

A NOTE ON FULL CREDIBILITY FOR ESTIMATING CLAIM FREQUENCY

J. ERNEST HANSEN*

Introduction

The conventional standards for full credibility are known to be inadequate. This inadequacy has been well treated in the Mayerson, et. al. paper¹ and in the ensuing discussions,² where the general problem of estimating pure premium was considered. However, in spite of this previous treatment, that old, familiar number, 1,082, still enjoys widespread patronage.

If, instead of estimating pure premium, we ignore claim severity and estimate only claim frequency, 1,082 claims, with the precision in estimation which it promises, is an acceptable standard, providing we are sampling from a homogeneous risk population and accept the usual assumption of mutual independence among risks having Poisson claim processes. However, we know the insureds are not a homogeneous population. We must provide for a distribution of the Poisson parameter over the population, referred to as the structure function.³ The structure function introduces additional variation into the claim process which reduces the precision of estimation promised by the conventional credibility standards.

The More General Model

Let m denote the number of claims of an insured selected at random, and let λ denote the parameter of his Poisson claim process for a given interval of time, i.e., the experience period. Then the unconditional probability distribution of m can be represented as:

* J. Ernest Hansen has submitted this paper in response to a presidential invitation. Mr. Hansen is a Research Associate with the Insurance Company of North America.

¹ A. L. Mayerson, D. A. Jones, N. L. Bowers, Jr., "On the Credibility of the Pure Premium", PCAS Vol. LV, p. 175.

² PCAS Vol. LVI, pp. 63-82.

³ H. Bühlmann, *Mathematical Methods in Risk Theory*, Springer-Verlag, New York, 1970, p. 65.

$$P(m) = \int_{\lambda=0}^{\infty} P(m|\lambda)f(\lambda)d\lambda, \quad m = 0, 1, 2, \dots$$

where $P(m|\lambda)$ is the Poisson claim process of an individual insured conditional upon λ , and $f(\lambda)$ is the structure function describing the manner in which λ is distributed over a population of insureds.

The gamma distribution is often used as an example of a structure function, where we have:

$$P(m) = \int_{\lambda=0}^{\infty} e^{-\lambda} \frac{\lambda^m}{m!} \cdot \frac{\alpha\beta}{(\beta-1)!} e^{-\alpha\lambda} \lambda^{\beta-1} d\lambda$$

where α and β are the parameters for the gamma distribution. Upon integrating, we have:

$$P(m) = \frac{(m + \beta - 1)!}{m! (\beta - 1)!} \left(\frac{\alpha}{\alpha + 1} \right)^\beta \left(\frac{1}{\alpha + 1} \right)^m, \quad m = 0, 1, 2, \dots$$

The above representation of $P(m)$ is in the form of a negative binomial distribution.

Therefore, if we assume that individual insureds have independent Poisson claim processes and that the Poisson parameter is gamma distributed over the population of insureds, and if we select insureds at random to observe m , then the number of claims from an insured, m has the negative binomial distribution. A number of researchers have found the negative binomial distribution satisfactorily fits automobile claims data⁴.

From the general representation of $P(m)$, again assuming a Poisson claim process for individual insureds, and using well known results from conditional probability, we can readily determine:

$$\begin{aligned} E(m) &= E[E(m)|\lambda] \\ &= E(\lambda) \end{aligned}$$

Also:

$$\begin{aligned} \text{Var}(m) &= \text{Var}[E(m|\lambda)] + E[\text{Var}(m|\lambda)] \\ &= \text{Var}(\lambda) + E(\lambda) \end{aligned}$$

⁴H. L. Seal, *Stochastic Theory of a Risk Business*, John Wiley and Sons, Inc., New York, 1969, p. 16.

These results are immediate when we remember $E(m|\lambda) = Var(m|\lambda) = \lambda$ for the Poisson variable m with parameter λ . Therefore, even though the mixed Poisson process $P(m)$ can be mathematically difficult to work with, depending upon what structure function is selected, the mean and variance of m are simply related to those of λ . Considering a random sample of n insureds, we have, by invoking the Central Limit Theorem, the consequence that the sample mean, \bar{m} , is approximately normally distributed with a mean of $E(\lambda)$ and a variance of $[Var(\lambda) + E(\lambda)]/n$.

The exponential distribution, with only one parameter, is a convenient choice for $f(\lambda)$ for a numerical example. If $E(\lambda) = .35$, i.e., we expect .35 claims per insured for the experience period, then $Var(\lambda) = E^2(\lambda) = .1225$ for an exponential distribution and \bar{m} is approximately normally distributed with the following parameter values:

$$\begin{aligned} E(\bar{m}) &= E(\lambda) = .35 \\ Var(\bar{m}) &= [Var(\lambda) + E(\lambda)]/n = (.1225 + .35)/n \\ &= .4725/n \end{aligned}$$

If we want the estimator \bar{m} to be within 5% of $E(m)$ with probability .90, we determine n as follows:⁵

$$\begin{aligned} \text{standard normal deviate, } Z &= \frac{\bar{m} - E(\bar{m})}{\sigma_{\bar{m}}} \\ 1.645 &= \frac{.05 \times .35}{\sqrt{\frac{.4725}{n}}} \\ n &= 4,175 \end{aligned}$$

The number of claims we could expect in a sample of this size would be:

$$n \cdot E(m) = 4,175 \times .35 = 1,461$$

Therefore, assuming an exponential structure function and an expectation of .35 claims per insured, we find that the standard for full credibility would be 1,461 claims in contrast to the conventional standard of 1,082 claims. The difference is attributable to the additional variation in m introduced by the structure function.

⁵ For an exponential structure function, \bar{m} is a maximum likelihood estimator.

However, it is difficult to estimate the shape of the structure function for a particular population of insureds since an insured's risk parameter is not an observable random variable. We can observe the number of claims of a particular insured over time for purposes of estimating this risk parameter, the insured's expected claim frequency, but the true expected claim frequency is never known with certainty.

The conventional standards for full credibility are derived by assuming the structure function is concentrated at a single point,⁶ i.e., the risk parameter λ is assumed to be constant over the population of insureds and, therefore, $Var(\lambda) = 0$. If we reconsider the previous numerical example, with $E(\lambda) = .35$, but assume the structure function is concentrated at this point, we have:

$$Var(m) = (0 + .35)/n$$

Then:

$$1.645 = \frac{.05 \times .35}{\sqrt{\frac{.35}{n}}}$$

$$n = 3,092$$

The number of claims we would expect in a sample of this size would be $3,092 \times .35 = 1,082$ claims. This is the answer we should have anticipated, the conventional standard for full credibility.

The conventional standard, being adequate only for an extreme and improbable case, violates one of the basic purposes for employing the techniques of statistical inference. This is to establish procedures for estimation which guarantee a level of precision in the estimates, e.g., a probability of at least .90 of being within 5% of the true average claim frequency. The levels of precision associated with conventional standards represent the most precision possible using these standards rather than the least; we have a ceiling on possible precision when what we want is a floor.

Choice of the Structure Function

An ideal choice for a structure function is one that leads to gen-

⁶ Simon has used the term "isohazardous" to characterize such a population of insureds in his paper, "The Negative Binomial and Poisson Distributions Compared", PCAS Vol. XLVII, p. 20.

erally conservative standards for estimating claim frequency. Toward this end, we can use the following result from reliability theory: the coefficient of variation for all distributions which have an Increasing Failure Rate is bounded above by that of the exponential distribution.⁷ In particular, gamma distributions which may be used as structure functions have increasing failure rates.⁸ For such distributions, $Var(\lambda)$ is maximized, and $Var(m) = Var(\lambda) + E(\lambda)$ is maximized by assuming an exponential structure function for a given value of $E(\lambda)$. The maximum variance of m will then be:

$$\max. Var(m) = E^2(\lambda) + E(\lambda)$$

Credibility standards based on this variance will be adequate for the entire set of structure functions; the standards will be based on the maximum possible rather than the minimum possible variance.

In practice, the actuary is sufficiently familiar with the data he works with to select an upper bound for $E(\lambda)$, the expected claim frequency. Then, using an exponential prior distribution, a more adequate standard for full credibility can be easily computed, as in the previous numerical example. Using the above expression for $\max. Var(m)$ and letting ϵ denote the tolerance of error as a proportion of $E(\lambda)$, we can rearrange the formula of the example as:

$$n = \frac{1 + E(\lambda)}{E(\lambda)} \cdot \frac{Z^2}{\epsilon^2}$$

⁷ R. Barlow and F. Proschan, *Mathematical Theory of Reliability*, John Wiley and Sons, New York, 1965, p. 33. A distribution, $f(x)$, is IFR, has an increasing failure rate, if $f(x)/[1-F(x)]$ increases as x increases. If we restrict a population of insureds to just those for which $\lambda > \lambda_0$, an arbitrary value of λ , then the chance that an insured is close to λ_0 is given by the conditional probability $f(\lambda_0)d\lambda/[1-F(\lambda_0)]$. If $f(\lambda)$ is IFR, this conditional probability increases as λ_0 increases.

⁸ Gamma distributions which are asymptotic to the vertical axis are not intuitively appealing candidates for structure functions.

The following table was constructed using this formula with $Z = 1.645$ and $\epsilon = .05$:

*Full Credibility Standards with a Tolerance of Error
of 5% and 90% Confidence*

<u>Upper Bound for Claim Frequency</u>	<u>Sample Size: No. of Exposure Units</u>	<u>Expected No. of Claims</u>
.05	22,731	1,137
.10	11,907	1,191
.15	8,298	1,245
.25	5,412	1,353
.35	4,175	1,461
.50	3,247	1,624
.75	2,526	1,894
1.00	2,165	2,165
1.50	1,804	2,706
2.00	1,624	3,247
3.00	1,443	4,330
5.00	1,299	6,494

Conclusion

The conventional standards for full credibility are known to be minimal for the estimation of claim frequency. They are adequate only when the structure function is concentrated at a point. The exponential distribution appears to present a reasonable bounding case with respect to the additional variance introduced by the structure function. With the assumption of an exponential structure function and the selection of a maximum possible mean value for the Poisson risk parameter, an adequate sample size for estimating claim frequency can be computed.