A Bayesian Approach to Negative Binomial Parameter Estimation

Richard M. Duvall, Ph.D., CPCU

A BAYESIAN APPROACH TO NEGATIVE BINOMIAL PARAMETER ESTIMATION

Richard M. Duvall, Ph.D., CPCU
Sedgwick Financial and Actuarial Consulting

Abstract

Some procedures that are used to calculate aggregate loss distributions and claim count distributions assume the claim count distribution is a negative binomial distribution. The parameters for the negative binomial distribution are often based on data from a small number of loss periods, and the estimates may have considerable error. A Bayesian procedure for parameter estimation allows the analyst to use some judgment when deriving the parameter estimate. This paper derives the Bayesian estimation procedure of the negative binomial parameter, p, under the assumption that the prior distribution for p is a beta distribution.

A BAYESIAN APPROACH TO NEGATIVE BINOMIAL PARAMETER ESTIMATION

Introduction

Consulting actuaries often calculate probability distributions of aggregate loss. Two methodologies, among others, are frequently used to arrive at the distribution of aggregate loss. One methodology is to use a theoretical distribution such as the log normal, gamma, or other distribution to approximate the aggregate loss distribution. A second methodology is to combine a distribution for the number of claims, usually the Poisson or negative binomial distribution, and a distribution of claim size. The use of either of these methodologies may require an estimate of the parameters of the negative binomial distribution. Usually, the actuary is working with a small number of years, and the parameter estimate for the claim count distribution may have considerable error. This paper provides a Bayesian procedure for estimating the negative binomial parameters that will provide some stability to the estimate.

When selecting a claim count distribution, an argument can be made that the negative binomial should be preferred to the Poisson in almost all situations. Two sets of assumptions are presented that lead to a negative binomial distribution as opposed to a Poisson distribution. First, assume that there are several populations that produce losses, for example, losses from the members of a pool or trust or from several divisions of a company. Next assume that the number of claims from each population has a Poisson distribution with parameter λ_r. If the λ's are gamma distributed, the claim count distribution for claims from all populations is negative binomial [1, p. 323-4]. A mathematically equivalent set of assumptions is to assume a Poisson distribution for the claim count distribution, and assume that the sampling errors in estimating the Poisson parameter have a gamma distribution. Then the claim count distribution including the parameter estimation error is negative binomial. In this situation the relationship between the negative binomial distribution and the Poisson distribution is analogous to the relationship between the t-distribution and the normal distribution. At least one of these sets of assumptions is reasonable in almost every situation involving the use of claim count distributions in producing a distribution of aggregate loss.

The following notation for number of claims, size of individual claim, and aggregate loss is adopted. Let n_t be the number of claims in period t; x_i is the size of the ith claim; and

$$y_t = \Sigma\, x_i,\ i = 1,\ldots,n_t \tag{1}$$

is the aggregate loss for period t. It is well known in the actuarial profession that the variance of the distribution of aggregate loss from a compound process of claim count and individual loss where claim size and the number of claims are independent is [1, p.319]

$$\sigma_y^2 = \mu_n\sigma_x^2 + \mu_x^2\sigma_n^2. \tag{2}$$

Thus, if the claim count distribution is negative binomial, the mean and variance of the aggregate distribution will depend on the parameters of the negative binomial. Whether a theoretical distribution is used to represent the aggregate distribution or the aggregate distribution is derived by combining the claim count distribution and the severity distribution, an estimate of the parameters of the claim count distribution is required.

Negative Binomial

There are several forms of the negative binomial. The form used here is

$$P(n) = \binom{n+k-1}{n} p^k (1-p)^n \qquad (3)$$

where $\mu_n = k(1 - p)/p$, and $\sigma_n^2 = k(1 - p)/p^2$. Solving these two relationships for p and k gives $p = \mu_n / \sigma_n^2$ and $k = \mu_n^2/(\sigma_n^2 - \mu_n)$. To emphasize the dependence on the parameter p, expression (3) may be written as

$$P(n|p) = \frac{\Gamma(n+k)}{\Gamma(n+1)\Gamma(k)} p^k (1-p)^n, \qquad (4)$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$

This is a conditional distribution for n given a specific value of the parameter p.

It is assumed that the actuary has made forecasts for the expected number of claims, μ_n, and the variance of the claim count distribution, σ_n^2. The method of moments can be applied using the relationships above to estimate the parameters p and k of the negative binomial distribution. However, this paper provides a procedure for modifying these estimates based on prior beliefs concerning these parameters. This procedure will provide some stability to the estimates and will cause extreme sample results from a small number of loss periods to be modified toward the actuary's preconceived notions which may be based on past experience.

Prior Distribution

Assume that the prior distribution of p is a beta distribution with parameters b and c. Thus, the prior distribution is [2;p. 255]

$$f(p) = \frac{p^{b-1}(1-p)^{c-1}}{B(b,c)}, \quad 0 < p < 1, \qquad (5)$$

where, $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function.

Let p_p represent the mean of this distribution. $p_p = b/(b+c)$. By choosing appropriate values for b and c, the actuary can have a subjective notion of the parameter, p, enter into the estimation process. For example, if b and c are both assigned values of one, the mean of the prior distribution of p, p_p is one-half; or if b=1 and c=3, p_p is one-fourth. The prior expected value will be modified based on the sample data for a final estimate which will be an average of the subjective prior estimate of the actuary and an estimate based on the sample data.

While the relative sizes of b and c determine the expected value of the prior distribution, the absolute size of the sum of b+c will influence the weight given to p_p when it is averaged with the estimate from the sample data. A procedure for determining the weight to be assigned the prior estimate is provided below. With this procedure the actuary can influence the relative weights given the prior estimate and the sample estimate based on the confidence placed in these estimates.

Posterior Distribution

To derive the posterior distribution of p, the joint distribution of p and the observed sample must first be calculated. Using (4), the probability of selecting the observed sample for a given value of p may written as

$$P(n_1,\dots,n_T|p) = \prod_{t=1}^{T} \frac{\Gamma(n_t+k)}{\Gamma(n_t+1)\Gamma(k)} p^k (1-p)^{n_t} \tag{6}$$

$$= \frac{p^{Tk}(1-p)^{\Sigma n_t}}{\Gamma(k)^T} \prod_{t=1}^{T} \frac{\Gamma(n_t+k)}{\Gamma(n_t+1)},$$

where T is the number of loss periods contained in the sample. Multiplying (6) by the prior distribution for p, (5), gives the joint distribution of the observed sample and p.

$$P(n_1,\dots,n_T,p) = \frac{1}{B(b,c)\Gamma(k)^T} p^{kT+b-1}(1-p)^{\Sigma n_t+c-1} \prod_{t=1}^{T} \frac{\Gamma(n_t+k)}{\Gamma(n_t+1)} \tag{7}$$

The probability distribution for an observed sample is obtained by integrating (7) over the entire range of p, $0<p<1$. When the joint distribution of the sample observations and p is divided by the marginal distribution for the sample, the conditional distribution of p given the observed sample is obtained. Thus, dividing (7), the joint distribution, by (7) integrated with respect to p, the distribution for the observed sample, provides the conditional distribution for p given the observed sample as

$$f(p|n_1,\dots,n_T) = \frac{p^{kT+b-1}(1-p)^{\Sigma n_t+c-1}}{\int_0^1 p^{kT+b-1}(1-p)^{\Sigma n_t+c-1}\,dp} \tag{8}$$

The denominator of (8) is a beta function, and may be written as $B[kT+b,\Sigma n_t+c]$, and (8) can be written as

$$f(p|n_1,\ldots,n_T) = \frac{p^{kT+b-1}(1-p)^{\Sigma n_t+c-1}}{B(kT+b,\Sigma n_t+c)}. \tag{8a}$$

(8a) is a beta distribution and is the posterior distribution of p given the observed sample.

For a squared error loss function the Bayes estimator of p is the mean of this posterior distribution.

$$E(p|n_1,\ldots,n_T) = p_B = \frac{\int_0^1 p^{kT+b}(1-p)^{\Sigma n_t+c-1}dp}{B(kT+b,\Sigma n_t+c)}$$

$$= \frac{B(kT+b+1,\Sigma n_t+c)}{B(kT+b,\Sigma n_t+c)}$$

$$= \frac{kT+b}{kT+b+c+\Sigma n_t}$$

$$= \frac{kT+b}{kT+b+c+Tm_n}. \tag{9}$$

To put the expression at the same level as the forecast value for the number of claims, the number of sample periods, T, times the forecast number, m_n, is substituted for the total number of claims in the sample period in the last step of the derivation.

An estimate of k is required to use expression (9) to calculate p_B. One choice is to use the estimates of μ_n and σ_n^2 from the sample data and the relationship $k = \mu_n^2/(\sigma_n^2 - \mu_n)$ to get an estimate of k for use in (9). Substituting this expression for k in (9) produces

$$p_B = \frac{b(1-p_s)+p_s m_n T}{(b+c)(1-p_s)+Tm_n} \tag{9a}$$

The Bayes estimate, p_B, is an average of the actuary's subjective estimate, $p_p = b/(b+c)$, and the sample data estimate, $p_s = m_n/s_n^2$. The weight given to these estimates depends partly on the sum b+c. Let w_p be the weight given to p_p and $w_s = 1 - w_p$ is the weight given to p_s. Then

$$w_p\, p_p + (1 - w_p)\, p_s = p_B. \tag{10}$$

Making substitutions for p_p and p_B and solving for w_p gives

$$w_p = (1-p_s)(b+c) / [(1-p_s)(b+c) + m_nT]. \tag{11}$$

The weight received by the prior estimate depends on the size of $(1-p_s)(b+c)$ relative to the forecast number of losses and the number of loss periods in the sample, m_nT, the weight given to the sample estimate.

The question to be answered is the value to be assigned to $(b+c)$. If an alternative question is answered, the value of $(b+c)$ will be determined. The relative weights of p_s and p_b can be made to depend only on the number of loss periods of sample data that is available. Suppose that it is determined that equal weights will be given to the two estimates when T_e periods of data are available. Under this assumption $(1-p_s)(b+c) = m_nT_e$, and the weight assigned to w_p is

$$\begin{aligned} w_p &= m_nT_e/(m_nT_e + m_nT) \\ &= T_e / (T_e + T). \end{aligned} \tag{12}$$

For example, if it is decided that the prior value and the sample value should receive equal weight when there are four years of sample data, then $T_e = 4$. When $T < 4$, the prior estimate receives more weight than the sample estimate, and vice versa when $T > 4$. The weights assigned to p_p and p_s by expression (12) for selected numbers of years in the sample are:

Number of Years in Sample:	2	4	6	8	10
Prior Estimate Weight (w_p):	.667	.500	.400	.333	.286
Sample Value Weight (w_s):	.333	.500	.600	.667	.714

When the value of T_e has been selected, the expression for estimating p_B becomes

$$p_B = (T_ep_p + Tp_s) / (T_e + T). \tag{13}$$

The estimate of k will need to be calculated such that the negative binomial distribution will have an expected value that equals the claim count forecast. The value for k may be obtained from the expression $k = p_Bm_n/(1-p_B)$, where m_n is the claim count forecast. Having estimates of p and k, the estimated variance of the claim count distribution is $s_n^2 = k(1 - p_B)/p_B^2$.

<u>An Example</u>

The first three columns of the following table show for each year in the sample the estimate of the ultimate number of claims and the exposure in terms of head count. The fourth column inflates the claim count to an exposure equivalent the exposure for the forecast period by multiplying the claim count for each year by the ratio of the forecast exposure to the loss year exposure. The variance for the sample data is calculated in the last column. The prior estimate of p is $p_p = .5$. p_s is the ratio of the expected claim count to the variance, $p_s = 298/881 = .338$. If $T_e = 4$, then using (13) the Bayes estimate is made as

$$p_B = \frac{[4(.5) + 6(.338)]}{(4+6)} = .403.$$

Year	Claim Count	Exposure (Head count)	Inflated Claim Count	Squared Differences
1992	204	1282	334	1319
1993	226	1455	326	805
1994	219	1455	316	337
1995	226	1623	293	26
1996	214	1622	277	453
1997	240	1942	260	1465
Expected	298	2100	298	881

Using this estimate, k is calculated using the relationship $k = p_B m_n/(1-p_B) = .403(298)/.597 = 201$. Then the variance of the claim count distribution is estimated using $s_n^2 = k(1-p_B)/p_B^2 = 201(.597)/.403^2 = 739$.

A Bayesian procedure has been applied to produce an estimate of the variance of the claim count distribution that contains information relative to the analyst's prior estimates and experience. If experience adds valid information, this should be a more reliable estimate than one based solely on the sample data.

Summary

When calculating a probability distribution for aggregate losses for an accident year, an estimate of the variance of the claim count distribution is often required. When the number of accident years in the sample period is relatively small, an estimate based solely on the sample data is not reliable. This paper presents a methodology for estimating the variance of the claim count distribution that is based on a Bayesian procedure assuming a squared error loss function. The mean of the posterior distribution is the estimator that minimizes the expected squared error loss. The mean of the posterior distribution is a weighted average of the mean of the prior distribution and the sample estimate based on the sample moments. It is suggested that the actuary can choose the weights assigned to the prior estimate and the sample estimate that depend on the number of loss periods of sample data so that appropriate weights will be given to the two estimates.

References

1. Bowers, N. L., H. U. Gerber, J. C. Hickman, D. A. Jones, and C. J. Nesbitt, Actuarial Mathematics, Itasca, IL: Society of Actuaries, 1986.

2. LaValle, Irving H., An Introduction to Probability, Decision, and Inference, New York: Holt, Rinehart and Winston, Inc., 1970.