THE NEGATIVE BINOMIAL APPLIED TO THE CANADIAN MERIT RATING PLAN FOR INDIVIDUAL AUTOMOBILE RISKS

BY

CHARLES C. HEWITT, JR.

1. *Summary of Current Theoretical Developments*

Dropkin[1] has shown that an excellent fit of actual automobile accident frequencies is obtained by the use of the negative binomial distribution.

The negative binomial distribution is justified on the assumption that, for a particular mean accident frequency, the Poisson distribution:

$$P(x; mt) = \frac{(mt)^x e^{-mt}}{x!} \tag{1.1}$$

will hold. In the above expression *m* represents the mean accident frequency for a particular unit period of time (normally it will be assumed that *one year* is the unit of time), and *t* represents the number of units of time exposed (years). The mean of this Poisson distribution is:

$$E(x; mt) = mt \tag{1.1a}$$

and the variance of this Poisson distribution is:

$$\sigma^2(x; mt) = mt \tag{1.1b}$$

A significant step in Dropkin's approach is the assumption that the mean accident frequency varies among drivers (or cars) and that this variation can be expressed by a Pearson Type III curve of the form:

$$T(m) = \frac{a^r}{\Gamma(r)} m^{r-1} e^{-am} \tag{1.2}$$

(*a* and *r* positive)

The mean of this frequency distribution is:

$$E(m) = \frac{r}{a} \tag{1.2a}$$

and the variance of this frequency distribution is:

$$\sigma^2(m) = \frac{r}{a^2} \tag{1.2b}$$

Dropkin shows that if the mean individual accident frequency *m* is a continuous random variable with a range from 0 to ∞ and with a frequency distribution T(m), then the group probability of exactly *x* accidents during a unit time interval may be obtained by:

[1] Dropkin, Lester, *Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records*, CAS XLVI, p. 165.

$$N(x; t) = \int_0^\infty P(x; mt) \cdot T(m)dm \tag{1.3}$$

This expression takes the form of the negative binomial:

$$(t=1) \qquad N(x; 1) = \left(\frac{a}{a+1}\right)^r \binom{-r}{x} \left(\frac{-1}{a+1}\right)^x \tag{1.4}$$

or more generally for a period of time *t*:

$$N(x; t) = \left(\frac{a}{a+t}\right)^r \binom{-r}{x} \left(\frac{-t}{a+t}\right)^x \tag{1.5}$$

The mean[2] of this latter distribution is:

$$E(x; t) = \frac{r}{a} t \tag{1.5a}$$

and the variance:[2]

$$\sigma^2(x; t) = \frac{r}{a} \frac{a+t}{a} t \tag{1.5b}$$

It is important to remember that, although the mean accident frequencies, *m*, (for unit time interval) may vary as among individuals, it is assumed that *m* remains constant over the period of time under consideration for a particular individual.

Bailey and Simon[3] have introduced the concept that the occurrence or non-occurrence of accidents during a particular period of time creates different groupings of individuals by inherent hazard, where the basis of grouping is the individual driving record during the period of time under consideration. These groupings based on driving record have frequency distributions of the group members by inherent hazard, which frequency distributions differ as among groupings and as against the original frequency distribution by inherent hazard of all drivers. This process of creating groupings based upon driving record is in effect a process of selection and is completely random.

It is quite possible that an inherently good risk through bad luck may find himself in a class with risks who have had one or more accidents during the time period under review; on the other hand, an inherently bad risk may find himself in a classification with risks who have had no accidents during a particular period of time. However, in the long run the process of selection on the basis of driving record will result in a greater frequency of good risks in the class with no accidents and a relatively greater frequency of poor risks in a class with one or more accidents.

---

[2]For further characteristics of the negative binomial distribution see Simon, L. J., *The Negative Binomial and Poisson Distributions Compared,* CAS XLVII, p. 20.

[3]Bailey, R. A. and Simon, L. J., *An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car,* Appendix 1, CAS XLVI, p. 164.

In an actuarial note currently being presented to the Society, Dropkin[4] shows analytically how this process of selection works and develops an expression for the frequency distribution by inherent hazard in a particular grouping as well as an expression for forecasting claim frequency for this grouping.

Using functions contained in the previous section, it can be shown that for the group of risks who have incurred exactly c accidents in the last s years, the distribution function by inherent hazard, m, for this particular group is given by:

$$T(m/c, s) = \frac{(a+s)^{r+c}}{\Gamma(r+c)} m^{r+c-1} e^{-(a+s)m} \tag{1.6}$$

In this system of notation the expression (c, s) will be assumed to stand for the grouping of risks who have had exactly c accidents in the last s years. The mean of this frequency distribution is given by:

$$E(m/c, s) = \frac{r+c}{a+s} \tag{1.6a}$$

and the variance is:

$$\sigma^2(m/c, s) = \frac{r+c}{(a+s)^2} \tag{1.6b}$$

From (1.6), (1.6a) and (1.6b) it is evident that $T(m/c, s)$ is, itself, a Pearson Type III distribution of the form contained in (1.2) where r is replaced by $r+c$ and a is replaced by $a+s$. Furthermore each grouping, (c, s), contains (from (1.5)):

$$N(c; s) = \left(\frac{a}{a+s}\right)^r \binom{-r}{c} \left(\frac{-s}{a+s}\right)^c \tag{1.6c}$$

risks from the initial population.

With respect to future (or forward) accident frequency, the probability of exactly x accidents in a future period of time t is given by a negative binomial expression:

$$N(x; t/c; s) = \left(\frac{a+s}{a+s+t}\right)^{r+c} \binom{-(r+c)}{x} \left(\frac{-t}{a+s+t}\right)^x \tag{1.7}$$

The mean number of accidents during the period of time t will be:

$$E(x; t/c; s) = \frac{r+c}{a+s} t \tag{1.7a}$$

and the variance:

$$\sigma^2(x; t/c; s) = \frac{r+c}{a+s} \cdot \frac{a+s+t}{a+s} t \tag{1.7b}$$

[4]Dropkin, Lester. *Automobile Merit Rating and Inverse Probabilities.* CAS XLVII, p. 37.

The expressions given in (1.6a) and (1.7a) are remarkably simple expressions for the mean forward accident frequency of a grouping of risks who have had exactly c accidents in the last s years. From the mean of all risks (1.2a):

$$E(m) = \frac{r}{a}$$

it is possible to develop an expression for an experience modification on the basis of driving record. This experience modification would have the form:

$$\overline{\text{Mod}}_{c,s} = \frac{\dfrac{r+c}{a+s}}{\dfrac{r}{a}} = \frac{a(r+c)}{r(a+s)} \tag{1.8}$$

and in the special case where the risk is accident-free during the time interval s, this, of course, indicates the "no claim bonus":[5]

$$1 - \frac{ar}{r(a+s)} = \frac{s}{a+s} \tag{1.8a}$$

Of course, such modification is based entirely on accident frequency and ignores the question of accident severity.

Parenthetically it might be noted here that data with respect to accident severity is rather limited. The Canadian information that is available indicates that except for those risks who have had one or more accidents in the past year, there is little or no variation in severity on the basis of driving record. For those risks who have had an accident in the past year, Canadian data indicates that accident severity is approximately 10% greater than for risks in the various accident-free classifications.

2. *Analytical Expressions for Canadian Merit Rating Classes*

The expressions developed in the previous section make is possible to forecast forward accident frequency for any group of risks where the accident history is known. However, the Canadian system of classifying by accident record does not always permit application of these formulae. The Canadian merit rating classes are as follows:

Class A – No claim within past three years (or more)
Class X – No claim within past two years
Class Y – No claim within past year
Class B – One or more claims within past year.

In the case of Class B, the exact value of c is not known since there will undoubtedly be some risks in this class which have had more than one claim in

[5] cf. Written discussion by Bailey, R. A. of *Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records*, Dropkin, L. ibid.

the past year. In the case of Class A, the exact value of s is not known since some risks will have been claim-free for more than the three years necessary to qualify for this class.

In order to permit more complete analysis of the available Canadian data, it is necessary to derive expressions (similar to those developed by Dropkin) with the Canadian merit rating classification system in mind.

The distribution function for risks which have had one or more claims (where the exact number of claims is not stated) during a period of s years may be developed as follows:

$$T(m/>0, s) = \frac{(1-e^{-ms})T(m)}{\int_0^\infty (1-e^{-ms})T(m)\,dm}$$

$$T(m/>0, s) = \frac{1-e^{-ms}}{1-\left(\frac{a}{a+s}\right)^r} T(m) = \frac{T(m)-T(m/0, s)}{1-\left(\frac{a}{a+s}\right)^r} \qquad (2.1)$$

In the above expressions the notation used follows the pattern of the previous analysis. It will be recognized that the number of risks falling into this particular classification is represented by the denominator of (2.1). The mean of this distribution function is given by the expression:

$$E(m/>0, s) = \frac{\frac{r}{a} - \left(\frac{a}{a+s}\right)^r \frac{r}{a+s}}{1-\left(\frac{a}{a+s}\right)^r} \qquad (2.1a)$$

The variance of this distribution function is quite involved and not important to subsequent analysis and is therefore not given here.

The probability of exactly x claims during some period of time t is given by the difference of two negative binomials:

$$\frac{\left(\frac{a}{a+t}\right)^r \binom{-r}{x}\left(\frac{-t}{a+t}\right)^x - \left(\frac{a}{a+s+t}\right)^r \binom{-r}{x}\left(\frac{-t}{a+s+t}\right)^x}{1-\left(\frac{a}{a+s}\right)^r} \qquad (2.2)$$

The mean value of the number of accidents during a future period of time t is:

$$\frac{\frac{r}{a} - \left(\frac{a}{a+s}\right)^r \frac{r}{a+s}}{1-\left(\frac{a}{a+s}\right)^r} t \qquad (2.2a)$$

Again, the variance with respect to this function is quite involved and is not given here. The denominator of expressions contained in (2.2) and (2.2a) represents the number of risks from the original population who fall into the group.

Now, it is necessary to develop a new distribution function consisting of those drivers who have had one or more claims (exact number of claims not stated) in an s year period and who have then been claim-free during a w year period immediately following. This function can be derived and expressed as follows:

$$T(m/\!/0, w/{>}0, s) = \frac{e^{-mw}T(m/{>}0, s)}{\int_0^\infty e^{-mw}T(m/{>}0, s)\,dm}$$

$$T(m/\!/0, w/{>}0, s) = \frac{[e^{-mw} - e^{-m(s+w)}]T(m)}{\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+s+w}\right)^r} \tag{2.3}$$

The notation $(0, w/{>}0, s)$ is intended to indicate that the risks in this particular group have had one or more claims in a time interval s and then have experienced no claims in the immediately subsequent time interval w.
The mean value of this frequency distribution is:

$$\frac{\left(\frac{a}{a+w}\right)^r \frac{r}{a+w} - \left(\frac{a}{a+s+w}\right)^r \frac{r}{a+s+w}}{\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+s+w}\right)^r} \tag{2.3a}$$

The probability that the risks in this group will have exactly x accidents in a period of t years in the future is given by the difference of two negative binomials:

$$\frac{\left(\frac{a}{a+w+t}\right)^r \binom{-r}{x}\left(\frac{-t}{a+w+t}\right)^x - \left(\frac{a}{a+s+w+t}\right)^r \binom{-r}{x}\left(\frac{-t}{a+s+w+t}\right.}{\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+s+w}\right)^r} \quad ($$

The mean value of this function is:

$$\frac{\left(\frac{a}{a+w}\right)^r \frac{r}{a+w} - \left(\frac{a}{a+s+w}\right)^r \frac{r}{a+s+w}}{\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+s+w}\right)^r}\,t \tag{2.4a}$$

In the two expressions immediately above, the number of risks out of the original population who fall into this particular group is expressed by the denominator:

$$\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+s+w}\right)^r \tag{2.4b}$$

It must be remembered that this expression does *not* hold for the group of risks who have had *no* claims in the period of time $s+w$. For this claim-free group, the number of risks is given by the expression:

$$N(0; s+w) = \left(\frac{a}{a+s+w}\right)^r \tag{2.4c}$$

In examining the results during an n-year period ($n = s+w$), the number of risks who have had one or more claims, but have then been claim-free for the most recent w (or more) years is [from (2.4b)]:

$$\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+n}\right)^r \tag{2.4d}$$

Similarly the number of risks who have had one or more claims, but have then been claim-free for the most recent $w+1\,(<n)$ years (or more) is:

$$\left(\frac{a}{a+w+1}\right)^r - \left(\frac{a}{a+n}\right)^r$$

Therefore the number of risks who have had one or more claims, but have then been claim-free for *exactly* the most recent w $(<n)$ years is:

$$\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+w+1}\right)^r \tag{2.5}$$

The average forward probability or claim frequency of such risks is given by the expression:

$$\frac{\left(\frac{a}{a+w}\right)^r \frac{r}{a+w} - \left(\frac{a}{a+w+1}\right) \frac{r}{a+w+1}}{\left(\frac{a}{a+w}\right)^r - \left(\frac{a}{a+w+1}\right)^r}\, t \tag{2.5a}$$

The weighted average forward probability as represented by the product of the number of risks in each such group times the average forward probability or claim frequency for the group is given by the following expression:

$$\left[\left(\frac{a}{a+w}\right)^r \frac{r}{a+w} - \left(\frac{a}{a+w+1}\right)^r \frac{r}{a+w+1}\right] t \tag{2.5b}$$

TABLE 1

ANALYTICAL EXPRESSIONS FOR CANADIAN PRIVATE PASSENGER AUTOMOBILE DATA

Merit Rating Class	Number of Risks	Average Forward Claim Frequency	Weighted Average Forward Claim Frequency (Number of Risks Times Forward Claim Frequency)
A	$\left(\frac{a}{a+3}\right)^r$	$\frac{r}{a+3}$	$\left(\frac{a}{a+3}\right)^r \frac{r}{a+3}$
X	$\left(\frac{a}{a+2}\right)^r - \left(\frac{a}{a+3}\right)^r$	$\dfrac{\left(\frac{a}{a+2}\right)^r \frac{r}{a+2} - \left(\frac{a}{a+3}\right)^r \frac{r}{a+3}}{\left(\frac{a}{a+2}\right)^r - \left(\frac{a}{a+3}\right)^r}$	$\left(\frac{a}{a+2}\right)^r \frac{r}{a+2} - \left(\frac{a}{a+3}\right)^r \frac{r}{a+3}$
Y	$\left(\frac{a}{a+1}\right)^r - \left(\frac{a}{a+2}\right)^r$	$\dfrac{\left(\frac{a}{a+1}\right)^r \frac{r}{a+1} - \left(\frac{a}{a+2}\right)^r \frac{r}{a+2}}{\left(\frac{a}{a+1}\right)^r - \left(\frac{a}{a+2}\right)^r}$	$\left(\frac{a}{a+1}\right)^r \frac{r}{a+1} - \left(\frac{a}{a+2}\right)^r \frac{r}{a+2}$
B	$1-\left(\frac{a}{a+1}\right)^r$	$\dfrac{\frac{r}{a} - \left(\frac{a}{a+1}\right)^r \frac{r}{a+1}}{1-\left(\frac{a}{a+1}\right)^r}$	$\frac{r}{a} - \left(\frac{a}{a+1}\right)^r \frac{r}{a+1}$
Total	1	$\left[\frac{r}{a}\right]$	$\frac{r}{a}$
A + X	$\left(\frac{a}{a+2}\right)^r$	$\left[\frac{r}{a+2}\right]$	$\left(\frac{a}{a+2}\right)^r \frac{r}{a+2}$
A + X + Y	$\left(\frac{a}{a+1}\right)^r$	$\left[\frac{r}{a+1}\right]$	$\left(\frac{a}{a+1}\right)^r \frac{r}{a+1}$

All that now remains is a summation from a particular value of w to n ($>w$) in order to arrive at expressions for risk groups which have been accident-free for a period of w years or more *including* the group which has been claim-free for the entire n-year period.

These expressions are as follows:

the number of risks who have been accident-free for w years or more is:

$$\left(\frac{a}{a+w}\right)^r \tag{2.6}$$

the average forward claim frequency of risks in this group will be:

$$\frac{r}{a+w} \tag{2.6a}$$

and finally the average weighted forward claim frequency (product of the number of risks times the forward claim frequency) will be

$$\left(\frac{a}{a+w}\right)^r \frac{r}{a+w} \tag{2.6b}$$

This makes it possible to produce analytical expressions for the various Canadian merit rating classes. These are set forth in tabular form in Table 1 opposite.

3. *Test of Analytical Expressions Against Canadian Data*

It is possible to make a test of the above expressions against actual Canadian data bearing in mind the method in which risks are permitted to enter a particular classification.[6]

The parameters r and a can be determined by solving the simultaneous equations:

$$\frac{r}{a} = \text{forward claim frequency (All classes)}$$

and

$$\frac{a+3}{a} = \frac{\text{forward claim frequency (All classes)}}{\text{forward claim frequency (Class A)}}$$

for the Canadian data for policy years 1957 and 1958.[7]

The parameters used to determine theoretical claim frequencies become:

Class	$\bar{r}$	$\bar{a}$
1	2.6047	30.076
2	4.3044	35.733
3	4.1665	29.251
4	4.3859	27.065
5	4.5776	41.751

[6] Wittick, H. E., *The Canadian Merit Rating Plan for Individual Automobile Risks,* CAS XLV, p. 214.

[7] Bailey, R. A., and Simon, L. J., ibid, Table 1.

The test of the theoretical expressions against the actual claim frequencies produces the results given in Table 2 immediately below:

TABLE 2

CANADA (EXCLUDING SASKATCHEWAN)
POLICY YEARS 1957 and 1958 (AS OF JUNE 30, 1959)
PRIVATE PASSENGER AUTOMOBILE LIABILITY - NON-FARMERS

Merit Rating Class	Private Passenger				
	Class 1	Class 2	Class 3	Class 4	Class 5
			Theoretical Claim Frequencies		
A	.0787	.1111	.1292	.1459	.1023
X	.1107	.1388	.1629	.1823	.1261
Y	.1142	.1425	.1681	.1887	.1290
B	.1180	.1465	.1738	.1955	.1320
Total	.0866	.1205	.1424	.1621	.1096
A + X	.0812	.1141	.1333	.1509	.1046
A + X + Y	.0838	.1172	.1377	.1563	.1071
			Actual Claim Frequencies		
A	.0787	.1111	.1292	.1459	.1023
X	.1055	.1384	.1698	.1725	.1206
Y	.1183	.1470	.1741	.1716	.1259
B	.1377	.1591	.2008	.2000	.1501
Total	.0866	.1205	.1424	.1621	.1096
A + X	.0800	.1126	.1316	.1486	.1034
A + X + Y	.0820	.1148	.1347	.1511	.1049

With respect to Classes 1 through 5, the standard deviation of risk frequencies in each class may be calculated:

Class	Standard Deviation
1	.0537
2	.0581
3	.0698
4	.0774
5	.0512
Total	.0611

It is interesting to note, in passing, that Classes 1, 2 and 5 produce a standard deviation less than the standard deviation for the entire group of risks while Classes 3 and 4 produce a standard deviation greater than the standard deviation for all risks. However, it should be noted that the claim frequency in Classes 3 and 4 is relatively high. Using a measure of relative dispersion (standard deviation divided by the mean), all of the Classes 1 through 5 show a smaller relative dispersion than the entire group of risks.

It is recognized that the poorest fit of the theoretical expressions for forward claim frequency occurs in merit rating Class B (those risks who have had an accident in the most recent year). Further analysis of the data in this class is certainly warranted. It is possible that part of the difference between theoretical and actual frequencies in this merit rating class is explainable by the inclusion within the class of vehicles where the operator has had a conviction within the past year.