CREDIBILITY FOR SEVERITY

CHARLES C. HEWITT, JR.

"THE UNSOLVED PROBLEM

"In casualty insurance, the inherent hazard of an insured, or of a classification of insureds, is the product of an inherent frequency of loss occurrence and an inherent average amount of loss, and it is the value of this product for which an estimate is desired. Such an estimate must be expressed in terms of the amounts of the individual losses which have occurred and the a priori *knowledge as to average frequencies, average amounts of losses, the distribution of frequencies and loss amounts about such averages and* a priori *knowledge as to the correlation between frequencies of loss and average loss amounts.*

"The expected value, or estimate, of such a product would, no doubt, be more complicated in form than the results obtained for the simpler cases studied herein. The form such an estimate should take would be very desirable information for the actuary to have, even though, at the present time, there is little or no knowledge as to the correlation between frequencies of loss and average loss amounts in casualty insurance. It is the hope of the writer that someone with a knowledge of the statistical behavior of products will undertake the development of the appropriate procedure."

— A. L. Bailey [1]

"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."

— Allen L. Mayerson [9]

This paper accepts the challenge laid down by these two distinguished actuaries.

Credibility Defined

The literature of this Society amply reflects the widespread use of the formula:

$$Z = \frac{n}{n+K}$$

where:

n equals the number of trials or exposure units
K equals a constant (whose derivation will be the principal subject of this discourse)
z equals *credibility* (with range *0 to 1*).

Credibility is applied in the following manner:

1) there is an hypothesis concerning the mean expectation from some observations,
2) for n trials or exposure units a mean value of the observations is established,
3) with two values to choose from, the question is asked, "To what extent do we believe the expectation and to what extent do we believe the observations?",
4) the degree of belief in the observations is expressed as a measure z, and the degree of belief in the hypothesis as the complement, $1 - z$,
5) or more formally in the linear relationship:

$$C = zR + (1 - z)H$$

where:

R equals the mean of observation (*R*esult)
H equals the mean of the *H*ypotheses, and
C equals the value to be used as a *C*ompromise estimate

Thus, *credibility is a linear estimate of the true* (or inherent) *expectation derived as the result of a compromise between hypothesis and observation.*

Bühlmann [3] has demonstrated that:

$$K \text{ equals } \frac{\text{Expected value of the process variance}}{\text{Variance of the hypothetical means}}$$

The implications of this derivation of K are discussed more fully in [8].

Credibility and Bayesian Estimation

Before proceeding further it will be helpful to underscore the relationship between credibility and *a posteriori* (or Bayesian) estimation and to weigh their relative advantage and disadvantages.

Credibility does not (necessarily) produce the optimum estimate.

Bayesian analysis produces the optimum estimate.

Credibility does produce the "least-squares" fit to the optimum (Bayesian) estimates for all possible outcomes weighted by the respective probabilities of those outcomes.

Both estimates — credibility and Bayesian — are "in-balance" for all possible outcomes.

The estimate resulting from the application of credibility always falls on or between the hypothetical mean and the observed result. The Bayesian resultant frequently does not satisfy this condition. Hence, the use of credibility often produces results more easily explained to the layman, whether he be a customer or an underwriter.

On the other hand the Bayesian resultant can never fall outside the range of hypotheses, whereas the credibility-produced estimate can fall outside the "realm of possibility," although such a happening is unlikely.

Determination of the Bayesian estimate can be extremely complex, even on one trial, and is predictably too complex to handle for more than a few trials (or exposure units). The theoretical part of using credibility is encompassed in the fixing of the value for K; once K is determined, credibility may be applied by a clerk and understood by virtually anyone concerned.

The Risk Process

The basic process in risk and insurance is the compounding of a number of events or occurrences (labelled claims) with a value assigned to each separate event or occurrence (called the amount of the claim). The value of the number of occurrences is discrete: 0, 1, 2, 3, ..., k, ... The amount assigned to each occurrence may be constant, or it may vary over a wide range, often considered to be a continuum for convenience of analysis. The compound process is discussed and expressions for moments of the compound process are derived in [10]. For purposes of this section it will be sufficient to draw the following expressions adapted from [10]:

$$_{x}E = {_{k}E} \cdot {_{x_1}E}$$

and
$$_{x}\sigma^2 = {_{k}E} \cdot {_{x_1}\sigma^2} + {_{k}\sigma^2} \cdot {_{x_1}E^2}$$

Where the prefatory subscripts have the following significance:

x — the compound process (or distribution)

k — the discrete process of determining the number of occurrences

x_1 — the distribution of the values of a single claim

Thus $_{k}E$ is the mean number of occurrences, $_{x_1}\sigma^2$ is the variance of the amounts of a single claim, etc. Of most importance are $_{x}E$ and $_{x}\sigma^2$ which are the mean and variance of the compound process.

The inhibitions placed upon the use of the expression for the mean and variance of the compound process are:

(a) the value attributed to each separate occurrence is independent of all other values so attributed,

(b) the values so attributed are drawn from the same probability distribution,

(c) the number of occurrences is statistically independent of the values attributed to the occurrences.

There are circumstances in practice under which these inhibitions are breached — non-reporting of smaller claims as opposed to the more likely reporting of larger claims is clearly in this category. However, there are many situations in insurance for which these inhibitions are not violated to any important degree. Judicious selection of the proper event-producing process and/or what properly constitutes a single event (or claim) will generally provide adequate reassurance that the expression for the compound process variance is quite satisfactory.

It is important, nay vital, to realize that these inhibitions which qualify the "process variance" do not, in any way, affect the "variance of the hypothetical means," which depends only upon defining the possible states and quantifying their *a priori* probabilities.

Two Illustrative Examples

In order to bridge the gap between the theoretical derivation by Bühlmann [3] of the credibility "K" and actual application thereof, two examples will be used to explain how the "variance of the hypothetical means" and the "expected value of the process variance" may be derived. In so doing the theoretically correct method for inclusion of the severity component will also be illustrated — for the first time as far as this author is aware.

A Discrete Example

A die is selected at random from a pair of "honest" dice. It is known that one die has one marked face and five unmarked faces and that the other die has three marked and three unmarked faces.

For reference we will define:

A_1 as the state of having drawn the die with one marked and five unmarked faces

A_2 as the state of having drawn the die with three marked and three unmarked faces

A spinner is selected at random from a pair of spinners.

It is known that one spinner has six equally-likely sectors five of which are marked *two* and one of which is marked *fourteen*, and that the other spinner has six equally-likely sectors three of which are marked *two* and three of which are marked *fourteen*.

For reference we will define:

B_1 as the state of having selected the spinner with five *twos* and one *fourteen*

B_2 as the state of having selected the spinner with three *twos* and three *fourteens*.

Initially it will be specified that the selection of the die and the spinner are completely independent.

Thus there are four *equally-likely* compound states:

$$A_1 \cap B_1$$

$$A_1 \cap B_2$$

$$A_2 \cap B_1$$

$$A_2 \cap B_2$$

The state once determined will remain the same throughout, but will be *unknown* to the participants.

First, the die which was drawn will be rolled. If a marked face appears uppermost this constitutes a claim; if not, there is no claim. If there is a claim, the selected spinner will be spun to determine the amount of the claim.

The process of rolling the die, once drawn, is assumed to be binomial, i.e either marked or unmarked face appears.

A Continuous Example

A private passenger automobile insurance risk will be chosen from a class of such risks. Each risk in the class has its own inherent measure of hazard which can never be exactly known to an insurer.

For frequency of occurrence of a claim we will define:

M as the state of having chosen a risk with the inherent frequency of claims in one exposure unit (one car-year) m (see [4], [5] and [6] for an earlier treatment of this situation)

The *a priori* probability of having chosen M is given by the gamma distribution:

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

where m varies between *zero* and (positive) infinity.

The process by which the number of claims is given is assumed to be Poisson for the particular risk, M. Thus the probability of n claims in one car-year is given by:

$$Pr(n) = \frac{m^n}{n!} e^{-m}$$

where n is any non-negative integer.

It is important to distinguish between the frequency process for the individual risk (Poisson as already stated) and the frequency process for the class of risks from which it is drawn. The latter process will be negative binomial as developed in [4], [5] and [6].

For severity of an individual claim, use will be made of a distribution of some auto property damage losses by size which are fitted quite precisely by a compounding of the log-normal and gamma distributions. (A paper describing the method of fit and the accuracy thereof is in preparation.) For simplicity here it will be assumed that the distribution of a single property damage claim (X) follows the log-normal pattern.

We will define:

θ as the state of having chosen a risk with the inherent severity (average amount of a single claim):

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

where the loss amounts for the risk θ are distributed log-normally:

$$\Lambda(X; \mu, \sigma)dX = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$x = \log_e X$ and varies from negative infinity to positive infinity. The *a priori* probability of having chosen μ is given by the normal distribution:

$$N(\mu; N, S)d\mu = \frac{1}{S\sqrt{2\pi}} e^{-\frac{(\mu-N)^2}{2S^2}} d\mu$$

where μ varies from negative infinity to positive infinity and it is assumed that σ does not vary from risk to risk in this particular class of risks.

(Appendix A demonstrates that the class distribution of amounts of a single claim (Y) is also log-normal:

$$\Lambda(Y; N, S, \sigma)dY = \frac{1}{\sqrt{2\pi(S^2+\sigma^2)}} e^{-\frac{(y-N)^2}{2(S^2+\sigma^2)}} dy$$

for the whole class.)

$$y = \log_e Y$$

Thus amounts of a single claim are distributed log-normally for both individual risks and the whole class. It is important, however, to distinguish among the parameters of the respective log-normal distributions.

For simplicity at this point it will be assumed that the inherent risk parameters m and μ are completely independent.

Thus there is an infinitude of compound states:

$$M \cap \theta$$

whose likelihood is given by the distribution:

$$T(m) \cdot N(\mu; N, S)d\mu dm$$

The inherent hazard of the risk will be assumed to remain the same throughout.

Exposure Units

Discrete Example — One roll plus, if necessary, one spin to determine the amount of the claim.

Continuous Example — One car-year.

Moments of Frequencies and Severities

Discrete Example

(Frequency) The probability of a claim for state A_1 is $1/6$ and for state A_2 is $3/6$. Remembering that the variance in the binomial process is given by *npq* and that *n* equals *one* for one for one exposure unit, the respective variances are:

A_1: $1 \cdot (1/6) \cdot (5/6) = 5/36$

A_2: $1 \cdot (3/6) \cdot (3/6) = 9/36 = 1/4$

Summarizing in the special notation adopted for the risk process:

State	${}_kE$	${}_k\sigma^2$
A_1	$1/6$	$5/36$
A_2	$3/6$ (or $1/2$)	$1/4$

(Severity) When a claim occurs the amount thereof is determined by use of the spinner. If B_1 has been selected, then the key moments of a single claim may be calculated as follows:

Amount of Claim (X)	Probability of Amount $f(X)$	$X\,f(X)$	$X^2\,f(X)$
2	5/6	5/3	10/3
14	1/6	7/3	98/3
Total	1	4	36

Thus the mean for state B_1 is *4* and the variance (mean of the squares less the square of the mean) is *20*.

In a similar manner it can be shown that the mean and variance of the amount of a single claim for state B_2 are *8* and *36* respectively. Summarizing in the special notation adopted for the risk process:

State	${}_{x_1}E$	${}_{x_1}\sigma^2$
B_1	4	20
B_2	8	36

Continuous Example

(Frequency) The frequency of property damage claims in one car-year for a risk with state M is m and since the process for an individual risk is Poisson the variance is also m.

Summarizing:

State	${}_kE$	${}_k\sigma^2$
M	m	m

(Severity) The severity (average amount of a single claim) for a risk with state θ is:

$$e^{\mu+\frac{\sigma^2}{2}}$$

And, for the log-normal distribution, the variance of a single claim (for state θ) is:

$$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$$

Summarizing:

State	${}_{x_1}E$	${}_{x_1}\sigma^2$
θ	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$

(The reader will surely recognize that the "sigmas" in the risk process notation are *not* the same as the "sigmas" in the log-normal distribution. There is a need to compromise between distinguishing symbols and still using familiar notation.)

Pure Premium

The product of frequency and severity (for one exposure unit) is commonly referred to as the pure premium (Π). When multiplied by the number of exposure units the pure premium indicates the charge necessary to cover expected losses. For the risk process the pure premium is given by:

$$\Pi = {}_xE = {}_kE \cdot {}_{x_1}E$$

Discrete Example

Table 1

Mean and Variance
of the
Pure Premium

State	Probability of State	Frequency	Severity	Pure Premium	Square of Pure Premium
$A_i \cap B_j$	$f(A_i \cap B_j)$	${}_kE$	${}_{x_1}E$	Π	Π^2
(1)	(2) 1/4*	(3) (As calculated)	(4)	(5) (3) x (4)	(6) (5) squared
$A_1 \cap B_1$	1/4	1/6	4	2/3	4/9
$A_1 \cap B_2$	1/4	1/6	8	4/3	16/9
$A_2 \cap B_1$	1/4	1/2	4	2	4
$A_2 \cap B_2$	1/4	1/2	8	4	16
Total	1	Weighted — by (2) — Total		2	50/9

* Remember that each state in the dice-spinner example is equally likely.

Thus, if one knew that the true state was $A_2 \cap B_2$, one would charge a pure premium of *4*. Since one never knows (in this example) which state one is dealing with, one would start *a priori* by charging the mean (of the hypotheses) pure premium of *2*.

The variance of the hypothetical means (mean of the squares less the square of the mean) is (14/9).

Continuous Example

Recalling that ${}_kE = m$ and ${}_{x_1}E = e^{\mu + \frac{\sigma^2}{2}}$

the pure premium for the state $M \cap \theta$ is:

$$\Pi = m \cdot e^{\mu + \frac{\sigma^2}{2}}$$

Also recalling that the probability of state $M \cap \theta$ is:

$$T(m) \cdot N(\mu; N, S)\, d\mu dm$$

the mean (class) pure premium is given by:

$$E(\Pi) = \int_0^\infty mT(m) \left[\int_{-\infty}^{\infty} e^{\mu + \frac{\sigma^2}{2}} N(\mu; N, S)\, d\mu \right] dm$$

$$E(\Pi) = \frac{r}{a}\; e^{N + \frac{S^2+\sigma^2}{2}}$$

In a similar manner the mean of the pure-premium-squared is given by:

$$E(\Pi^2) = \int_0^\infty m^2 T(m) \left[\int_{-\infty}^{\infty} e^{2\mu+\sigma^2} N(\mu; N, S) d\mu \right] dm$$

$$E(\Pi^2) = \frac{(r+1)r}{a^2}\, e^{2N+2S^2+\sigma^2}$$

Thus the *variance of the hypothetical means* is:

$$\sigma^2(\Pi) = \frac{r}{a^2}\, e^{2N+S^2+\sigma^2} [(r+1)e^{S^2} - r]$$

Expected Value of the Process Variance

In the risk process each separate state has its own process. Recall that the variance of the compound process is given by:

$${}_x\sigma^2 = {}_kE \cdot {}_{x_1}\sigma^2 + {}_k\sigma^2 \cdot {}_{x_1}E^2$$

Discrete Example

Table 2

Process Variance

State	Probability of State	Frequency: Mean	Frequency: Variance	Severity: Variance	Severity: Mean-Squared	Process Variance
$A_i \cap B_j$	$f(A_i \cap B_j)$	${}_kE$	${}_k\sigma^2$	${}_{x_1}\sigma^2$	${}_{x_1}E^2$	${}_x\sigma^2$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
	1/4	(As calculated)		(As calculated)		[(3) × (5)] + [(4) × (6)]
$A_1 \cap B_1$	1/4	1/6	5/36	20	16	50/9
$A_1 \cap B_2$	1/4	1/6	5/36	36	64	134/9
$A_2 \cap B_1$	1/4	1/2	1/4	20	16	14
$A_2 \cap B_2$	1/4	1/2	1/4	36	64	34
Total	1	Weighted — by (2) — Total				154/9

It is pertinent to note at this point that the *process variance* of the individual states does not change, but, as actual experience is obtained, the probabilities of the individual states may change. New probabilities, after experience is obtained, are *a posteriori.* The mean of the process variances using the *a priori* probabilities is the *expected value of the process variance* — in this example $(154/9)$

Continuous Example

Recalling that $\quad {}_kE = {}_k\sigma^2 = m \quad$ and that

$$ {}_{x_1}\sigma^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1), \quad \text{and} \quad {}_{x_1}E^2 = e^{2\mu+\sigma^2} $$

$$ {}_x\sigma^2 = m[e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) + e^{2\mu+\sigma^2}] $$

$$ {}_x\sigma^2 = m \cdot e^{2(\mu+\sigma^2)} $$

This is the process variance for each individual risk with state $M \cap \theta$. Weighting each such variance by the *a priori* probability of the respective states and integrating over m and μ is:

$$ E({}_x\sigma^2) = \int_0^\infty mT(m)\left[\int_{-\infty}^{\infty} e^{2(\mu+\sigma^2)}N(\mu; N, S)d\mu\right]dm $$

$$ E({}_x\sigma^2) = \frac{r}{a}\, e^{2(N+S^2+\sigma^2)} $$

(Expected value of the process variance)

Credibility "K"

From the definition of credibility (Bühlmann)

$$ K = \frac{\text{Expected value of the process variance}}{\text{Variance of the hypothetical means}} $$

Discrete Example

$$ K = \frac{(154/9)}{(14/9)} = 11, $$

Now

$$ Z = \frac{n}{n+K} $$

so that credibility for one exposure unit (one roll — and spin, if required)

$$ Z = \frac{1}{1+11} = \frac{1}{12} $$

and for two rolls — 2/13, three rolls — 3/14, etc.

Continuous Example

$$K = \frac{\frac{r}{a} e^{2(N+S^2+\sigma^2)}}{\frac{r}{a^2} e^{2N+S^2+\sigma^2}[(r+1)e^{S^2} - r]} = \frac{ae^{S^2+\sigma^2}}{(r+1)e^{S^2} - r}$$

If the individual risks in the class being investigated were assumed to have the same severity, then S would be zero and K would reduce to:

$$ae^{\sigma^2}$$

Furthermore, if severities were ignored and the same amount used for each claim, then σ would be zero and K would be further simplified to a, which is the same value determined in [6] and [8] when severities had not yet been introduced into credibility formulas.

It should not be inferred that theoretical loss distributions are necessary for this method to work. First and second moments of raw data may also be used as estimators.

Recapitulation

This concludes the major thesis of this paper:

To demonstrate how to determine theoretically correct credibilities for the pure premium by making use of the Bühlmann definition of credibility and the formula for the variance of the compound process.

The steps involved are summarized below:

To Determine the Variance of the Hypothetical Means

(1) Enumerate all of the possible states,

(2) Assign an *a priori* probability to each state,

(3) For *each state separately* assign a mean of the number of occurrences or events (labelled claims) per exposure unit,

(4) For *each state separately,* assign a mean value for one occurrence of an event (labelled a claim),

(5) The product of the values assigned in (3) and (4) for *each state separately* is weighted by the *a priori* probability of that state,

(6) The sum of the weighted products in (5) is the mean of the hypothetical means, elsewhere referred to as H, the mean of the hypotheses,

(7) The products described in (5) are squared for *each state separately* and weighted by the *a priori* probability of that state,

(8) The sum of the weighted products in (7) less the square of the mean of the hypothetical means is the *variance of the hypothetical means*, as used in the expression for K in the credibility formula.

To Determine the Expected Value of the Process Variance

Utilize the values obtained in Steps (1) through (4) in determining the variance of the hypothetical means.

(5) Square the value obtained in (4) for *each state separately,*

(6) Obtain the variance of the mean number of occurrences, per exposure unit, and the variance of the amount of a single claim for *each state separately,*

(7) Obtain the product of the value in (3) with the variance of the amount of a single claim for *each state separately,*

(8) Similarly, obtain the product of the value in (5) with the variance of the mean number of occurrences, per exposure unit, for *each state separately,*

(9) The sum of the products for *each state separately* obtained in (7) and (8) is the process variance for each state respectively,

(10) The sum of the process variances weighted by the *a priori* probability of each respective state is the *expected value of the process variance.*

The balance of the paper is devoted to applications of this new approach and to a comparison of credibility results with the results of Bayesian estimation.

Credibility vis-a-vis Bayesian Estimation

While the discrete example of the dice and the spinners is fresh, it is instructive to compare credibility with the results of Bayesian analysis. To start, the probabilities of obtaining all possible outcomes under all possible states are set forth in Table 3.

Table 3
Probabilities of States and Outcomes

State	Outcome for One Trial in 144ths: 0	2	14	Marginal Total
$A_1 \cap B_1$	30	5	1	36
$A_1 \cap B_2$	30	3	3	36
$A_2 \cap B_1$	18	15	3	36
$A_2 \cap B_2$	18	9	9	36
Total	96	32	16	144

To illustrate the method of determining the values in the above table:

Probability of state $A_2 \cap B_1$ equals $1/4$

Probability of a claim given this state equals $3/6$

Probability of claim amount "2" given a claim and given this state equals $5/6$

Therefore, Probability $\underline{(2 \cap A_2 \cap B_1)}$ equals $1/4 \times 3/6 \times 5/6$ equals $15/144$

The inverse, or Bayesian, probabilities derived from the above table given the outcome of one trial but not knowing the true state, are obtained by dividing the individual cell probabilities by the probability of the outcome (column) in which the cell falls. The results, plus a refresher on the pure premiums of each individual state, are given in Table 4.

Table 4
Inverse Probabilities

State	Given Outcome: 0	2	14	Pure Premium [Table 1, Column (5)]
$A_1 \cap B_1$	5/16	5/32	1/16	2/3
$A_1 \cap B_2$	5/16	3/32	3/16	4/3
$A_2 \cap B_1$	3/16	15/32	3/16	2
$A_2 \cap B_2$	3/16	9/32	9/16	4
Total	1	1	1	

The probabilities in each of the "Given Outcome" columns represent the *a posteriori* probabilities of the respective individual states if the outcome is as given. They are, then, a revision of the *a priori* state probabilities found in Column (2) of Tables 1 and 2. The pure premiums in the last column above weighted by the *a posteriori* probabilities in each respective "Given Outcome" column are the Bayesian estimates, e.g. for the outcome "2":

$$\text{Pure Premium/Given "2"} = \frac{(5 \times 2) + (3 \times 4) + (15 \times 6) + (9 \times 12)}{96}$$

$$= \frac{55}{24} = 2\tfrac{7}{24}$$

In a similar manner the Bayesian estimates for the outcomes "0" and "14" are $1\frac{3}{4}$ and $2\frac{11}{12}$ respectively.

It is interesting to note that credibility has a stochastic aspect. For one could use the *a posteriori* probabilities of the individual states to recalculate the *variance of the hypothetical means* and the *expected value of the process variance* and hence "K." The new "least squares" line so determined would, of necessity, pass through the new Bayesian estimate of the pure premium ($2\frac{7}{24}$ instead of the *a priori* 2, if the outcome had been a "2"), and would no longer necessarily produce the same estimate after a second trial as the original "K" would for *two* trials. In practice it is doubtful if the stochastic approach would be used.

However, the purpose of this section is to compare the credibility-produced estimates with the Bayesian pure premiums. This is done in Table 5.

Table 5
Pure Premiums
Bayesian vs. Credibility-Produced

Outcome	Credibility-Produced Estimate	Bayesian Estimate	Square of the Differences
(1)	(2)	(3)	(4)
	[(1/12) × (1)] + [(11/12) × 2]	See Narrative	[(2) − (3)] squared
0	11/6	7/4	1/144
2	2	55/24	49/576
14	3	35/12	1/144

The reader who has persisted with this discrete example this far is encouraged to weight the "difference-squares" in Column (4) above by the probabilities of the outcomes from Table 3: 2/3, 2/9, and 1/9 respectively — take the sum and verify for himself by trying alternatives that the credibility of 1/12 does, in fact, produce a "least squares" fit to the Bayesian pure premiums in Column (3).

Finally, it should be observed in this example that, as was pointed out much earlier, the Bayesian estimate (55/24) does not fall between the observed result (2) and the hypothetical mean (also 2). While this is easy enough for the probabilist to understand, it is awfully difficult to interpret for a layman.

Interdependence of Frequency and Severity

In the discrete example it was originally specified that the selection of the die (determining frequency) and the spinner (determining severity) are completely independent. Such a specification is not necessary in order to obtain credibility "K" using the method described herein.

In fact, assume that state B_1 can only be associated with state A_1 and that state B_2 can only be associated with state A_2 — total interdependence of frequency and severity. It is only necessary to make a change in the state probabilities in the second column of Tables 1 and 2. ($A_1 \cap B_1$ becomes 1/2 as does $A_2 \cap B_2$, while $A_1 \cap B_2$ and $A_2 \cap B_1$ become zero.) Greater familiarity with the method could be obtained by the reader by making this substitution and carrying out the balance of the steps. (For the record, the new pure premium is $2\frac{1}{3}$ and K equals 7.12.)

The important point is that hypothetical frequencies and severities may be interrelated without vitiating this method of determining the credibility of the pure premium.

Auto Merit Rating — Application of Method (See Continuous Example)

In private passenger automobile insurance the theory with respect to merit rating ([5] and [6]) is pretty well established, if severity is ignored entirely. The connection between merit rating and credibility has been pointed out ([2] and [7]), but for frequency of occurrence only.

When severity is ignored, it has been shown in [6] and [8] that credibility for Canadian private passenger data — Class 1 (Adult-Pleasure Use) — is determinable from the parameters:

$$r = 2.62$$

$$(K =) \quad a = 30.1$$

$$\frac{r}{a} = .08704 \quad \text{(frequency-class)}$$

$$Z = \frac{1}{1 + 30.1} = .032$$

Bringing in the additional dimension of severity, for para-realistic automobile property damage data (log-normally distributed-Appendix A):

$$N = 5.289$$

$$S^2 + \sigma^2 = 0.738$$

$$S^2 = 0.01932$$

$$e^{N + \frac{S^2 + \sigma^2}{2}} = \$286.60 \quad \text{(severity-class)}$$

$$E(\Pi) = \frac{r}{a}\, e^{N + \frac{S^2 + \sigma^2}{2}} = \$24.95 \quad \text{(pure premium-class)}$$

From the continuous example:

$$K = \frac{a e^{S^2 + \sigma^2}}{(r + 1) e^{S^2} - r} = \frac{(30.1) e^{0.738}}{(3.62) e^{.01932} - (2.62)} = 58.8$$

$$Z = \frac{1}{1 + 58.8} = .017$$

So the second dimension (severity) has the effect of *halving* credibility *in this instance.* For a coverage with wider dispersion of loss values (of a single claim), say bodily injury, there would have been an even greater reduction in credibility.

Rating Plans With Normal/Excess Loss Splits

Workmen's Compensation Insurance has a multi-split experience rating plan and many forms of commercial insurance have single-split experience rating plans. As a change-of-pace from symbols, the results for a single-split experience rating plan have been calculated using the same data as in the previous section to illustrate how a plan with (complete) credibility, i.e. for amounts as well as occurrences, might work. (See Appendix B for complete details.)

To illustrate the effect of splitting losses upon credibility values it was assumed that all risks would have the same size-of-loss distribution. It would be possible with a computer and Monte Carlo methods to determine K for variable size-of-loss distributions with splitting, but this is beyond the needs of this paper.

How would experience rating work for the values thus obtained? Table 6 below sets forth the credibilities and experience-rated pure premiums for *no losses*, and for a *single loss* in the amount of $100, $500, $1,000 and $3,000 respectively:

Table 6
Class Pure Premium $24.95
Experience-Rated Pure Premium

Amount of (one) Loss	Loss Split-Point $50	$100	$250	$500	$1,000	None
(z-normal)	.03184	.03095	.02739	.02298	.01901	.01557
(z-excess)	.01088	.00757	.00289	.00083	.00014	0
$ 0	$24.59	$24.57	$24.50	$24.47	$24.50	$24.56
100	26.73	27.67	27.23	26.77	26.40	26.12
500	31.08	30.70	32.07	35.96	34.00	32.35
1,000	36.52	34.49	33.51	36.37	43.51	40.13
3,000	58.28	49.63	39.29	38.03	43.79	71.27

This table tells only a small part of a much larger story. For example, if the amounts of loss used above were divided among two or more separate occurrences, the resulting experience-rated pure premium would be different. However, it is clear that the individual risk suffers most (relative to the amount of loss) when the single loss is right at the split point. Also the inconsistency in values, read horizontally for a particular amount of loss, illustrates the fact that credibility does not necessarily produce an optimum estimate but rather is a "least squares" value fitted to a series of optimum estimates. It also shows the inconsistency of subsuming credibility into *normal* and *excess*, thus implying that there is no interrelationship between the empirical number and amount of losses in each category.

Conclusion

Credibility is theoretically justifiable and eminently practicable when amount of loss is considered in addition to frequency of occurrence. The results produced by so using credibility are "least squares" approximations to Bayesian estimates.

The necessary "tools" are:

1) *A priori* probabilities of all possible states (frequency and severity may or may not be interdependent),
2) The *first* and *second moments* for each state of the —
   a) Discrete process which determines the number of occurrences, and of the —
   b) Amount of a single claim.

APPENDIX A

Log-Normal (Class) Distributions by Size-of-Loss

Given X = amount of claim:
$x = \log_e X$

then the p. d. f. of X is:

$$\Lambda(X;\mu,\sigma)dX = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

Also given that μ varies from risk to risk according to:

$$N(\mu;N,S)d\mu = \frac{1}{S\sqrt{2\pi}}\, e^{-\frac{(\mu-N)^2}{2\sigma^2}}\, dx$$

While σ does not vary from risk to risk.

$$Pr(X) = \sum_{\mu} Pr(X \mid \mu)\cdot Pr(\mu)$$

$$= \int_{-\infty}^{\infty} \Lambda(X;\mu,\sigma)N(\mu;N,S)d\mu$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\cdot \frac{1}{S\sqrt{2\pi}}\, e^{-\frac{(\mu-N)^2}{2S^2}}\, d\mu$$

Combining exponents and completing the square produces:

$$Pr(X) = \frac{e^{-\frac{(x-N)^2}{2(S^2+\sigma^2)}}}{\sqrt{2\pi(S^2+\sigma^2)}}\int_{-\infty}^{\infty} \frac{1}{\frac{\sigma S}{\sqrt{S^2+\sigma^2}}\sqrt{2\pi}}\, e^{-\frac{\left(\mu-\frac{S^2x+\sigma^2N}{S^2+\sigma^2}\right)^2}{2\frac{\sigma^2S^2}{S^2+\sigma^2}}}\, d\mu$$

$$= \frac{1}{\sqrt{2\pi(S^2+\sigma^2)}}\, e^{-\frac{(\mu-N)^2}{2(S^2+\sigma^2)}} \quad \text{(also log-normal)}$$

where S^2 is the variance of the individual μ's about N, the mean of μ's.

As mentioned earlier some special-purpose auto property damage data, available to the writer will be used for illustrative purposes. For the log-normal fitted to this data the parameters are:

$$N = 5.289 \quad ; \quad (S^2 + \sigma^2) = 0.738$$

For purposes of illustrating method arbitrarily assume that the severity of an individual risk, with a μ which is $2S$ below N, is equal to 75% of the average severity of the class. Then:

$$e^{N - 2S + \frac{\sigma^2}{2}} = 75\% \; e^{N + \frac{S^2 + \sigma^2}{2}}$$

and: $$-2S = \log_e 0.75 + \frac{S^2}{2}$$

and solving for S: $S = 0.139$

$$S^2 = 0.01932$$

APPENDIX B

A Numerical Illustration for Normal/Excess Split Plans

From Appendix A assume, for simplicity, that all risks in a class have the same size-of-loss distribution. Then $S = 0$ and:

$$N = 5.289$$

$$\sigma^2 = 0.738$$

And from the paper itself:

$$r = 2.62$$

$$a = 30.1$$

making for the class of risks:

$$\text{frequency} = \frac{r}{a} = .08704$$

$$\text{severity} = e^{N + \frac{\sigma^2}{2}} = \$286.60$$

$$\text{pure premium} = \frac{r}{a} e^{N + \frac{\sigma^2}{2}} = \$24.95$$

$$K = ae^{\sigma^2} = 63.0 \quad \text{(no split)}$$

Table B-1 and the explanation which follows show how credibility should be calculated for a single split plan.

TABLE B-1

KEY PARAMETERS

Split Point	Cumulative Frequency	$_{x_1}E$	$_{x_1}E^2$	$_{x_1}\sigma^2$	$_x\sigma^2$	$\sigma^2(\pi)$	K
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X	$F(X)$	from Log-Normal*	$(3)^2$	from Log-Normal*	$(.087) \times [(4)+(5)]$	$[(.00288) \times (4)]$	$\frac{(6)}{(7)}$
			Normal (Below Split-Point)				
$ 50	.05447	$ 49.26	2,427	14.32	212.40	6.99	30.4
100	.2131	92.72	8,597	297.40	773.80	24.76	31.3
250	.6064	178.05	31,702	5,577.00	3,243.00	91.30	35.5
500	.8595	238.40	56,835	23,142.00	6,958.00	163.68	42.5
1,000	.97029	272.60	74,311	52,678.00	11,048.00	214.00	51.6
No Limit	1.000	286.60	82,140	89,640.00	14,945.00	236.60	63.0
			Excess (Above Split-Point)				
	$1-(2)_N$				$[(.087) \times (2)]$ $[(4)+(5)]$	$(.00288) \times$ $(2)^2 \times (4)$	
$ 0	1.000	286.60	82,140	89,640	14,945	236.60	63.0
50	.94553	251.00	63,001	116,093	14,739	162.22	90.9
100	.7869	246.40	60,713	146,284	14,179	108.25	131.0
250	.3936	275.80	76,066	265,654	11,687	33.94	344.3
500	.1405	343.10	117,718	535,686	7,985	6.692	1,193
1,000	.02971	471.20	222,029	1,285,577	3,890	0.5644	6,892

* Calculation not reproduced here

The numbers in Table B-1 have very little value per se, but the method of computing *K* for split losses is illustrated. Some explanation is required, particularly for anyone unfamiliar with rating plans in use in the United States on commercial insurance. "Normal losses" for a split-point of, say, $500, would be the total amount of any claim below $500 and the first $500 on any claim above that point. "Excess losses" would be that portion of any loss over $500 in excess of $500.

Table B-1, Column (2) indicates that 85.95% of all losses will be $500 or less and consequently that only 14.05% (see under Excess) of all losses will be excess losses. Thus the frequency of excess losses is obtained by multiplying the regular frequency of (.087) by the probability that, given a loss, it is an excess loss. It has been stated by Verbeek [11] that excess losses taken from a Poisson process also follow a Poisson process. Thus the variance in Column (7) must be multiplied by the square of the probability of an excess loss in Column (2).

Column (3) indicates that the average value of a loss limited to $500 is $238.40 and also that the 14.05% of losses which are excess have an average (excess) value of $343.10. Columns (4) and (5) are self-explanatory. Column (6) illustrates the calculation of the *expected value of the process variance.* Since all risks have the same severity the variance of the hypothetical means can be obtained (Column (7)) by multiplying the variance of the hypothetical frequencies by the square of the average severity. By now Column (8) should be self-explanatory also.

In summary, if claim amounts are disregarded, *K* has the value *a* of 30.1. For low split-points — $50 and $100 — the effect is not much different from just counting claims as far as normal losses are concerned. But as the split-point is increased *K* increases and credibility given to normal losses would therefore decrease, until with no limit on the split-point *K* equals the previously calculated value of *63.0* for a no-split rating plan.

On excess losses, *K* starts out at *63.0* with the split-point at $0, as might have been expected. However, when the split-point increases the credibility for excess loss approaches the vanishing point (*K* equals 1,193 for excess of $500 losses).

In summary, credibility is greatest when severity is ignored entirely (as has been the case in the past); when severity is introduced, credibility can be retained by limiting the value for which a loss enters the rating, but credibility decreases as more and more of the value of the individual claim enters the rating until it reaches a fixed value when all loss amounts are subject to inclusion. If losses above a certain value (excess losses) are rated, credibility has a maximum at the same fixed value applying to the rating of all losses and then decreases to zero as the excess point moves upward toward infinity.

REFERENCES

[1] Bailey, Arthur, L. "Credibility Procedures" (shortened title), *PCAS* XXXVII (1950), pp. 7-23

[2] Bailey, Robert A. Review — "Some Considerations on Automobile Rating Systems" (shortened title), Dropkin, *PCAS* XLVII (1960), pp. 152-156

[3] Bühlmann, Hans "Experience Rating and Credibility," IV *ASTIN Bulletin* (1967), pp. 199-207 (particularly p. 207)

[4] Dropkin, Lester "Automobile Rating Systems" (shortened title), *PCAS* XLVI (1959), pp. 165-176

[5] Dropkin, Lester "Automobile Merit Rating and Inverse Probabilities," *PCAS* XLVII (1960), pp. 37-40

[6] Hewitt, Charles C., Jr. "The Negative Binomial" (shortened title), *PCAS* XLVII (1960), pp. 55-65

[7] Hewitt, Charles C., Jr. Review — "A Bayesian View of Credibility," Mayerson, *PCAS* LII (1965), pp. 121-127

[8] Hewitt, Charles C., Jr. Review — "On the Credibility of the Pure Premium," Mayerson, Jones, and Bowers, *PCAS* LVI (1969), pp. 78-82

[9] Mayerson, Allen L. "A Bayesian View of Credibility," *PCAS* LI (1964), pp. 85 et seq.

[10] Mayerson, Jones, and Bowers "On the Credibility of the Pure Premium," *PCAS* LV (1968), p. 179

[11] Verbeek, H. G. "The Claim-Counting Distribution of Large Claim-Realizations," 8th ASTIN Colloquium (1969)