

CREDIBILITY PROCEDURES
LAPLACE'S GENERALIZATION OF BAYES' RULE
AND THE COMBINATION OF
COLLATERAL KNOWLEDGE WITH OBSERVED DATA

BY ARTHUR L. BAILEY

"If thou canst believe, all things are possible
to him that believeth."

Mark 9:23

The casualty insurance business has used credibility formulas or procedures for many years in making rates or in experience rating plans. These formulas have been used to determine the weight to be given to the indications of actual observations in a combination of such indications with *a priori* expectations which were based either on other actual data, on prior knowledge or on reasonable assumptions made before actual observations were available. Such formulas have invariably provided that the weight to be given to actual observations increase as the volume of such observations increases.

Last December the discussion of a paper by Mr. T. O. Carlson, entitled "Statistical and Actuarial Procedures in Liability Insurance"*[1], pointed out that casualty insurance underwriters and actuaries believe that they are not devoid of knowledge before they have acquired any statistics from observed data, and that this belief results in the use of credibility formulas to produce weighted averages of that prior knowledge and the information provided by the observed data. The remarks made at that time were general and unsupported by any demonstration. In fairness to other statisticians and to students of casualty insurance, it appears desirable to present a complete development from basic principles to show exactly the basis upon which credibility formulas rest and to make evident the point at which the classical statistical theory, particularly that of statistical estimation, departs from that used by casualty actuaries.

The basis for these credibility formulas has been a profound mystery to most people who have come in contact with them. The actuary finds them difficult to explain and, in some cases, even difficult to understand. Paradoxical as it may be, the more contact a person has had with statistical practices in other fields or the more training a person has had in the theory of mathematical statistics, the more difficult it has been to understand these credibility procedures or the validity of their application.

The credibility formulas for casualty insurance have been accepted in the past, and continue to be accepted at the present time, because it appears to most people to be logical and reasonable to give the indications of a large volume of data more consideration or weight than the indications of a small volume of data. How much weight the indications of specific volumes of data

* Numerals in brackets refer to the References at the end of this paper.

are to be given, in the casualty business, has continued to be a matter of individual judgment.

In addition to the relatively simple concept that more consideration or weight should be given to a greater volume of observational data, the casualty actuaries have devised credibility procedures to give more weight to the frequent occurrence of small losses than to the occasional or fortuitous occurrence of large losses of the same total amount. (It should be noted that negative losses can not occur.) For example, the rate making procedure for workmen's compensation insurance separates the actual losses into "Serious," "Non-serious" and "Medical" losses and uses three differing schedules of credibility for the three components of the total loss. Several experience rating plans give a greater schedule of credibility to the first G dollars of each loss than is given to the excess of any loss over G dollars. The "Multi-split Experience Rating Plan" for workmen's compensation insurance carries this even further by providing, in effect, a separate schedule of credibilities for each interval of G dollars of which a loss is composed.

It is at this point in the discussion that the ordinary individual has to admit that, while there seems to be some hazy logic behind the actuaries' contentions, it is too obscure for him to understand. The trained statistician cries "Absurd! Directly contrary to any of the accepted theories of statistical estimation." The actuaries themselves have to admit that they have gone beyond anything that has been proven mathematically, that all of the values involved are still selected on the basis of judgment, and that the only demonstration they can make is that, in actual practice, it works. Let us not forget, however, that they have made this demonstration many times. It does work!

It is the purpose of the technical portion of this paper (1) to show that it is proper to give greater weight to larger volumes of observed data and why; (2) to show that under certain conditions, specifically those prevailing in casualty insurance, it is proper to give greater weight to frequently occurring small values than to infrequently occurring large values and why; and (3) to show that these procedures are universally applicable to all fields of observation and are not peculiar to casualty insurance.

HISTORICAL COMMENTS

It will be realized that all of the problems in which credibilities are used are problems in statistical estimation and that the problem of statistical estimation is a very old problem. One of the first steps in the solution of this problem was made by Bayes [2] resulting in what is known as Bayes' Rule. That rule was initially produced as the solution of a specific case in which, *a priori*, all possible events were equally likely to occur [3]. It appears that statisticians of that day grasped at this as being better than no solution even when the basic condition as to equality of *a priori* probabilities was not met. Laplace in an early paper [4] advocated just that, and the practice appears to have become so well established that the Laplace generalization of Bayes' Theorem [5] (published in 1820) was given very little attention. Laplace's generalization actually provided the solution when, *a priori*, the possible events had varying probabilities of occurring.

It is interesting to note here that the Rev. Richard Price, who presented

Bayes' essay for publication in 1763, was closely connected with the insurance industry and would now be called an actuary. The following quotation from his introductory comments to Bayes' essay is so true today that it could not be improved to introduce the subject at hand:

"Every judicious person will be sensible that the problem now mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to assure foundation for all our reasonings concerning past facts, and what is likely to be hereafter. Common sense is indeed sufficient to show us that, from the observation of what has in former instances been the consequence of a certain cause or action, one may make a judgment what is likely to be the consequence of it another time, and that the larger number of experiments we have to support a conclusion, so much the more reason we have to take it for granted. But it is certain that we cannot determine, at least not to any nicety, in what degree repeated experiments confirm a conclusion, without the particular discussion of the beforementioned problem; which therefore, is necessary to be considered by any one who would give a clear account of the strength of analogical or inductive reasoning; concerning which, at present, we seem to know little more than that it does sometimes in fact convince us, and at other times not; and that, as it is the means of acquainting us with many truths, of which otherwise we must have been ignorant; so it is, in all probability, the source of many errors, which perhaps might in some measure be avoided, if the force that this sort of reasoning ought to have with us were more distinctly and clearly understood."

From 1763 to the present time there has been continual argument over the propriety of using Bayes' Theorem in its original form and, possibly because of its apparent complexity, little use made of Laplace's generalization. The advocates of the use of Bayes' original theorem have formalized the process, with its assumption that all possibilities are equally likely, into what they describe as the application of the "Principle of Insufficient Reason." Their opponents have in turn characterized it as the "Assumption of the Equal Distribution of Ignorance," or the "Theory of Equal Ignorance." R. A. Fisher has modified it slightly to produce the "Method of Maximum Likelihood." Others have developed the "Best Unbiased Estimate" by methods which assume that there is only one possibility rather than several or many.

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 [6] 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 [7], 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.

In our own *Proceedings* we have some astounding paradoxes which only serve to show the extent to which the teaching of the Principle of Insufficient Reason has been embedded in the minds of even our own actuaries. In 1918 Mr. Whitney [8] presented the first comprehensive development of credibilities to appear in our *Proceedings*. He assumed that the inherent hazards differed among classifications of risks and assumed a knowledge of the distribution of such hazards. However, in the course of the mathematical development he used Bayes' Rule to obtain a solution, thus reversing his assumption in the middle of the development. Mr. Arne Fisher, in discussing Mr. Whitney's paper [9], took Mr. Whitney to task for using Bayes' Rule, quoted many authorities against the use of it, and then suggested another approach which was based on the same philosophy, if not directly on the same theorem.

From the foregoing it will be appreciated that anyone advocating a return all the way back to the fundamental principles of Laplace's generalization of Bayes' Theorem must look for opposition from many sides. However, Mr. Kendall's recent survey [10] of the current position of probability theory and his plea for progress along practical lines has been accepted by the writer as a definite encouragement to present such a development of the credibilities or weights to be given to observed data in its combination with collateral data or with *a priori* knowledge. Let us be clear in one thing however. Use will be made of Laplace's generalization of Bayes' Rule and not of the original Bayes' Rule.

GENERAL DISCUSSION

Let us define the problem of statistical estimation as that in which it is desired to obtain $E(x | H)$, the expected value of a statistic x which corresponds to the origin or cause of an observed event H . Such an expected value is the sum of the products of all possible values of x and the probabilities $P(x | H)$; where $P(x | H)$ is the probability that the value x was the value corresponding to the origin or cause of the observed event H .

In the insurance business such an expected value is obviously desirable in setting insurance rates, in order that there will be a balance between premiums and losses. The use by an actuary of a "maximum likelihood" estimate would be suicidal because, in many cases, the most likely event is the complete absence of loss. Thus the maximum likelihood estimate would provide nothing for losses and the premium would be, to say the least, inadequate.

The expected value $E(x | H)$ is an unbiased estimate of x for a particular value of H . It should be noted, however, that the "Best Unbiased Estimate" of the literature is unbiased for a particular value of x , not of H , under the tacit assumption that there is only one possible value of x , as yet unknown but having a probability or certainty of existing. This only serves to bring out a major difference of approach. The actuary knows that there is more than one possible value of x and is willing to assume that he can approximate the *a priori* probabilities of the existence of such possible values.

The expected value $E(x | H)$ will be the "best" estimate, from the least squares point of view, because, if it is used as the estimate of x for all of the

possible cases for which H may occur, the sum of the squares of the errors, $(x - E')$, will be a minimum when $E' = E(x | H)$.

It will be noted that $x' = E(x | H)$ is the true regression of x on H and that it may be a series of discrete points or a continuous curve, but not necessarily a straight line. In specific cases the discrete points may fall on a straight line or the continuous curve may actually be a straight line. (Several such special cases will be discussed herein.) Whatever the form of the true regression, it will be possible to obtain the best linear regression of x on H , and it is such best linear regressions that have been previously discussed by the writer [11].

The Laplace generalization of Bayes' Rule [3] states that if an event, H , has been produced by one of the mutually exclusive conditions, F_1, F_2, \dots, F_j , and if $K(x)$ is the *a priori* probability that F_x existed, and if $P(H | x)$ is the *a priori* probability that when F_x exists the event H will occur, the *a posteriori* probability $P(F_a | H)$ that the particular condition F_a was the origin or cause of an observed event H is:

$$P(F_a | H) = K(a) \cdot P(H | a) \left/ \sum_x K(x) \cdot P(H | x) \right. \quad (1)$$

When the mutually exclusive conditions F_1, F_2, \dots, F_j are the conditions under which a statistic has the values 1, 2, \dots, j , the value of $E(x | H)$ can be written as

$$E(x | H) = \sum_x x \cdot P(x | H) = \sum_x x \cdot K(x) \cdot P(H | x) \left/ \sum_x K(x) \cdot P(H | x) \right. \quad (2)$$

It is important to note that, if the event H is the simultaneous occurrence of events H_1, H_2, \dots, H_n , then $E(x | H)$ is *not* the average value of $E(x | H_i)$, but:

$$E(x | H) = \frac{\sum_x x \cdot K(x) \cdot P(H_1 | x) \cdot P(H_2 | x) \cdot \dots \cdot P(H_n | x)}{\sum_x K(x) \cdot P(H_1 | x) \cdot P(H_2 | x) \cdot \dots \cdot P(H_n | x)} \quad (3)$$

In the following developments either formulas (2) or (3) will be utilized as the case requires. For simplicity of expression J_H will be used at times in place of $E(x | H)$, $F(H | x)$ in place of $K(x) \cdot P(H | x)$, and $F(H)$ in place of $\sum_x K(x) \cdot P(H | x)$.

The error variance, $\sigma_{e,H}^2$, of using J_H as an estimate of the true value of x in all of the possible cases for which a particular value of H may occur is:

$$\sigma_{e,H}^2 = \sum_x (x - J_H)^2 \cdot F(H | x) \left/ F(H) \right.,$$

or

$$\sigma_{e,H}^2 = \frac{\sum_x x^2 \cdot F(H | x)}{F(H)} - J_H^2 \quad (4)$$

The error variance, σ_e^2 , of using J_H as an estimate of the true value of x in all of the possible cases for which all possible values of H may occur is:

$$\sigma_e^2 = \sum_H F(H) \cdot \sigma_{e,H}^2 \quad \text{as} \quad \sum_H F(H) = 1. \quad (5)$$

The mean error for each possible value of H is zero.

At this point it will be desirable to let the mean and variance of the $K(x)$ distribution be indicated by m and σ^2 respectively and to let T^2 indicate the variance of H for all values of x . Thus:

$$\begin{aligned} m &= \sum_x x \cdot K(x) \\ \sigma^2 &= \sum_x x^2 \cdot K(x) - m^2 \\ T^2 &= \sum_H H^2 \cdot F(H) - \left[\sum_H H \cdot F(H) \right]^2 \\ \sigma^2 + m^2 &= \sum_H \left[\sum_x x^2 \cdot F(H | x) \right]. \end{aligned}$$

Combining (5) with (4) and using this new notation

$$\sigma_e^2 = \sigma^2 + m^2 - \sum_H J_H^2 \cdot F(H). \quad (6)$$

In the special cases for which $\sum_H H \cdot P(H | x) = Ax + B$ the regression of H on x is the line $H' = Ax + B$. Irrespective of the form of the true regression of x on H , the best fitting straight line can be obtained readily from knowledge as to the relationship of the coefficients of the best linear regression of x on H and the coefficients of the best linear regression of H on x . Thus, the best fitting straight line can be written as

$$x' = Z \left(\frac{H - B}{A} \right) + (1 - Z)m \quad (7)$$

where

$$Z = A^2 \sigma^2 / T^2. \quad (8)$$

If the true regression of x on H is a straight line, it obviously must be the line expressed in (7), and x' then becomes J_H or $E(x | H)$. Such special cases will be considered later, but it should be noted here that in such cases, (6) reduces* to:

$$\sigma_e^2 = \sigma^2(1 - Z). \quad (9)$$

* The algebra of the derivation of (9) from (6) is straightforward but is not shown here because of its length.

It should be noted that equation (7) provides for the combination of the indications of the data, summarized by $(H - B)/A$, and the *a priori* knowledge, summarized by m , through a weighting procedure in which Z is the weight given to the indications of the data and $(1 - Z)$ is the weight given to the *a priori* knowledge.

WHEN $P(H | x)$ FOLLOWS THE BINOMIAL DISTRIBUTION

When the event H is the occurrence of H successes out of n trials for each of which the *a priori* probability of a success was the same and equal to x , the value of $P(H | x)$ follows the Binomial distribution and is $\binom{n}{H} x^H (1 - x)^{n-H}$. The problem is to obtain the expected value or estimate of the true value of x from the observation that H successes occurred out of n trials and the *a priori* knowledge of the probabilities $K(x)$ of various possible values of x ; summarized if possible by the mean, m , and the variance σ^2 of the probability function $K(x)$.

The best straight line regression of x on H may be obtained by the following reasoning.

For a particular value of x , the mean value of H is nx , the variance of H is $nx(1 - x)$ and the mean square of H is $n(n - 1)x^2 + nx$. For all values of x the mean square of H is $n(n - 1)(\sigma^2 + m^2) + nm$; the mean value of H is nm ; and T^2 , the variance of H , is $n(n - 1)\sigma^2 + nm(1 - m)$. The value of $\sum_H H \cdot P(H | x)$ is nx , so that the values of A and B to be used in (7) and (8) are $A = n$ and $B = 0$. Thus Z can be obtained from (8) as:

$$Z = \frac{n \sigma^2}{(n - 1) \sigma^2 + m(1 - m)} \tag{10}$$

This value of Z can be inserted in equation (9) to obtain the best fitting straight line to the regression of x on H .

In general $E(x | H)$ will consist of $n + 1$ discrete points which can be calculated from (2) for any known values of $K(x)$. There is one special case for which these $n + 1$ points will all fall on a single straight line. This case occurs when $K(x)$, the *a priori* probabilities of the existence of x , follow the Hardy [12] distribution* as suggested by E. C. Molina in 1946 [6].

Let

$$K(x) = Kx^{a-1}(1 - x)^{b-1} \tag{11}$$

where

$$c = \frac{m(1 - m)}{\sigma^2} - 1$$

and

$$a = mc, \quad b = (1 - m)c$$

and

$$K = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}$$

so that x has a mean of m and a variance of σ^2 .

* Note that this is the particular case of the Pearsonian Type I distribution for which the range of x is from 0 to 1 and is also known as the Beta distribution [13].

Inserting these values of $K(x)$ and the Binomial distribution values of $P(H | x)$ in (2), except for constants common to both numerator and denominator, gives:

$$E(x | H) = \int_0^1 x^{H+a}(1-x)^{n-H+b-1} dx \Bigg/ \int_0^1 x^{H+a-1}(1-x)^{n-H+b-1} dx. \quad (12)$$

$$= \frac{B(H+a+1, n-H+b)}{B(H+a, n-H+b)} = \frac{H+a}{n+a+b}$$

where $B(x, y)$ is the Beta function equal to $\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. When the values of a and b from (11) and the value of Z from (10) are used, (12) becomes:

$$E(x | H) = Z \frac{H}{n} + (1-Z)m. \quad (13)^*$$

The value Z is thus seen to be the credibility, or percentage of total weight, to be given to the observed ratio of successes to trials in its combination with the *a priori* expectation, m . From (10) it is seen that when n is one, $Z = \sigma^2/m(1-m)$ and that Z increases as n increases, approaching unity as n approaches infinity.

WHEN $P(H | x)$ FOLLOWS THE POISSON DISTRIBUTION

When H is the number of events observed in n units of time or space throughout which events are randomly distributed with an average frequency of x events per unit, the value of $P(H | x)$ follows the Poisson distribution and is $(nx)^H e^{-nx}/H!$. The problem is to estimate x by obtaining its expected value from the observed value of H and the *a priori* knowledge of the probabilities, $K(x)$, of various possible values of x ; summarized if possible by the mean m and the variance σ^2 of the probability function $K(x)$.

The best straight line regression of x on H may be obtained by the following reasoning. For a particular value of x , the mean value of H is nx , the variance of H is nx and the mean square of H is $nx + n^2x^2$. For all values of x the mean square of H is $nm + n^2(\sigma^2 + m^2)$; the mean value of H is nm ; and T^2 , the variance of H , is $nm + n^2\sigma^2$. The value of $\sum_H H \cdot P(H | x)$ is nx , so that the values of A and B to be used in (7) and (8) are $A = n$ and $B = 0$. Thus Z can be obtained from (8) as:

$$Z = \frac{n\sigma^2}{n\sigma^2 + m} \quad (14)$$

which can be inserted in equation (9) to obtain the best fitting straight line to the regression of x on H .

* Note: If the Principle of Insufficient Reason is applied in this case the assumption would be that all values of x from 0 to 1 were equally likely. This would produce $E(x | H) = (H+1)/(n+2)$ and not $E(x | H) = H/n$ as is frequently used.

In general $E(x | H)$ will consist of discrete points corresponding to the discrete values of H from 0 to ∞ . These discrete points can be calculated from (2) for any known values of $K(x)$. There is one special case for which these points will all fall on a single straight line. This case occurs when $K(x)$ follows the Pearsonian Type III distribution having a range of x from 0 to ∞ ;* specifically when:

$$K(x) = \frac{g^{mg} x^{mg-1} e^{-gx}}{\Gamma(mg)} \quad \text{where } g = \frac{m}{\sigma^2} \tag{15}$$

Inserting this value of $K(x)$ and the Poisson distribution values of $P(H | x)$ in (2), except for constants common to both numerator and denominator, gives:

$$E(x | H) = \frac{\int_0^\infty x^{H+mg} e^{-x(n+g)} dx}{\int_0^\infty x^{H+mg-1} e^{-x(n+g)} dx} \tag{16}$$

$$= \frac{\Gamma\left(H + \frac{m^2}{\sigma^2} + 1\right)}{\left(n + \frac{m}{\sigma^2}\right)^{H+(m^2/\sigma^2)+1}} \cdot \frac{\left(n + \frac{m}{\sigma^2}\right)^{H+(m^2/\sigma^2)}}{\Gamma\left(H + \frac{m^2}{\sigma^2}\right)} = \frac{H + \frac{m^2}{\sigma^2}}{n + \frac{m}{\sigma^2}}$$

Using the value of Z in (14), this becomes:

$$E(x | H) = Z \cdot \frac{H}{n} + (1 - Z)m. \tag{17} \dagger$$

The value Z is seen to be the credibility, or percentage of total weight, to be given to the observed number of events per unit of time or space in its combination with the *a priori* expectation, m . From (14) it is seen that when n is one, $Z = \sigma^2 / (\sigma^2 + m)$ and that Z increases as n increases, approaching unity as n approaches infinity.

WHEN H IS THE SUM OF THE INDEPENDENT VARIABLES x AND h

When H consists of the simultaneous occurrence of the values H_1, H_2, \dots, H_n for the sum of a single value of x and n random values of a variable h , and

* It will be noted that this distribution is closely related to the Chi-square distribution with $2m^2/\sigma^2$ degrees of freedom (see reference [13] and is called the Gamma distribution. It was used by R. Keffer [14] in 1929 with $m = 1$.

† Note: If the Principle of Insufficient Reason is applied in this case it would produce $E(x | H) = (H + 1)/n$ and not H/n as is frequently used.

when h is independent of x with a mean of B , a variance of S^2 and a frequency distribution of $\varphi(h)$, the value of $P(H | x)$ may be expressed as:

$$P(H | x) = \varphi(H_1 - x) \cdot \varphi(H_2 - x) \cdot \cdots \cdot \varphi(H_n - x). \quad (18)$$

The problem is to estimate the value of x included in each of the sums, H_1, H_2, \dots, H_n by obtaining its expected value from the values H_1, H_2, \dots, H_n and the *a priori* knowledge of the probabilities, $K(x)$, of various possible values of x ; summarized if possible by the mean m and variance σ^2 of the probability function $K(x)$ and the mean, \bar{H} , of the values H_1, H_2, \dots, H_n .

Consider the special case when both $K(x)$ and $\varphi(h)$ are normal distributions. Inserting the values of $K(x)$ and of $\varphi(H_1 - x)$ in (3), except for constants common to both numerator and denominator, gives:

$$E(x | H) = \frac{\int_{-\infty}^{\infty} x e^{-\left[\frac{(x-m)^2}{2\sigma^2} + \frac{(H_1-x-B)^2 + (H_2-x-B)^2 + \cdots + (H_n-x-B)^2}{2S^2} \right]} dx}{\int_{-\infty}^{\infty} e^{-\left[\frac{(x-m)^2}{2\sigma^2} + \frac{(H_1-x-B)^2 + (H_2-x-B)^2 + \cdots + (H_n-x-B)^2}{2S^2} \right]} dx} \quad (19)$$

The numerator of (19) is of the form $C \int_{-\infty}^{\infty} U \cdot dV$ where

$$C = e^{-\frac{1}{2} \left[\frac{m^2}{\sigma^2} + \frac{nB^2}{S^2} + \frac{H_1^2 + H_2^2 + \cdots + H_n^2}{S^2} + \frac{2B(H_1 + H_2 + \cdots + H_n)}{S^2} \right]},$$

$$U = e^{x \left[\frac{m}{\sigma^2} + \frac{H_1 + H_2 + \cdots + H_n}{S^2} - \frac{nB}{S^2} \right]}, \text{ and}$$

$$dV = x e^{-\frac{x^2}{2} \left[\frac{1}{\sigma^2} + \frac{n}{S^2} \right]} dx.$$

Thus the numerator of (19) may be expressed as $C \cdot U \cdot V \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} C \cdot V \cdot dU$.

The value of $C \cdot U \cdot V \Big|_{-\infty}^{\infty}$ is nil as $C \cdot U \cdot V$ is zero at both ∞ and $-\infty$.

The value of $-\int_{-\infty}^{\infty} C \cdot V \cdot dU$ is the denominator of (19) multiplied by the following quantity, which thus becomes the value of $E(x | H)$:

$$E(x | H) = \frac{\frac{m}{\sigma^2} + \frac{H_1 + H_2 + \dots + H_n}{S^2} - \frac{nB}{S^2}}{\frac{1}{\sigma^2} + \frac{n}{S^2}}. \quad (20)$$

This may be expressed as:

$$E(x | H) = Z(\bar{H} - B) + (1 - Z)m \quad (21)^*$$

where

$$Z = \frac{n \sigma^2}{n \sigma^2 + S^2}. \quad (22)^*$$

Not only is this special case one for which the true regression of x on H is a straight line when H is a single observed sum of x and h ; but it is also one for which all of the knowledge pertinent to the determination of $E(x | H)$ is contained in \bar{H} when H is the simultaneous occurrence of n values with such an average. Knowledge of the individual values of H_1, H_2, \dots, H_n would add nothing to the knowledge provided by \bar{H} .

This concentration of knowledge in \bar{H} is the result of assuming that $\varphi(h)$ is a normal distribution function. If that assumption is continued without the assumption that $K(x)$ is a normal distribution function, the regression of x on \bar{H} will not be a straight line although the best fitting straight line will be the line provided by (21) and (22). If $\varphi(h)$ is not a normal distribution function, $E(x | H)$ can be calculated from (3) and will involve the individual values of H_i and not only \bar{H} .

$E(x | H)$ IN TERMS OF \bar{J} AND \bar{H}

Before proceeding to other cases, it will be helpful to investigate the possibility of expressing $E(x | H)$ in terms of \bar{J} where \bar{H} is the concurrent observation of events H_1, H_2, \dots, H_n and \bar{J} is the average value of $E(x | H_1), E(x | H_2), \dots, E(x | H_n)$.

In the case where $P(H | x)$ follows the Binomial distribution and $K(x)$ follows the Beta distribution, a value $E(x | H_i) = J_i$ could be obtained for the result of each of the n trials. It would be:

$$J_i = \frac{\sigma^2}{m(1-m)} \cdot H_i + \left(1 - \frac{\sigma^2}{m(1-m)}\right) \cdot m. \quad (23)$$

Thus

$$\bar{J} = \frac{\sigma^2}{m(1-m)} \cdot \frac{H}{n} + \left(1 - \frac{\sigma^2}{m(1-m)}\right) \cdot m \quad (24)$$

OR

* Note that the Principle of Insufficient Reason would, in effect, assume that $\sigma^2 = \infty$ so that Z would become 1 and $E(x | H)$ would equal $\bar{H} - B$. Note also that, under that assumption, (8) would produce a larger value, namely $\sigma_0^2 = S^2/n$ instead of $\sigma_0^2 = S^2/(n + \frac{S^2}{\sigma^2})$.

$$m = \frac{m(1 - m)\bar{J} - \sigma^2 \cdot \frac{H}{n}}{m(1 - m) - \sigma^2}. \quad (25)$$

Substituting this value of m and the value of Z given in (10) in equation (13) produces:

$$E(x | H) = \frac{n - 1}{n} \cdot Z \cdot \frac{H}{n} + \left(1 - \frac{n - 1}{n} \cdot Z\right) \bar{J}. \quad (26)$$

In the case where $P(H | x)$ follows the Poisson distribution and $K(x)$ follows the Gamma distribution, a value of J_i could be obtained from the number of events observed in each of the n units of time or space. It would be

$$J_i = \frac{\sigma^2}{\sigma^2 + m} \cdot H_i + \left(1 - \frac{\sigma^2}{\sigma^2 + m}\right) \cdot m. \quad (27)$$

Thus

$$\bar{J} = \frac{\sigma^2}{\sigma^2 + m} \cdot \frac{H}{n} + \left(1 - \frac{\sigma^2}{\sigma^2 + m}\right) \cdot m \quad (28)$$

or

$$m = \frac{(\sigma^2 + m)\bar{J} - \sigma^2 \cdot H/n}{m}. \quad (29)$$

Substituting this value of m and the value of Z given in (14) in equation (17) produces:

$$E(x | H) = \frac{n - 1}{n} \cdot Z \cdot \frac{H}{n} + \left(1 - \frac{n - 1}{n} \cdot Z\right) \cdot \bar{J}. \quad (30)$$

In the case where $P(H | x) = P(H_1 | x) \cdot P(H_2 | x) \cdot \dots \cdot P(H_n | x)$ and where $P(H_i | x)$ and $K(x)$ are both normal distributions, a value of J_i could be obtained for each of the n values of H_i . It would be:

$$J_i = \frac{\sigma^2}{\sigma^2 + S^2} (H_i - B) + \left(1 - \frac{\sigma^2}{\sigma^2 + S^2}\right) \cdot m. \quad (31)$$

Thus

$$\bar{J} = \frac{\sigma^2}{\sigma^2 + S^2} (\bar{H} - B) + \left(1 - \frac{\sigma^2}{\sigma^2 + S^2}\right) \cdot m \quad (32)$$

or

$$m = \frac{(\sigma^2 + S^2)\bar{J} - \sigma^2(\bar{H} - B)}{S^2}. \quad (33)$$

Substituting this value of m and the value of Z given in (22) in equation (21) produces:

$$E(x | H) = \frac{n - 1}{n} \cdot Z \cdot (\bar{H} - B) + \left(1 - \frac{n - 1}{n} \cdot Z\right) \cdot \bar{J}. \quad (34)$$

It is noted that the coefficient $(n - 1)Z/n$ is common to equations (26), (30) and (34) and that, although Z is different in each case, Z is the coefficient in the best straight line regression of x on \bar{H} . When J_i is not a linear function of H_i , the coefficient $(n - 1)Z/n$ is obviously proper both when n is one and when n approaches infinity. This suggests that this general relationship is either always true or that it represents a close approximation to the truth even when J_i is not a linear function of H_i . This will be assumed to be the case although it will be clearly understood that it has not been proven.

WHEN H IS THE PRODUCT OF THE INDEPENDENT VARIABLES x AND h

When H consists of the simultaneous occurrence of the values H_1, H_2, \dots, H_n as the product of a single value of x and n random values of a variable h , and when h is independent of x with a mean of 1, a variance of S^2 and a frequency distribution of $\varphi(h)$, the value of $P(H | x)$ may be expressed as:

$$P(H | x) = \frac{1}{x^n} \varphi\left(\frac{H_1}{x}\right) \cdot \varphi\left(\frac{H_2}{x}\right) \cdot \dots \cdot \varphi\left(\frac{H_n}{x}\right). \tag{35}$$

This condition is of frequent occurrence in practical applications for which both x and h can have only positive values. The problem is to obtain the expected value of the x which is included in each of the products, H_1, H_2, \dots, H_n , from those values and the *a priori* knowledge of the probabilities, $K(x)$, of various possible values of x ; summarized if possible by the parameters of $K(x)$ and $\varphi(h)$ and the means \bar{H} and \bar{J} for the values H_1, H_2, \dots, H_n .

The best fitting straight line to the regression of x on \bar{H} can be shown to be:

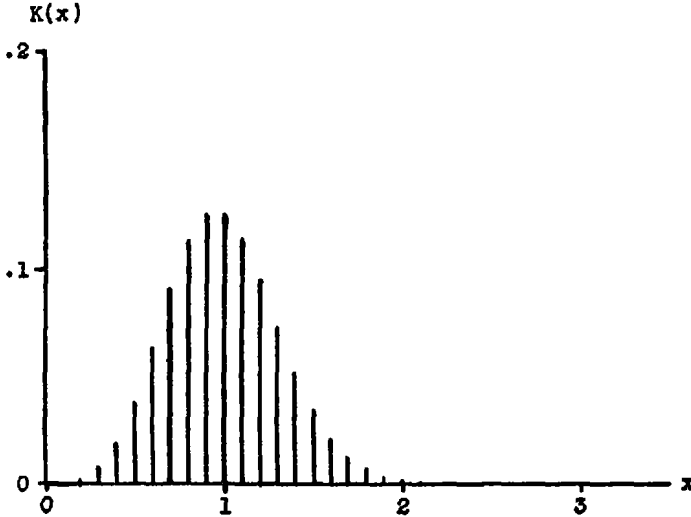
$$x' = Z \cdot \bar{H} + (1 - Z)m \tag{36}$$

where

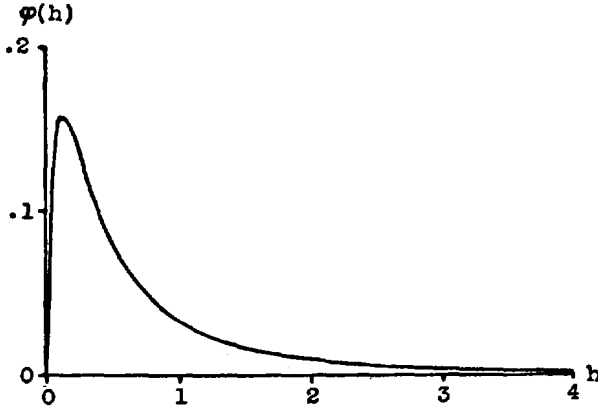
$$Z = \frac{n \cdot \sigma^2}{n \cdot \sigma^2 + S^2(\sigma^2 + m^2)}. \tag{37}$$

This straight line can not, however, be depended upon to give a reliable estimate of x for small values of n for two reasons: first, the true regression must be expected to be far from a straight line in most practical applications; and second, there is usually much more information in the individual values of H_i than is summarized in the average \bar{H} .

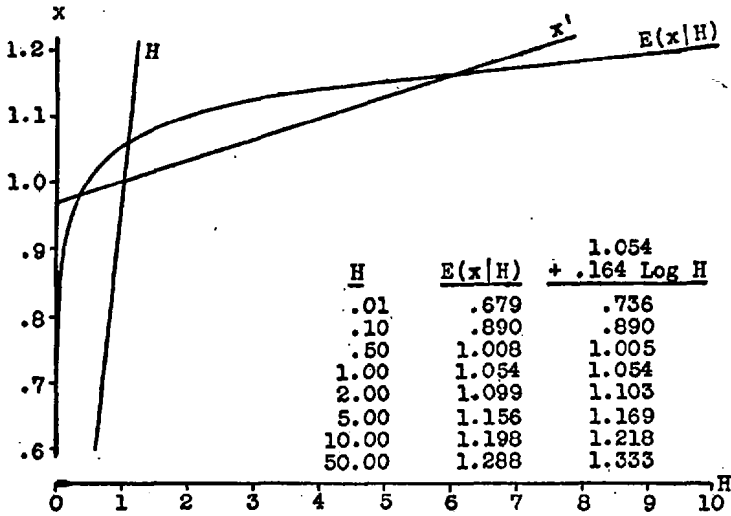
To show the departure from a straight line regression when $n = 1$, an example has been selected in which $\varphi(h)$ is typical of the distribution of losses by size of loss for casualty insurance and in which $K(x)$ is typical of the distribution of classification average claim costs expressed as a percentage of the average for all classes. To simplify calculations $K(x)$ has been taken so that it has values only at the discrete intervals of $x = n/10$ where n is an integer. With $K(x) = e^{-10} \cdot 10^{10x} / (10x)! K(x)$ has a mean of 1, a variance of $\sigma^2 = 1/10$ and the distribution shown in the following diagram. $P(H | x)$ has been chosen as equal to $\varphi(\frac{H}{x})/x$ where $\varphi(h)$ has a mean of 1 and a variance of $S^2 = 3$ and



with $\varphi(h)$ following the normal logarithmic distribution shown in the following diagram.



$E(x | H)$ has been calculated for a sufficient number of values of H to indicate the relationship to H shown in the following diagram. The lines $x = H$, produced by the application of the Principle of Insufficient Reason, and $x' = ZH + (1 - Z)m$, produced as the best straight line regression, are also shown on the diagram for comparison with the curve $x = E(x | H)$. The calculated values of $E(x | H)$ are shown together with the values of $1.054 + .164 \log_{10} H$ which appear to reasonably approximate the values of $E(x | H)$ in this example.



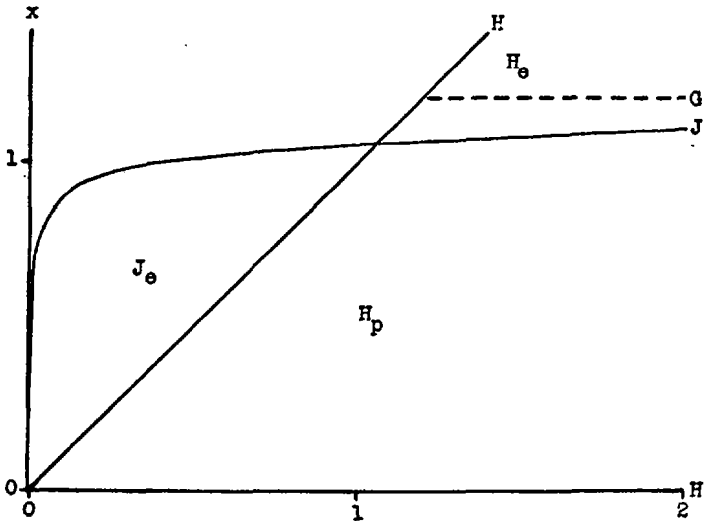
PRIMARY AND EXCESS VALUES

When H_i is the product of the independent variables x and h_i , the value of $E(x | H)$ may be expressed in terms of \bar{H} and \bar{J} as:

$$E(x | H) = \frac{n-1}{n} Z\bar{H} + (1 - \frac{n-1}{n} Z)\bar{J} \tag{38}$$

where Z has the value shown in (37). This relationship is exact when the true regression of x on H is a straight line. Let it be assumed to hold when that regression is not a straight line.

Consider now the portions of H_i and J_i illustrated by areas on the following diagram and as defined below:



H_p = the primary portion of H , defined as:
 $H_p = H$ if $H \leq J$ and $H_p = J$ if $H > J$;

H_e = the excess portion of H , defined as:
 $H_e = 0$ if $H \leq J$ and $H_e = H - J$ if $H > J$;

J_e = the excess portion of J , defined as:
 $J_e = J - H$ if $H \leq J$ and $J_e = 0$ if $H > J$.

Noting that $\bar{H} = \bar{H}_p + \bar{H}_e$ and that $\bar{J} = \bar{H}_p + \bar{J}_e$, equation (38) can be written:

$$E(x | H) = \bar{H}_p + \frac{n-1}{n} Z \bar{H}_e + (1 - \frac{n-1}{n} Z) \bar{J}_e. \quad (39)$$

It is found that the average of the primary portions of the observations H_1, H_2, \dots, H_n should be given full credibility (a weight of unity) and that the excess portions of those observations should be given a lesser weight of $(n-1)Z/n$. This coincides with the beliefs of casualty actuaries as expressed in practice in the Multi-split Experience Rating Plan for workmen's compensation insurance. As the *a priori* expected value of \bar{J}_e is equal to that of \bar{H}_e , the actuaries have replaced \bar{J}_e in (39) with the *a priori* expected value of \bar{H}_e . It is obvious from (39) that such a replacement impairs the accuracy of the estimate of x although such impairment may not be appreciable.

From the diagram it will be seen why the single split of observed values at G may be a sufficiently close approximation. With such a split the definition of primary and excess values would be $H_p = H$ if $H \leq G$ and $H_p = G$ if $H > G$, and $H_e = 0$ if $H \leq G$ and $H_e = H - G$ if $H > G$.

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. It is for that person's encouragement that Jesus' statement was initially quoted.

REFERENCES

- [1] T. O. Carlson: "Statistical and Actuarial Procedures in Liability Insurance." Proceedings of American Association of University Teachers of Insurance (1950).
- [2] Thomas Bayes: "An Essay towards Solving a Problem in the Doctrine of Chances." Philosophical Transactions, Vol. LIII (1763) p. 370. See discussion by E. C. Molina in "Bayes' Theorem, An Expository Presentation," Annals of Mathematical Statistics.
- [3] Hugh H. Wolfenden: "The Fundamental Principles of Mathematical Statistics" (Macmillan Company of Canada Ltd., Toronto, 1942).
- [4] Laplace: "Memoire sur les Probabilités" (1774). See discussion by E. C. Molina in "Theory of Probability, Some Comments on Laplace's Théorie Analytique," Bulletin of the American Mathematical Society, Vol. 36 (June 1930) pp. 369-392.
- [5] Laplace: "Théorie Analytique des Probabilités" (Third Edition of 1820) p. 182.
- [6] E. C. Molina: "Some Fundamental Curves for the Solution of Sampling Problems." Annals of Mathematical Statistics, Vol. 17, No. 3 (September 1946).
- [7] Bertrand Russell: "Human Knowledge, Its Scope and Limits" (Simon and Schuster, New York, 1948) Part 5, Chapter 6, p. 380, "Degrees of Credibility."
- [8] A. W. Whitney: "The Theory of Experience Rating." Proceedings of the Casualty Actuarial Society, Vol. IV (1918) p. 274.
- [9] Arne Fisher: Written Discussion of [8]. Proceedings of the Casualty Actuarial Society, Vol. V (1919) p. 139.
- [10] M. C. Kendall: "On the Reconciliation of the Theories of Probabilities." Biometrika, Vol. 36, Parts I and II (June 1949).
- [11] A. L. Bailey: "A Generalized Theory of Credibility." Proceedings of the Casualty Actuarial Society, Vol. 32 (1945) p. 13.
- [12] G. F. Hardy: Transactions, Faculty of Actuaries, Vol. 8 (1920) p. 181.
- [13] H. Cramér: "Mathematical Methods of Statistics" (Princeton University Press, Princeton, N. J., 1946) pp. 233-249.
- [14] R. Keffer: "An Experience Rating Formula." Transactions of the Actuarial Society of America, Vol. 30 (1929) p. 130 (also pp. 593-611).