A BAYESIAN VIEW OF CREDIBILITY

ALLEN L. MAYERSON

Until recently, the credibility procedures used by casualty actuaries, and their theoretical justification, were developed apart from, and in isolation from, the methods used by statisticians. Arthur Bailey could write, in 1950:

> "At present, practically all methods of statistical estimation appearing in textbooks on statistical methods or taught in American universities are based on an equivalent to the assumption that any and all collateral information or a priori knowledge is worthless. There have been rare instances of rebellion against this philosophy by practical statisticians who have insisted that they actually had a considerable store of knowledge apart from the specific observations being analyzed. Philosophers have recently discussed the credibilities to be given to various elements of knowledge, thus undermining the accepted philosophy of the statisticians. However, it appears to be only in the actuarial field that there has been an organized revolt against discarding all prior knowledge when an estimate is to be made using newly acquired data." [14]

In 1950 the actuary stood nearly alone in his use of statistical techniques to modify his prior knowledge, instead of treating each new set of data as a separate statistical problem, to be used by itself if the volume of data was large enough to be statistically significant, or discarded if the contrary was the case. Because statistical techniques were not adequate to solve the actuary's problems, he developed his own methods. He ingeniously developed a credibility Z which was used to weight his prior knowledge B, with the current available statistical data A, by the formula $ZA+(1-Z)B$. But to determine Z, since there were no statistical techniques available, he has had to depend on empirical methods which, though they worked in practice, were hard to explain to non-actuaries and even harder to justify mathematically.

Statistical theory has now caught up with the actuary's problems. Starting with the 1954 book by Savage [8], and buttressed by the 1959 volume by Schlaifer [9] and the 1961 book by Raiffa and Schlaifer [7], there has been, among probabilists and statisticians, an organized revolt against the classical approach and a trend toward the use of prior knowledge for statistical inference. Instead of using credibility procedures, however, the Bayesian school of statisticians relies on Bayes theorem to merge the distribution representing prior knowledge with the statistical indications to produce a posterior distribution which reflects both.

At the same time as this revolution in the foundations of statistics, which formally reinstates prior opinion in statistical theory, advances have been made in probability and stochastic processes which result in mathematical techniques which lend themselves to the solution of actuarial problems and which can more easily be used by actuaries.

The relationship between Bayes theorem and credibility was first noticed by Arthur Bailey [14] who showed that the formula $ZA+(1-Z)B$ can be derived from Bayes theorem, either by assuming that the number of claims follow a Bernoulli process, with a Beta prior distribution on the unknown parameter p, or by assuming that the number of claims follow a Poisson process, with a Gamma prior distribution on the unknown parameter m. (The formula for Z differs, however, depending on whether a Bernoulli or a Poisson process is assumed.)

It seems appropriate, in view of the growing interest among statisticians in the Bayesian point of view, to attempt to continue the work started 15 years ago by Bailey, and, using modern probability concepts, try to develop a theory of credibility which will bridge the gap that now separates the actuarial from the statistical world. The purpose of this paper is to summarize the Bayesian point of view, to show its relevance to credibility theory, and to express credibility concepts in terms which are meaningful to a mathematical statistician.

THE "CLASSICAL" VIEW OF CREDIBILITY

As expounded by Whitney [38] in 1918, Perryman [33] and, more recently, Longley-Cook [30], the credibility theory now in use in the United States for fire and casualty insurance ratemaking rests on the following premises:

1. The formula $ZA+(1-Z)B$ can be used to modify the actuary's prior knowledge B (usually the rate currently being charged for a particular classification or, in experience rating, the manual rate) by the latest year's statistical data for the classification or risk in question, A.
2. The probability of an accident is the same for all insureds, namely q, and the total number of claims for n insureds follows a Poisson distribution

   $$f(x) = \frac{e^{-m}\ m^x}{x\,!}$$

   which has mean and variance both equal to $m = nq$.
3. The Poisson distribution may be approximated by a normal distribution. The normal distribution is a two parameter distribution, but for credibility work it is customary to assume that the mean and variance are both equal to m. Then, if P is the probability that

the actual number of accidents will be within $100k\%$ of the expected number,

$$P = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{\frac{-t^2}{2}} dt$$

where

$$x = \frac{(m + km) - m}{\sqrt{m}} = k\sqrt{m}.$$

For selected values of P and k, we may determine the value of x from tables of normal curve areas. From the relationship $m = \frac{x^2}{k^2}$ we can then obtain m, the level of expected claims for which the probability is P that the observed number of claims is within $100k\%$ of the expected number.

4. There is a certain number of expected claims which deserves a credibility of 1, and this number is the m determined from the normal curve calculations.
5. If the actual number of claims observed is equal to m, as calculated in 4 above, this set of data may be assigned a credibility of 1.
6. We can ignore the distribution of claim size, or loss severity, and use the number of claims, or loss frequency, to determine our credibilities. Or, if we wish to recognize the fact that the variation in claim severity is at least as great and usually greater than the variation in number of claims, we can do so by using a higher value of P or a lower value of k, thus stiffening our requirements for full credibility.
7. Once the full credibility point m has been settled, partial credibilities, for a volume of data yielding r claims, not large enough to merit full credibility, can be assigned by the formula $Z = \sqrt{\frac{r}{m}}$ or $Z = \frac{r}{r+k}$, where k is a normalizing constant.

It has recently been recognized (Dropkin [23], Simon [37], Bailey and Simon [17]) that assumption 2 is open to question. For example, in automobile insurance the claim frequency varies for different drivers. If we assume that the number of accidents for each driver is Poisson distributed, and that the means of these accident distributions are themselves random variables distributed according to a gamma distribution, the total number of accidents follows the negative binomial distribution, and the probability of exactly x accidents is:

$$f(x) = \binom{x+r-1}{x} p^r (1-p)^x$$

which has mean $\frac{r(1-p)}{p}$ and variance $\frac{r(1-p)}{p^2}$.

Furthermore, the data studied by Harwayne [26] and Dropkin [23] show a mean accident frequency of .163 and variance .193, which casts some doubt upon assumption 3 above. The mean and variance of the distribution of claim frequency are not equal. The data studied by Hewitt [27] also indicates a variance, in each of the classes studied, which differs somewhat from the mean.

THE BAYESIAN VIEWPOINT

The Bayesian view of statistical inference can best be summarized by a quotation from a recent paper by Edwards, Lindman and Savage [25]:

> "Probability is orderly opinion, and inference from data is nothing other than the revision of such opinion in the light of relevant new information."

This view of probability differs radically from that used by most classical statisticians. Most authors define probability in terms of symmetry or as the limit of a series of relative frequencies. For example, one classical definition of probability is:

> "The probability of the occurence of a given event is equal to the ratio between the number of cases which are favorable to this event, and the total number of possible cases, provided that all these cases are mutually symmetric." (Cramer [2])

Another way of expressing this definition is:

> "If an event can occur in N mutually exclusive and equally likely ways, and if n of these outcomes have an attribute A, then the probability of A is the fraction $\frac{n}{N}$." (Mood [6] p. 7)

Some authors embody the limit concept in their definition thus:

> "The proportion of the time that an event takes place is called its relative frequency, and the relative frequency with which it takes place in the long run is called its probability." (Freund [3] p. 124)

Even when probability is treated in the more modern terms of sets and sample spaces, it is usually defined in terms of symmetry i.e. equally likely elementary outcomes:

"The probability that an event A will occur is the ratio of the number of sample points that correspond to the occurrence of A to the total number of sample points." (Hoel [4] p. 6)

To a believer in an objectivistic definition of probability, the probability of an event may only be estimated by observing a series of trials of the event in question. Such questions as whether it will rain tomorrow, or whether there will be more automobile accidents next year than this year are considered, by holders of the relative frequency view of probability, to be completely outside the scope of probability. Such questions, they would say, have no meaning in probability terms.

By contrast, Bayesians believe that probability concepts may be used to express either the uncertainty of a future event or the uncertainty of unknown existing conditions. For a Bayesian, the probability of an event A is the largest price he would be willing to pay in exchange for the promise of a dollar if A turns out to be true. The probability that it will rain tomorrow is ⅓, for you, if you are willing to pay $.33 for the right to receive a dollar if, in fact, it does rain tomorrow.

The consistency among the probabilities an individual would assign to various events can be obtained by his being unwilling to accept a combination of bets that assures a loss no matter what happens. Bayesians avoid the apparent contradiction between scientific objectivity and irrational human behavior by postulating an ideal individual who is consistent in this sense. Such a man will confront each of his probabilities with his other beliefs and will maintain consistency between them. The actuary will want to work with a consistent set of probabilities; this is equivalent to requiring that the probabilities assigned to the various events obey the usual mathematical rules of probability.

Such a reasonable and prudent man will not only maintain consistency among his opinions, but will be willing to change them when confronted with new evidence. Furthermore, if there are two reasonable men who initially assign different probabilities (prior probabilities) to a given event, their revised probabilities (posterior probabilities) will draw closer together when they are confronted with external evidence as to the truth or falsity of a given event or proposition. If the evidence is overwhelming (has credibility one), their posterior probabilities will tend to merge, given some degree of initial open-mindedness, no matter how far apart they were before they saw the evidence.

The mechanism by which prior probabilities can be confronted by evidence is Bayes theorem, which states that the conditional probability that the hypothesis H is true, given that data D have been observed, $P(H|D)$, can be expressed as:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

where $P(H)$ is the prior probability for hypothesis H. The denominator, $P(D)$ can be expressed as

$$\sum_i P(D|H_i)P(H_i)$$

where $\{H_i\}$ represents a set of exhaustive and mutually exclusive hypotheses of which H is the particular one under examination. If we are only interested in whether H is true or false, then the set H_i compromises only two members, H and $\bar{H}$, and

$$P(D) = P(D|H)P(H) + P(D|\bar{H})P(\bar{H}).$$

The partition $\{H_i\}$ is often arbitrary. For example if H is the hypothesis "the average paid claim cost C for automobile bodily injury liability is $796 in 1963," the set may consist of only two other members, besides H, namely $C<796$ and $C>796$, or it may consist of a continuum of numbers x, with initial probability densities $f(x)$, such that

$$P(D) = \int P(d|x)f(x)dx$$

where H is the particular interval

$$795.5 < x < 796.5.$$

Bayesians emphasize decision making as the purpose of most statistical work; the purpose of obtaining a statistical estimate of μ is to decide on a certain course of action (e.g. what premium to charge) rather than merely to assert something about μ. By contrast, many statisticians believe that their function is limited to an analysis of the data and that decision making is a separate function; the decision maker, in their view, must combine the statistical results with his own judgment and other relevant factors in deciding what action to take.

The above short explanation of personal probabilities and the use of Bayes theorem is not intended to change the view of anyone who now holds the frequentist view of probability. A more extended and convincing treatment can be found in [25], [21] and [36].

CONJUGATE PRIOR DISTRIBUTIONS

The actuary is rarely interested in testing whether a hypothesis H is true or false. In most problems involving credibility he wants to determine, after seeing claim data for the latest calendar or policy year, whether the current manual rate needs to be modified. Or, his problem may be whether a particular insured should be charged a premium different from the manual rate. His initial point estimate H will usually be the current premium rate in the class under review. He would like to determine whether H must be modified, and to what extent, by the observed data D. Rarely, however, can he decide on the distribution of $P(H)$, the prior probability he is willing to assign to H, purely by introspection.

Fortunately, there is a way out of this dilemma, at least partially, through the theory of conjugate prior distributions, studied in detail by Raiffa and Schlaifer [7]. A prior distribution is said to be conjugate to an experiment when the prior distribution is so related to the conditional distribution that the posterior distribution is of the same type as the prior. For example, if D is viewed as the outcome of a Bernoulli process, and $P(D|H)$ is the binomial distribution, then the choice of a Beta distribution for $P(H)$ will result in a Beta distribution for $P(H|D)$ also, but with different parameters. If D is viewed as the outcome of a Poisson process, and $P(H)$ is chosen as a Gamma distribution, $P(H|D)$ will also be a Gamma distribution. If D is interpreted as the mean of independent normal observations with known variance, and $P(H)$ is assumed to be normal, then $P(H|D)$ will also be normal, but with smaller variance.

Arthur Bailey [14] studied both the Beta-Binomial and the Gamma-Poisson conjugate distributions and showed that, under either assumption, a credibility Z can be obtained, of the form $Z = \frac{n}{n+k}$, so that

$$E(H|D) = Z\,M_D + (1-Z)\,M_H.$$

Contrary to usual actuarial usage, where k is taken as an arbitrary normalizing constant, Bailey's formulas require that k be a specific function of the mean m and variance σ^2 of the prior distribution $P(H)$. If $P(H)$ is taken as a Beta distribution, and $P(D|H)$ is a binomial distribution, then

$$k = \frac{m - m^2 - \sigma^2}{\sigma^2}.$$

If $P(H)$ is assumed to be a Gamma distribution, and $P(D|H)$ is a Poisson distribution, then $k = \frac{m}{\sigma^2}$. It should be noted that Whitney [38] realized that k is not constant, but accepted an invariant k on grounds of expediency and simplicity.

The existence of conjugate prior distributions makes the actuary's job easier. If he thinks that the claim data he observes result from a Bernoulli process, he may, with a sufficient degree of approximation, be able to take $P(H)$ to be a Beta distribution. If he believes that his claim data come from a Poisson process, he may be able to assume that $P(H)$ is a Gamma distribution. In either case, he must choose m and σ^2, the mean and variance of $P(H)$, hence the parameters of the prior distribution, so that $P(H)$ adequately reflects his belief about H before seeing the observed data D. If there is a sufficient amount of data, the posterior distribution will not depend heavily on the exact form of the prior distribution.

The choice of m will, as a rule, be simple. m will be the pure premium, claim frequency, average claim cost, or whatever other actuarial function H is intended to test, e.g. if H is the hypothesis "the average paid claim cost C for automobile bodily injury liability is \$796 in 1963" m would be taken as 796.

The choice of σ^2 is much more difficult, and in its use lies a major departure from present actuarial practice. At present, the current claim frequency, pure premium, etc. is taken as fixed and assigned a credibility $1-Z$, where Z depends only on the number of claims or the amount of losses observed in D. Actually, the current premium rate, or its component claim frequency or claim cost, is itself only a parameter chosen to represent a distribution which has not only a mean, but also a variance and other moments. The classical view takes H to be an unknown constant which may be estimated, but holds that it is meaningless to speak of probabilities concerning H. The Bayesian, on the other hand, is willing to treat H as a random variable, with a distribution which reflects his current uncertainty regarding H.

THE CONCEPT OF FULL CREDIBILITY

In order to use credibility theory in ratemaking, an actuary must first determine the number of claims required for full credibility (Longley-Cook [30] p. 199). He then uses a formula, often based on the ratio of the number of actual claims in the observed data to some function of the number of claims which would be entitled to full credibility, to assign credibilities to data comprising fewer claims than this magic "full credibility" number. If the observed data for a particular classification results in a greater number of claims than the number required for full credibility, the data are taken at face value and are used for ratemaking, without reference to the previous manual rate or any other auxiliary information.

The concept of full credibility has always been rather difficult, philosophically. Some actuaries believe that no data are entitled to 100% credibility and that the credibility curve should approach 1 asymptotically, without ever reaching it. In Bayesian terms, however, the concept merges with that of partial credibilities in a natural and logical way. The Bayesian poses his credibility problem as that of modifying his prior opinion H by some observed data D. If the data are few, there is no reason for him to change H. $P(H|D)$ remains very close to $P(H)$. As the volume of data increases, $P(H|D)$ becomes more and more dependent on D and, finally, $P(H|D)$ comes to depend almost entirely on D. For a large enough volume of data, the posterior distribution is generally almost independent of the prior distribution.

Thus the Bayesian would pose the question of full credibility as: "For what prior distributions are these data fully credible, i.e. for what prior distributions can we say that, for practical purposes, the posterior distribution is independent of the prior distribution because of overwhelming data?" As we increase the volume of data, we increase the family of prior distributions for which this independence of posterior from prior is substantially true. There will, however, always be some prior distributions for which this is not true. For example, if the actuary (or insurance commissioner) chooses a prior distribution which is rather narrow, with all its mass concentrated in an interval close to last year's claim frequency, no amount of data will be sufficient to make him give up his prior opinion entirely, though he may be willing to modify it somewhat.

Most reasonable people will, however, alter their original beliefs if the data do not appear to support them. In actuarial work in particular, one must be exceedingly stubborn to hold to a narrow prior distribution, in the face of contrary evidence, because of the possibility, or even probability, of trends or secular changes in the underlying situation. Accident rates, average claim costs and other such quantities change with time, and the actuary is not, as a rule, surprised to find that this year's data differs somewhat from last year's.

CHOOSING PRIOR PROBABILITIES

The actuary's choice of his prior probability distribution has traditionally been that underlying the previous rate for the classification in question. In experience rating, he takes the manual rate as the mean of the prior distribution, to be modified by the experience of the individual risk. He has never, however, faced the question: "How much confidence do I have in the current rate?"

One way to achieve meaningful results would be to estimate not only the mean but the variance of the present rate level. Then, after choosing a distribution, we can solve for the number of claims to which our prior knowledge is equivalent, which is a function of the mean and variance.

For example, if we believe our data are the result of a Bernoulli process we may assume that $P(H)$ is a Beta function

$$p(h) = K\, h^{r'}(1-h)^{n'-r'}$$

with r' favorable and n'-r' unfavorable outcomes. Then h has mean

$$m = \frac{r'+1}{n'+2}$$

and variance

$$\sigma^2 = \frac{r'+1}{(n'+2)^2} \cdot \frac{(n'-r'+1)}{(n'+3)} = \frac{m\,(1-m)}{n'+3}$$

(See Raiffa and Schlaifer [7] p. 216). Thus

$$n' = \frac{m\,(1-m)}{\sigma^2} - 3$$

represents the validity of the prior knowledge, and a comparison of n' with the number of exposure units in D will indicate the credibility that D deserves relative to H.

If we assume, for example, that the mean accident frequency underlying our present rates is .10 with $\sigma = .005$ then

$$n' = \frac{(.10)\,(.90)}{.000025} - 3 = 3597.$$

If we assume that $m = .10$ but $\sigma = .02$ we would, of course, have much less confidence in H than in the previous example. Here

$$n' = \frac{(.10)(.90)}{.0004} - 3 = 222.$$

If $P(H)$ is a Beta function with parameters n' and r', and if our data has a binomial conditional distribution with parameters n and r, we can approach the credibility problem by treating our prior knowledge H as a sample of size n' and our data as a sample of size n. We will then have

$$m_H = \frac{r'+1}{n'+2} \text{ and } m_D = \frac{r}{n}.$$

If we then combine the two sets of "data" into a single "sample" of size $n + n'$, we have:

$$m_{H|D} = \frac{r+r'+1}{n+n'+2}$$

$$= \frac{n}{n+n'+2} \cdot \frac{r}{n} + \frac{n'+2}{n+n'+2} \cdot \frac{r'+1}{n'+2}$$

which has the form $ZM_D + (1 - Z) M_H$. It should be noted that these expressions for $m_{H/D}$ and for Z are the same as those derived later by means of Bayes Theorem.

It should be noted that n', which measures the validity of the prior knowledge, varies directly with m and inversely with σ. This agrees with our intuitive notion that it takes fewer units exposed to risk to produce a given level of credibility if the claim frequency is high than if it is low. (Here this principle results in greater validity for the underlying experience in a classification with a high claim frequency than in one with a lower frequency.) It also seems logical that the validity of any estimate should vary inversely with its standard deviation, since a larger standard deviation indicates a smaller degree of confidence that the values are clustered around the mean.

USING BAYES THEOREM

Once a prior distribution has been chosen, it is necessary to combine the prior distribution with the distribution of the data, in order to obtain a posterior distribution. This was done in 1950 by Arthur Bailey [14] but, since his notation is rather complicated, it will be helpful to restate his results in terms of modern statistical concepts, in an endeavor to show what assumptions actually underly credibility theory.

Let H be the random variable whose value we would like to estimate and let $p(h)$ be the prior distribution of H, before the data have been obtained. Let D be a random variable whose value can be observed, the data, and $f(H|d)$ the posterior distribution of H, given that $D = d$. Let m_D and m_H be the means of D and H and σ_D^2 and σ_H^2 their respective variances. Let ρ be the correlation coefficient between D and H and σ_{HD} the covariance. $g(D|h)$ is the conditional density of D given that $H = h$. To apply Bayes theorem $p(h)$ and $g(D|h)$ must be known or assumed. $g(D|h)$ will reflect the type of chance variation of the data around the "population parameter" h and $p(h)$, since it is a prior distribution, can be chosen to reflect the actuary's prior knowledge and beliefs about the random variable to be estimated. As we shall see, it is convenient to choose $p(h)$ as a conjugate distribution to $g(D|h)$.

It should be noted that the notion of a correlation coefficient between H and D would not be acceptable in classical statistics, since the former is a "parameter" and the latter a "statistic." In the Bayesian view, however, such a correlation is permissible so long as it makes sense to talk about the joint distribution $f(h,d)$ of H and D.

Bailey suggests that we take as our estimator of H:

$$h' = E(H|d) = \int_{-\infty}^{\infty} h\, f(h|d)\, dh$$

$$= \frac{\int_{-\infty}^{\infty} h\, f(h,d) dh}{f_1(d)}$$

$$= \frac{\int_{-\infty}^{\infty} h\, f(h)\, g(d|h)\, dh}{\int_{-\infty}^{\infty} f(h)\, g(d|h)\, dh}$$

It should be noted that the conditional expectation $E(H|d)$ is a function of d alone. $E(H|d)$ may be called the regression function of H on D (Hogg & Craig [5] p. 212). $E(H|d)$ may not be linear. Let $x\, m_H + y\, d$ be the "best fitting" approximation to $E(H|d)$, i.e. choose x and y to minimize

$$\int_{-\infty}^{\infty} [E(H|d) - x\, m_H - y\, d]^2\, f_1(d)\mathbf{d}d$$

where

$$f_1(d) = \int_{-\infty}^{\infty} g(d|h)\, p(h)\, dh$$

is the marginal distribution of d. The minimum is obtained by taking

$$x = 1 - \frac{m_D}{m_H}\frac{\sigma_{HD}}{\sigma_D^2} = 1 - \frac{m_D}{m_H} \cdot \rho \cdot \frac{\sigma_H}{\sigma_D}$$

and

$$y = \frac{\sigma_{HD}}{\sigma_D^2} = \rho\, \frac{\sigma_H}{\sigma_D}.$$

Thus

$$E(H|d) \approx \left[1 - \frac{m_D}{m_H} \cdot \rho \cdot \frac{\sigma_H}{\sigma_D}\right] m_H + \rho \cdot \frac{\sigma_H}{\sigma_D} \cdot d$$

Let $A = \frac{\rho^2}{y} = \rho\, \frac{\sigma_D}{\sigma_H}$ and $B = m_D - A m_H$, then

$$E(H|d) \approx (1-\rho^2)\, m_H + \rho^2 m_H - \rho^2\, \frac{m_D}{A} + \rho^2\, \frac{d}{A}$$

$$= \rho^2 \left(\frac{d-B}{A}\right) + (1-\rho^2)\, m_H$$

and this result is exact if $E(H|d)$ is linear. A and B are the coefficients of the regression line of D on H. In particular, if $A = n$ and $B = 0$, which will be seen to be the case if $E(D|h) = nh$ (which is true if $g(d|h)$ is either a binomial or a Poisson distribution):

$$E(H|d) = \rho^2 \frac{d}{n} + (1 - \rho^2)\, m_H$$

$$= Z\, m_D + (1 - Z)\, m_H.$$

It can be seen, then, that the Z used by actuaries as the credibility to be given to observed data, when the data are combined with prior knowledge, is the square of the correlation coefficient between H and D (called by some the "coefficient of determination"). It should be noted that Z has the desired property $Z \leqq 1$ and, because the "best fitting" approximation $h' = x\, m_H + yd$ is defined in an analogous fashion to a least squares regression line, the error variance $E(h - h')^2$ is minimized by this choice of Z. The exact form of Z depends, in any particular case, on the conditional distribution $g(d|h)$ and on the prior distribution $p(h)$.

THE BETA-BINOMIAL

If $g(d \mid h)$ is a binomial distribution, as appears to be true in many branches of insurance,

$$g(d \mid h) = \frac{n!}{d!\,(n-d)!}\, h^d (1-h)^{n-d} \qquad (d \text{ "successes" in } n \text{ trials})$$

then

$$E(D \mid h) = \sum_d d \cdot g(d \mid h)$$

$$= nh = Ah + B,$$

hence $A = n$ and $B = 0$ and

$$E(D^2 \mid h) = \sum_d d^2 \cdot g(d|h)$$

$$= \sum_d [d^{(2)} + d]\, g(d \mid h)$$

$$= n(n-1)\, h^2 + nh$$

and

$$\sigma^2_{D|H} = E(D^2|h) - [E(D|h)]^2$$

$$= n\, h\, (1 - h)$$

If we sum over all values of h we get:

$$E(D) = \sum_h E(D \mid h)\, p(h) = \sum_h n\, h\, p(h) = n\, m_H$$

$$\begin{aligned} E(D^2) &= \sum_h E(D^2 \mid h)\, p(h) \\ &= \sum_h n(n-1)\, h^2\, p(h) + \sum_h nh\, p(h) \\ &= n(n-1)\, E(H^2) + n\, m_H \\ &= n(n-1)(\sigma_H^2 + m_H^2) + n\, m_H \end{aligned}$$

and

$$\sigma_D^2 = n(n-1)\,\sigma_H^2 + n\, m_H (1 - m_H)$$

Although nothing has been said so far about the form of $p(h)$, the prior distribution, it will be helpful to take $p(h)$ as a Beta distribution. If $g(d \mid h)$ is binomial, and $p(h)$ has a Beta distribution, $f(h \mid d)$ will also have a Beta distribution. (See Raiffa & Schlaiffer [7] p. 53.)

If

$$g(d \mid h) = \frac{n!}{d!\,(n-d)!}\, h^d (1-h)^{n-d}$$

and

$$p(h) = K h^{r'} (1 - h)^{n'-r'},$$

then

$$\begin{aligned} E(H \mid d) &= \frac{\int_0^1 h g(d \mid h)\; p(h)\; dh}{\int_0^1 g(d \mid h)\; p(h)\; dh} \\ &= \frac{K \int_0^1 h^{r'+d+1} (1-h)^{n+n'-d-r'}\, dh}{K \int_0^1 h^{r'+d} (1-h)^{n+n'-d-r'}\, dh} \\ &= \frac{B(r'+d+2,\quad n+n'-d-r'+1)}{B(r'+d+1,\quad n+n'-d-r'+1)} \\ &= \frac{r'+d+1}{n+n'+2}. \end{aligned}$$

Since $p(h)$ has mean $m_H = \frac{r'+1}{n'+2}$ and variance $\sigma_H^2 = \frac{m(1-m)}{n'+3}$, and since $E(H \mid d)$ is linear in d, we may write:

$$E(H \mid d) = \frac{n}{n+n'+2} \cdot \frac{d}{n} + \frac{r'+1}{n+n'+2}$$
$$= \frac{n}{n+n'+2} \cdot \frac{d}{n} + \frac{n+2}{n+n'+2} \cdot m_H$$
$$= Z\frac{d}{n} + (1-Z)\, m_H,$$

where $Z = \frac{n}{n+n'+2}$. For a fixed n', Z approaches 1 as n gets very large. We may rewrite Z in terms of the mean and variance of the prior distribution:

$$Z = \frac{n\sigma_H^2}{(n+n'+2)\sigma_H^2}$$
$$= \frac{n\sigma_H^2}{(n-1)\sigma_H^2 + (n'+3)\sigma_H^2}$$
$$= \frac{n\sigma_H^2}{(n-1)\,\sigma_H^2 + m_H(1-m_H)}$$
$$= \frac{n}{n+k},$$

where

$$k = \frac{m_H(1-m_H) - \sigma_H^2}{\sigma_H^2},$$

as previously stated.

THE GAMMA-POISSON

If $g(d|h)$ has the Poisson distribution

$$g(d|h) = \frac{(nh)^d\, e^{-nh}}{d!}$$

then:

$$E(D|h) = \sum dg(d|h) = nh$$
$$E(D^2|h) = \sum d^2\, g(d|h) = n^2 h^2 + nh$$
$$\sigma_{D|h}^2 = \text{nh}$$

Summing over all values of h,

$$E(D) = nE(h) = nm_H$$
$$E(D^2) = n^2\, E(h^2) + n\, E(h)$$
$$= n^2\, (\sigma_H^2 + m_H^2) + nm_H$$
$$\sigma_D^2 = n^2\sigma_H^2 + nm_H$$

If we take $p(h)$ to be the Gamma distribution

$$p(h) = \frac{a^r}{\Gamma(r)} \cdot h^{r-1} \cdot e^{-ah}$$

which has mean $\frac{r}{a}$ and variance $\frac{r}{a^2}$, for $h \geq 0$, letting

$$K = \frac{n^d \, a^r}{d! \, \Gamma(r)}$$

we have:

$$E(H|d) = \frac{\int_0^\infty h \cdot p(h) \cdot g(d|h) \cdot dh}{\int_0^\infty p(h) \cdot g(d|h) \cdot dh}$$

$$= \frac{K \int_0^\infty h^{d+r} \cdot e^{-(n+a)h} \, dh}{K \int_0^\infty h^{d+r-1} \cdot e^{-(n+a)h} \, dh}$$

$$= \frac{\int_0^\infty (n+a)^{d+r} \cdot h^{d+r} \cdot e^{-(n+a)h} \cdot (n+a) \cdot dh}{(n+a) \int_0^\infty (n+a)^{d+r-1} \cdot h^{d+r-1} \cdot e^{-(n+a)h} \cdot (n+a) \, dh}$$

$$= \frac{\int_0^\infty x^{d+r} \cdot e^{-x} \cdot dx}{(n+a) \int_0^\infty x^{d+r-1} \cdot e^{-x} \cdot dx}$$

$$= \frac{(d+r)!}{(n+a)\,(d+r-1)!}$$

$$= \frac{d+r}{n+a}\text{, which is linear in } d.$$

Since $p(h)$ has mean $m_H = \frac{r}{a}$ and variance $\sigma_H^2 = \frac{r}{a^2}$, we may rewrite $E(H|d)$ as:

$$E(H|d) = \frac{d+r}{n+a}$$

$$= \frac{n}{n+a} \cdot \frac{d}{n} + \frac{a}{n+a} \cdot \frac{r}{a}$$

$$= Z\frac{d}{n} + (1-Z)\, m_H,$$

where $Z = \frac{n}{n+a}$ and $a = \frac{m_H}{\sigma_H^2}$.

THREE UNSOLVED PROBLEMS

The credibility tables commonly used in the United States are based on the normal approximation to the Poisson distribution. As has been shown by Harwayne [26], Dropkin [23], Hewitt [27] and others, the two parameter negative binomial distribution provides a better fit to the data than the Poisson. This would seem to indicate the need for new credibility tables in many branches of insurance.

Such new tables could be based on the negative binomial or on the Beta-binomial distribution. However, both of these ignore a very important factor, the distribution of claim size. Most credibility formulas in use today measure the credibility of a given number of claims. What is really needed, however, is the credibility of the pure premium, which depends on claim severity as well as claim frequency.

Let $X_1, X_2, \ldots, X_n$ represent the amounts of the n claims that occur during a given time period. Let us assume that the amount of each claim is independent of the size of any other claim (which might not be true, for example, in a class of policies containing an aggregate limit on benefits paid) and that the X_i are identically distributed. Let $F(x)$ represent the distribution function of the amount of a single claim. $F(x)$ is the probability that the amount of a claim is $\leq x$, given that a claim has occurred. Let N be the number of claims occurring during the time period in question and $p(n)$ represent the probability that $N = n$. The distribution of the total amount of claims paid by the company, i.e. the probability that this amount is $\leq x$, is

$$\sum_{n=0}^{\infty} p(n)\, \overset{n*}{F}(x)$$

where $\overset{n*}{F}(x)$ is the n-fold convolution of $F(x)$ with itself.

If we assume that N follows a Poisson or a negative binomial distribution, and $F(x)$ an appropriate claim distribution (which might vary by line of business), the distribution of the total amount of claims or, better still, the distribution of the pure premium, should produce more accurate credibilities than those now in use.

In many branches of property insurance, the distribution of claim size seems to follow a log-normal distribution (Benckert [19] and Bailey [16]). The convolutions of the log-normal, unfortunately, cannot be obtained in closed form. Mathematically, the easiest distribution to use is the Gamma distribution since, if $F(x)$ has a Gamma distribution with parameters (r,a), then $F^{n*}(x)$ has a Gamma distribution with parameters (nr,a). However, the log-normal distribution has greater skewness than the Gamma. Other distributions that may be useful for claim severity are the Pearson Types V and VI and the Pareto distribution, which is a special case of the Pearson Type VI. An analysis of these and many other distributions will be found in Kupper [29].

Unsolved problem number 1 is a statistical problem—to calculate, from insurance company records, the claim distribution $F(x)$ for various branches of property and casualty insurance.

Unsolved problem number 2 is to work out the convolutions, thus determining the joint distribution of claim frequency and claim severity. The Esscher approximation (See Cramér [1] p. 33) is one method of calculating the convolutions. Several methods of numerical integration, using electronic calculators, are described by Bohman and Esscher [20] who used one of these methods, based on the characteristic function of $F(x)$, to compute some claim distributions for life insurance and fire insurance in Sweden.

Unsolved problem number 3 is to obtain the pure premium distribution from the distribution of total claims and use it to compute credibility tables. Presumably it would be possible to choose a prior distribution for the pure premium, obtain some data, and apply Bayes Theorem to compute the posterior distribution from the prior distribution and the conditional distribution of the data. Since the analysis would be more complicated, mathematically, than an analysis involving Poisson, binomial, Beta and Gamma distributions, it may be necessary to use approximate methods. If mathematically more tractable distributions are substituted for the unruly empirical distributions that may result, it will then be necessary to obtain a measure of the error thus introduced.

CONCLUSION

Bayesian statistics, a new approach to the foundations of statistics, has at last enabled the casualty actuary to derive a sound theoretical foundation for his own work in credibility theory and related fields. This paper will have achieved its purpose if it has pointed the way towards the construction of such a foundation and if it has encouraged others to take up the work.

BIBLIOGRAPHY

BOOKS

[1] Cramér, Harald — *Collective Risk Theory,* Skandia Jubilee Volume, Stockholm (1955).

[2] Cramér, Harald — *The Elements of Probability Theory,* John Wiley & Sons (1955)

[3] Freund, John C. — *Modern Elementary Statistics,* 2nd Ed., Prentice-Hall (1960)

[4] Hoel, Paul G. — *Introduction to Mathematical Statistics,* 2nd Ed., John Wiley & Sons (1954)

[5] Hogg, R. V. and Craig, A. T. — *Introduction to Mathematical Statistics,* Macmillan (1959)

[6] Mood, Alexander M. & Graybill, Franklin A. — *Introduction to the Theory of Statistics,* 2nd ed., McGraw-Hill (1963)

[7] Raiffa, Howard and Schlaifer, Robert — *Applied Statistical Decision Theory,* Harvard University (1961)

[8] Savage, Leonard J. — *The Foundations of Statistics,* John Wiley & Sons (1954)

[9] Schlaifer, Robert — *Probability and Statistics for Business Decisions,* McGraw-Hill (1959)

PAPERS

[10] Ammeter, H. — Le Théorie Collective du Risque et l'Assurance de Choses, LIV *Bulletin de l'Association des Actuaires Suisses,* 185 (1954)

[11] Ammeter, H. — Experience Rating, A New Application of the Collective Theory of Risk. II *ASTIN* 261 (1962)

[12] Ammeter, H., Depoid, P. DeFinetti, B. — L'Etude Mathématique des Assurances non viagères dans. L'Europe Continentale Occidentale, I *ASTIN* 46 (1959)

[13] Bailey, A. L. — A Generalized Theory of Credibility. XXXII *PCAS* 13 (1945)

[14] Bailey, A. L. — Credibility Procedures. XXXVII *PCAS* 7 (1950) [See also discussion XXXVII *PCAS* 94 (1950)]

[15] Bailey, R. A. — Discussion of No. 23. XLVII *PCAS* 152 (1960)

[16] Bailey, R. A. — Experience Rating Reassessed. XLVIII *PCAS* 60 (1961)

[17] Bailey, R. A. and Simon, LeR. J. — Two Studies in Automobile Insurance Ratemaking. XLVII *PCAS* 1 (1960) (Reprinted in I *ASTIN* 192)

[18] Beard, R. E. — Analytical Expressions of the Risks Involved in General Insurance. II 15th *TICA* 230 (1957)

[19] Benckert, L. G. — The Lognormal Model for the Distribution of One Claim. II *ASTIN* 9 (1962)

[20] Bohman, H. and Esscher, F. — Studies in Risk Theory With Numerical Illustrations Concerning Distribution Functions and Stop Loss Premiums. *Skandinavisk Aktuarietidskrift* (1963)

[21] De Finetti, Bruno — Probability, Philosophy and Interpretation. *International Encyclopedia of the Social Sciences* (1963)

[22] Delaporte, P. — Un Problème de Tarification de l'Assurance Accidents d'Automobiles Examiné par la Statistique Mathématique. II 16th *TICA* 121 (1960)

[23] Dropkin, L. B. — Some Considerations on Automobile Rating Systems Utilizing Individual Driving Records. XLVI *PCAS* 165 (1959)

[24] Dropkin, L. B. — Automobile Merit Rating and Inverse Probabilities. XLVII *PCAS* 37 (1960)

[25] Edwards, Ward; Lindman, Harold; Savage, Leonard J. — Bayesian Statistical Inference for Psychological Research. 70 *Psychological Review* 193 (1963)

[26] Harwayne, F. — Merit Rating in Private Passenger Automobile Liability Insurance and the California Driver Record Study. XLVI *PCAS* 189 (1959)

[27] Hewitt, C. C., Jr. — The Negative Binomial Applied to the Canadian Merit Rating Plan for Individual Automobile Risks. XLVII *PCAS* 55 (1960)

[28] Hurley, R. L. — A Credibility Framework for Gauging Fire Classification Experience. XLI *PCAS* 161 (1954)

[29] Kupper, Josef — Wahrscheinlichkeitstheoretische Modelle in der Schadenversicherung. *ETH*, Zurich (1962)

[30] Longley-Cook, L. H. — An Introduction to Credibility Theory. XLIX *PCAS* 194 (1962)

[31] Martin, D. B. — Automobile Insurance Rate-Making in Canada. II 16th *TICA* 37 (1960)

[32] Matthews, A. N. — Recent Developments in Private Passenger Automobile Ratemaking in the United States. II 16th *TICA* 102 (1960)

[33] Perryman, F. S. — Some Notes on Credibility. XIX *PCAS* 65 (1932)

[34] Perryman, F. S. — Experience Rating Plan Credibilities. XXIV *PCAS* 60 (1937)

[35] Roberts, L. H. — Credibility of 10/20 Experience as Compared with 5/10 Experience. XLVI *PCAS* 235 (1959)

[36] Savage, Leonard J. — The Foundations of Statistics Reconsidered, *Proceedings of the Fourth Berkeley Symposium on Mathematics and Probability*, p. 575 (1960)

[37] Simon, LeR. J. — The Negative Binomial and Poisson Distributions Compared. XLVII *PCAS* 20 (1960)

[38] Whitney, A. W. — The Theory of Experience Rating. IV *PCAS* 274 (1918)

[39] Wittick, H. E. — The Canadian Merit Rating Plan for Individual Automobile Risks. XLV *PCAS* 214 (1958)