and over the population as a whole these average to m. Then the second moments for each individual are also p_i, and over the population as a whole these also average to m. Therefore, the variance of the population mean death rate is $m\ (1 - m)$.

The Bernoulli distribution allows only one occurrence, while the Poisson distribution has no limit on the number of occurrences. What if the number of possible occurrences is finite but is greater than one, such as with the binomial distribution?[2]

[2] The following result for the binomial distribution was shown to me by Dr. Rodney Kreps, an actuary at Fireman's Fund Insurance Companies.

Theorem: Suppose the accident frequency is modeled by a binomial distribution with parameters p_i and n, i.e.,

$$f(x) = \binom{n}{x} p_i^x (1 - p_i)^{n-x}$$

Further, suppose n is fixed for all individuals in the population, but p_i varies according to a p.d.f. $g(p)$, which has mean m and variance s^2. For each individual, the mean is $n\, p_i$, and the variance is $n\, p_i\, (1 - p_i)$. For the population as a whole, the mean is $n\, m$, and the variance is $n\, m\, (1 - m) + s^2\, n\, (n - 1)$.

Proof: For each individual, the mean is $n\, p_i$, and the second moment is $n\, p_i - n\, p_i^2 + n^2\, p_i^2$.

Therefore, the mean for the population is

$$\int_0^1 (n\, p)\, g(p)\, dp = n\, m.$$

The second moment for the population is

$$\int_0^1 (n\, p - n\, p^2 + n^2\, p^2)\, g(p)\, dp$$

$$= n\, m - n\, SM + n^2\, SM \text{ (where } SM \text{ is the second moment of } g(p))$$
$$= n\, m + n\, (n - 1)\, (SM - m^2) - n\, (n - 1)\, m^2$$

Subtracting the square of the mean, we get

$$= n\, m\, (1 - m) + n\, (n - 1)\, s^2, \text{ which is the desired result.}$$

Thus, the more that the number of possible occurrences for each individual (n) increases, the more the variance for the population as a whole depends upon the variance of $g(p)$.

A useful application of this result is in Bayesian estimation. Generally, in performing a Bayesian estimation, the accident frequency is chosen as Poisson or binomial, and the prior distribution as gamma or beta. However, the problem of selecting the parameters of the prior distribution can be serious, and the choice of these parameters will influence the resultant posterior distribution [4].

The above result provides a method of selecting parameters. Suppose the binomial distribution is chosen for accident frequency, with a given n. Then if the population mean is u, the mean of the prior distribution of the p_i is $u/n = m$. Similarly, if the variance of observed results is VAR, the variance of the prior distribution of the p_i is

$$s^2 = (VAR - n\, m\, (1 - m)) \,/\, (n\, (n - 1)).$$

Given these values of m and s^2, the two parameters of the prior beta distribution are easily determined.

This discussion may help clarify two apparently misleading items in Dropkin's paper. First, Dropkin's criterion for choosing whether to model accident frequency by a Poisson or negative binomial distribution is the observed relation of the population variance to the population mean. If the population variance is approximately equal to the binomial variance, i.e., $p\,(1 - p)$, where p is the population mean, then use a Poisson distribution; if it is significantly larger, then use a negative binomial distribution.

Presumably, this criterion should be, "If the population variance is approximately equal to the Poisson variance, i.e., p, where p is also the population mean, then use a Poisson distribution . . ." Dropkin's statement at first seems logical if he is referring to the probability of having one or more accidents, rather than to the number of accidents per exposure unit. But then the accident frequency is a Bernoulli distribution, and the population variance will be independent of the underlying distribution of mean accident frequencies.

Second, Dropkin implies that the only choices for modeling the accident frequency are the Poisson and the negative binomial. He shows that his data has a variance and mean incompatible with the Poisson distribution, and he concludes:

> We can, however, go further. Since a Poisson distribution is not indicated for the distributions by number of accidents, a negative binomial is indicated [3].

This is hardly so. His actual data only indicates that the Poisson distribution does not provide a perfect fit. It in no way indicates that a negative binomial distribution is better than other two-parameter distributions. The negative binomial distribution is only "indicated" if one assumes that each individual has a Poisson accident frequency and the mean accident frequencies among individuals are gamma distributed. One may test this by calculating the third moment of the observations and comparing it to the hypothetical third moment of the negative binomial distribution; unfortunately, Dropkin does not do this.[3]

---

[3] George Phillips, an actuary with the Transamerica Corporation, has recommended to me that use of other statistical methods, such as percentile matching, may give better results than examination of the third moment.

Other users of Dropkin's results have adopted this reasoning, such as Mayerson, et al., in "On the Credibility of the Pure Premium" [1] (though since the authors' purposes there are heuristic and not intended for practical applications, one can hardly fault them). They take the first two moments from Dropkin's accident frequency distribution, assume that it can be modeled by a negative binomial distribution, and calculate the third moment. But until one compares the derived third moment with the observed third moment, there is no evidence that the negative binomial provides an appropriate model. (I must reiterate, though, that the purpose of this paper is only to show how to apply a theory, not to provide firm credibility tables, and for such heuristic purposes, the assumptions are entirely plausible.)

To sum up: the Poisson distribution is a theoretically appealing model for accident frequencies for each individual. The accident frequency distribution for the population as a whole will depend upon the distribution of mean accident frequencies among the individuals in the population. The negative binomial distribution for the population accident frequency is indicated only if the individual mean accident frequencies are gamma distributed. The form of the mean accident frequency distribution may depend upon the line of insurance, class of risk, and so forth; in any case, there is no easy way to test it. Rather, one may test the first three (or more) moments of the observed results. In Dropkin's case, the first two moments provide the parameters of the negative binomial distribution as well as of the underlying gamma distribution. The observed third moment would then test whether the negative binomial provides an appropriate model. If not, a different two parameter population accident frequency distribution may be assumed. If the individual accident frequencies are Poisson distributed, this implies an underlying distribution of mean accident frequencies among individuals that is not gamma. Once more, the observed third moment can test whether this population accident frequency model is appropriate.

Of course, the more complex the distribution chosen, the better it may agree with observed results, but the less mathematically tractable it may be—and a mathematically intractable model is hardly useful. A great advantage of the Poisson distribution is its simplicity; the negative binomial distribution is also quite versatile. Nevertheless, there is a need to test the hypothetical models, to strike a balance between simplicity and accuracy.

REFERENCES

[1] A. L. Mayerson, D. A. Jones, and N. L. Bowers, Jr., "On the Credibility of the Pure Premium," *PCAS* LV, 1968, p. 175.

[2] Society of Actuaries, "Risk Theory," Part V Examination Study Note.

[3] Lester B. Dropkin, "Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records," *PCAS* XLVI, 1959, p. 165.

[4] Dick London, *Graduation: The Revision of Estimates*, Winsted and Abington, Connecticut, ACTEX Publications, 1985, ch. 5.