CREDIBILITY: THE BAYESIAN MODEL VERSUS BÜHLMANN'S MODEL

THOMAS N. HERZOG

ABSTRACT

"Credibility" is becoming as much of a catchword for life and health insurance actuaries as it is for casualty actuaries. Life and health insurance actuaries are better acquainted with Bayesian credibility models for estimating pure premiums, while casualty actuaries are better acquainted with the models of Hans Bühlmann. This paper examines the compatibility of the two models, in the hope that this exploration will prompt a greater cross-fertilization of ideas between the actuarial branches.

---

1. INTRODUCTION

According to Longley-Cook [19, p. 3]: "The word *credibility* was originally introduced into actuarial science as a measure of the credence that the actuary believes should be attached to a particular body of experience for ratemaking purposes." Thus, we might write:

$$C = ZR + (1 - Z)H \qquad (1.1)$$

where
- R is the mean of the current observations (for example, the data)
- H is the prior mean (for example, the estimate based on the actuary's prior data and/or opinion)
- C is the compromise estimate to be calculated
- Z is the credibility factor, satisfying $0 \le Z \le 1$.

In this case, an application of credibility theory produces a linear estimate of the true value, derived as the result of a compromise between the current observations and the actuary's prior opinion. The symbol Z denotes the weight assigned to the (current) data and $(1 - Z)$ the weight assigned to the prior data. In insurance terms, the new insurance rate, C, is a weighted average of the old insurance rate, H, and the observations, R, for the most recent period of observation. An alternative interpretation of Equation (1.1) is to let C be the insurance rate of a particular class of business, let R be the recent experience for that class, and let H be the insurance rate for all classes combined.

This paper consists of two principal parts. First, Sections 2 and 3 present a detailed discussion of some fundamental concepts of Bayesian analysis, which forms the basis for an important approach to insurance ratemaking. Second, Section 4 describes a credibility model suggested by Bühlmann [2].

Fuhrer [8] recently applied some Bühlmann-type models to health insurance problems. Klugman [18] uses a full Bayesian approach to analyze actual data on worker's compensation insurance. He investigates two problems. First, he calculates the joint posterior distribution of the relative frequency of claims in each of 133 rating groups. He employs three distinct prior distributions and shows that the results are virtually identical in all three instances. Second, Klugman analyzes the loss ratio for three years of experience in 319 rating classes in Michigan. He uses these data to construct prediction intervals for future observations, that is, the fourth year. He then compares his predictions to the actual results.

2. BAYES' THEOREM AND HEWITT'S EXAMPLES

This section consists of Bayes' Theorem (the foundation of a branch of statistics called Bayesian statistics,* which is useful for solving a wide range of actuarial problems), as well as two examples from the important paper of Hewitt [12].

Bayes' Theorem:

Let A and B be events such that $P[B] > 0$. Then:

$$P[A|B] = \frac{P[B|A]P[A]}{P[B]}.$$

Example 2.1:

A die is selected at random (that is, with probability 1/2) from a pair of "honest" dice. One die has one marked face and five unmarked faces, and the other die has three marked faces and three unmarked faces. We define:

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

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

---

*Edwards, Lindman, and Savage [6] summarize the Bayesian view of statistics as follows: "Probability is orderly opinion, and inference from data is nothing other than the revision of such opinion in the light of relevant new information."

U_i as the result of having an unmarked side showing on the i-th toss for $i = 1, 2, \ldots$

M_i as the result of having a marked side showing on the i-th toss for $i = 1, 2, \ldots$.

By definition of ''honest'' die, we simply mean that

$$P[U_i|A_1] = 5/6 \text{ and } P[U_i|A_2] = 3/6.$$

Example 2.2:

A spinner is selected at random (that is, with probability 1/2) from a pair of spinners. It is known that: (1) one spinner has six equally likely sectors, five of which are marked ''two'' and one of which is marked ''fourteen,'' and (2) the other spinner has six equally likely sectors, three of which are marked ''two'' and three of which are marked ''fourteen.'' We define:

B_1 as the state of having drawn the spinner with five ''twos'' and one ''fourteen''

B_2 as the state of having drawn the spinner with three ''twos'' and three ''fourteens''

S_i as the random variable representing the result of the i-th spin, $i = 1, 2, \ldots$.

3. DISCRETE FREQUENCY-SEVERITY INSURANCE MODEL UNDER INDEPENDENCE

This section presents a simple two-stage model of an insurance operation. The model is based on Examples 2.1 and 2.2. Specifically, we assume that there is a single insured whose claim experience is modeled as follows: First, a die and spinner are selected independently and at random. So, using the notation of Section 2, for $i = 1,2$ and $j = 1,2$,

$$P[A_i \text{ and } B_j] = P[A_i] P[B_j] = (1/2)(1/2) = 1/4. \tag{3.1}$$

(Once selected, the die and spinner, which determine the risk characteristics of the insured, are not replaced.) Second, the random claims process starts when the die selected is rolled. If a marked face appears, this constitutes a claim; if not, there is no claim. Third, if there is a claim, the selected spinner will be spun to determine the amount of the claim. Each roll of the die and spin of the spinner, if necessary, constitute a single period of observation. We use X_i to denote the random variable representing the (aggregate) amount of claims during the i-th period of observation, $i = 1, 2, \ldots$.

In this section, we first compute the initial *pure* (or net) *premium*, $E[X_1]$, using the initial (that is, prior) probabilities of $P[A_i$ and $B_j]$. Then, having observed the result of trial 1, we compute a revised pure premium estimate, $E[X_2|X_1]$, based on the revised (that is, posterior) probabilities of $P[A_i$ and $B_j]$ given the result of the first period of observation.

3.1 The Initial Pure Premium

Because X_1 takes only the values 0, 2, and 14 with positive probability, we may write $E[X_1]$ as:

$$E[X_1] = 0P[X_1 = 0] + 2P[X_1 = 2] + 14P[X_1 = 14]. \quad (3.2)$$

Now,

$$P[X_1 = 0] = P[U_1] = P[U_1|A_1] P[A_1] + P[U_1|A_2] P[A_2] = 2/3. \quad (3.3)$$

Also, using M_1 as defined in Example 2.1, we obtain:

$$P[X_1 = 2] = P[M_1 \text{ and } (S_1 = 2)]$$

$$= P[M_1] P[S_1 = 2]$$

$$= (P[M_1|A_1] P[A_1] + P[M_1|A_2] P[A_2]) \times$$

$$(P[S_1 = 2|B_1] P[B_1] + P[S_1 = 2|B_2] P[B_2]) = 2/9. \quad (3.4)$$

Finally, it is left as an exercise for the reader to verify that

$$P[X_1 = 14] = 1/9. \quad (3.5)$$

We summarize the results just obtained in column 2 of Table 1.

TABLE 1

INITIAL PURE PREMIUM ESTIMATE

(1) Value of x	(2) Initial Probability $P[X_1 = x]$	(3) $xP[X_1 = x]$ (1) × (2)
0	2/3	0
2	2/9	4/9
14	1/9	14/9
Total	1	2

Thus, the initial estimate of the pure premium as per Equation (3.2) is equal to 2, the result of adding the entries in column 3 of Table 1.

We may also write $E[X_1]$ as:

$$E[X_1] = \sum_{i=1}^{2} \sum_{j=1}^{2} E[X_1|A_i \text{ and } B_j] \, P[A_i \text{ and } B_j]. \tag{3.6}$$

Given the compound state A_i and B_j, the mean claim amount is equal to the product of (1) the mean number of claims and (2) the mean severity amount, given that a claim occurs. In symbols:

$$E[X_1|A_i \text{ and } B_j] = E[I_1|A_i] \, E[S_1|B_j],$$

where $I_1 = \begin{cases} 1 \text{ if the first toss of the die produces a marked side} \\ 0 \text{ if otherwise.} \end{cases}$

Hence, $E[I_1|A_i] = P[M_1|A_i]$. From Examples 2.1 and 2.2, we obtain the results shown in Table 2. Substituting the values of column (4) as well as $P[A_i \text{ and } B_j] = 1/4$ into Equation (3.6) above again results in $E[X_1] = 2$.

TABLE 2

PURE PREMIUMS BY STATE
(RISK CHARACTERISTICS)

(1) State A_i and B_j	(2) Frequency $E[I_1|A_i]$	(3) Severity $E[S_1|B_j]$	(4) Pure Premium $E[X_1|A_i \text{ and } B_j]$ (2) × (3)
A_1 and B_1	1/6	4	2/3
A_1 and B_2	1/6	8	4/3
A_2 and B_1	1/2	4	2
A_2 and B_2	1/2	8	4

3.2   Revised Pure Premium Estimates

In this section we estimate the pure premium for the second period of observation given the result of the first period. In symbols, we seek to calculate:

$$E[X_2|X_1 = k] \qquad \text{for } k = 0, 2, \text{ and } 14.$$

Because:

(1) Once a die has been chosen, the result of each toss of the die is independent of the results of all other tosses, and

(2) Once a spinner has been selected, the result of each spin of the spinner is independent of the results of all other spins,

the random variable X_2 only depends upon the result of the first trial through the probabilities of the states A_i and B_j; in other words, the only effect is through the revised state probabilities:

$$P[A_i \text{ and } B_j | X_1 = k].$$

As in the previous section, we present two ways of calculating the desired expected value. The first also yields the posterior probabilities of X_2; the second yields some intermediate results to be used in Section 4.

3.3 Pure Premiums and Predictive Distributions

By analogy with Equation (3.2), we may write for $k = 0$, 2, and 14:

$$E[X_2 | X_1 = k] = 0P[X_2 = 0 | X_1 = k] + 2P[X_2 = 2 | X_1 = k] \\ + 14P[X_2 = 14 | X_1 = k] \quad (3.7)$$

Our goal is to calculate the posterior expected claim amount (pure premium) after having observed a claim amount of k during the first period. Now, for the reasons given in the preceding section, we may write for $m = 0$, 2, and 14:

$$P[X_2 = m | X_1 = k] = \\ \sum_{i=1}^{2} \sum_{j=1}^{2} P[X_2 = m | A_i \text{ and } B_j] \, P[A_i \text{ and } B_j | X_1 = k] \quad (3.8)$$

We first calculate the probabilities $P[X_2 = m | A_i \text{ and } B_j]$ starting with $m = 0$. Now,

$$P[X_2 = 0 | A_i \text{ and } B_j] = P[X_2 = 0 | A_i] = P[U_2 | A_i]$$

$$= \begin{array}{ll} 30/36 & i = 1 \\ 18/36 & i = 2 \end{array} \quad (3.9)$$

Next, we consider $P[X_2 = 2 | A_i \text{ and } B_j]$. Now

$$P[X_2 = 2 | A_i \text{ and } B_j] = P[M_2 | A_i] \, P[S_2 = 2 | B_j]. \quad (3.10)$$

Hence the values of $P[X_2 = 2|A_i \text{ and } B_j]$ may be tabulated as shown in Table 3, making use of Equation (3.10).

TABLE 3

$P[X_2 = 2|A_i \text{ AND } B_j]$

State	(1) $P[M_2|A_i]$	(2) $P[S_2 = 2|B_j]$	(3) $P[X_2 = 2|A_i \text{ and } B_j]$ (1) × (2)
A_1 and B_1	1/6	5/6	5/36
A_1 and B_2	1/6	1/2	3/36
A_2 and B_1	1/2	5/6	15/36
A_2 and B_2	1/2	1/2	9/36

The results of Equations (3.9) and (3.10) can be summarized as shown in Table 4.

TABLE 4

PROBABILITIES OF CLAIM OUTCOME GIVEN
STATE (RISK CHARACTERISTICS)—
$P[X_2 = m|A_i \text{ and } B_j]$—IN 36THS

State	Outcome for One Trial			Total
	0	2	14	
A_1 and B_1	30	5	1	36
A_1 and B_2	30	3	3	36
A_2 and B_1	18	15	3	36
A_2 and B_2	18	9	9	36

The results of column (3) of Table 3 constitute column 2 of Table 4.

In order to evaluate each of the terms on the right-hand side of Equation (3.8), it remains to compute

$$P[A_i \text{ and } B_j|X_1 = k] \text{ for } k = 0, 2, 14. \qquad (3.11)$$

By Bayes' Theorem,

$$P[A_i \text{ and } B_j|X_1 = k] = \frac{P[X_1 = k|A_i \text{ and } B_j] \, P[A_i \text{ and } B_j]}{P[X_1 = k]} \qquad (3.12)$$

Since (1) $P[A_i \text{ and } B_j] = 1/4$ for all i and j, (2) the values of $P[X_1 = k]$ constitute column (2) of Table 1, and (3) the values of $P[X_1 = k | A_i \text{ and } B_j]$ constitute Table 4, it is an easy matter to evaluate $P[A_i \text{ and } B_j | X_1 = k]$. The results are summarized in Table 5.

TABLE 5

POSTERIOR DISTRIBUTION OF RISK CHARACTERISTICS —
VALUE OF $P[A_i \text{ AND } B_j | X_1 = k]$

	Value of k		
State	0	2	14
A_1 and B_1	5/16	5/32	1/16
A_1 and B_2	5/16	3/32	3/16
A_2 and B_1	3/16	15/32	3/16
A_2 and B_2	3/16	9/32	9/16
Total	1	1	1

Thus, as an example,

$$P[A_1 \text{ and } B_2 | X_1 = 14] = \frac{P[X_1 = 14 | A_1 \text{ and } B_2]\, P[A_1 \text{ and } B_2]}{P[X_1 = 14]} = 3/16.$$

Tables 4 and 5 contain all the results needed to evaluate the conditional probabilities specified by Equation (3.8). For example:

$$P[X_2 = 2 | X_1 = 14] =$$
$$\sum_{i=1}^{2} \sum_{j=1}^{2} P[X_2 = 2 | A_i \text{ and } B_j]\, P[A_i \text{ and } B_j | X_1 = 14] = 35/144.$$

Table 6 contains all the values of the conditional probabilities of X_2 given X_1. This is called the *predictive distribution of the random variable X_2 given the value of X_1*. Given the result of the first trial, the appropriate column of Table 6 contains the probability of each possible outcome of the second trial.

TABLE 6

$P[X_2 = m | X_1 = k]$

	Value of k		
m	0	2	14
0	17/24	7/12	7/12
2	7/36	85/288	35/144
14	7/72	35/288	25/144
Total	1	1	1

Finally, we are able to use Equation (3.7) and the entries of Table 6 to calculate the conditional expectation of X_2 given X_1. For example, for $k = 2$:

$$E[X_2|X_1 = 2] = 2P[X_2 = 2|X_1 = 2] + 14P[X_2 = 14|X_1 = 2] = 55/24.$$

TABLE 7

k	$E[X_2|X_1 = k]$
0	7/4
2	55/24
14	35/12

3.4 *Additional Remarks*

The reader should compare the prior probability estimates of Table 1 with the probability estimates of the predictive distribution of Table 6. The reader should also compare the initial pure premium estimate $E[X_1] = 2$ with those of Table 7 and note how the observed claim amount in the first period modifies the amount of claims expected in the second period.

In this work, "predictive distribution" refers only to that of a random variable given previous outcomes of one or more random variables, as in Table 6. We use the term "posterior distribution" to refer to other conditional distributions, that is, those involving one or more parameters or "states," as in Table 5. For example, in Section 4.6, we derive the *predictive* distribution of X_{m+1} given $X_1, \ldots, X_m$. To complete this derivation, we have to calculate the *posterior* distribution of a parameter (for example, state) θ given $X_1, \ldots, X_m$.

4. THE BAYESIAN MODEL VERSUS BÜHLMANN'S MODEL

4.1 *Introduction*

In this section we compare the Bayesian model described in Section 3 to a credibility model suggested by Bühlmann [2]. In Section 4.4, we show that the credibility estimates produced by the Bühlmann model are the "best" linear approximations to the corresponding Bayesian estimates. We then show that, under certain conditions, the Bühlmann credibility estimates are identical to the corresponding Bayesian estimates.

4.2 *Mathematical Preliminaries*

Let N be a random variable representing the number of claims of an insurance portfolio. So, N takes only nonnegative integer values with positive probability, that is,

$$\sum_{i=0}^{\infty} P[N=i] = 1.$$

For $i = 1, 2, \ldots$, let Y_i be a random variable representing the amount of the i-th claim, and let X represent the aggregate claim amount:

$$X = \sum_{i=1}^{N} Y_i.$$

Then we have:

Theorem 4.1:

If the Y_i are mutually independent with identical first and second moments and if the number of claims is independent of their amounts, then

(1) $E[X] = E[N]E[Y_1]$

and

(2) $\mathrm{Var}[X] = E[N]\,\mathrm{Var}[Y_1] + \mathrm{Var}[N]\,(E[Y_1])^2$

where $\mathrm{Var}[X]$ denotes the variance of the random variable X.

4.3 *Bühlmann's Credibility Formula*

If the actuary has a good deal of prior knowledge as well as substantial technical and computational resources, then he or she should probably construct a predictive distribution of the aggregate claims during the next period of interest given prior aggregate claim amounts. An alternative approach that does not explicitly require prior information or as many resources has been suggested by Bühlmann [2]. His approach is to employ a point estimate, C, of the form of Equation (1.1):

$$C = ZR + (1 - Z)H \qquad\qquad (1.1)$$

where Z is defined by

$$Z = \frac{n}{n + k} \tag{4.1}$$

and satisfies $0 \le Z \le 1$; also, n is the number of trials or exposure units, and

$$k = \frac{\text{expected value of the process variance}}{\text{variance of the hypothetical means}} \tag{4.2}$$

where the numerator and denominator of Equation (4.2) are defined in the next two sections. Henceforth, we refer to credibility estimates satisfying Equations (1.1), (4.1), and (4.2) as "Bühlmann credibility estimates."

Such estimates are of interest, because, among other properties, they are the "best" linear approximations to the Bayesian estimates. More is said about this property in Section 4.4 after we discuss the numerator and denominator of k.

4.3.1 Variance of the Hypothetical Means

In general, each *hypothetical mean* refers to the average frequency, average severity, or average aggregate claim amount (that is, pure premium) of an individual combination of risk characteristics. The hypothetical mean is the conditional expectation given that combination of risk characteristics.

For an initial example, we consider the die example of Section 2 in which one of two dice is selected with equal probability of 1/2 at random and then tossed once. We assume that we have a claim for \$1 if a marked face appears; otherwise, there is no claim. There are two combinations (or sets) of risk characteristics—namely, one corresponding to each die. The (hypothetical) mean amount of claims is equal to 1/6 for the die with one marked face and five unmarked faces, and 1/2 for the die with three marked and three unmarked faces. Because each die is selected with equal probability of 1/2, the expected claim amount (that is, the expected value of the hypothetical means) is:

$$(1/2)(1/6) + (1/2)(1/2) = 1/3.$$

So, the variance of the hypothetical means is:

$$(1/2)[(1/6) - (1/3)]^2 + (1/2)[(1/2) - (1/3)]^2 = 1/36.$$

In other words, the variance of the hypothetical means is the variance of the conditional expectations over the various sets of risk characteristics.

In the die-spinner example, the hypothetical means are the pure premium estimates for each of the four compound states or risk characteristics; see column (4) of Table 2. The variance of the hypothetical means is the weighted sum of the squared differences between each of the four pure premium estimates and the prior mean, $E[X_1] = 2$. (The prior mean is the expected value of the hypothetical means.) Again, each weight is equal to 1/4. The result is:

$$(1/4)[(2/3 - 2)^2 + (4/3 - 2)^2 + (2 - 2)^2 + (4 - 2)^2]$$

$$= 14/9 = \text{variance of hypothetical means.} \quad (4.3)$$

4.3.2 Expected Value of the Process Variance

In general, each *process variance* refers to the variance of the frequency, severity, or aggregate claim amount of an individual combination of risk characteristics. The term "process" refers to the process generating the number of claims and/or the amount of the claims. Thus, the process variance is the conditional variance given the combination of risk characteristics.

For an initial example, we again consider the die example of Section 2. We assume that for each die the number of claims has a binomial distribution. So, for a single toss of the die with one marked face the (process, that is, conditional) variance is

$$npq = (1)(1/6)(5/6) = 5/36.$$

For a single toss of the die with three marked faces, the (process) variance is

$$npq = (1)(1/2)(1/2) = 1/4.$$

Since each die is selected with probability 1/2, the expected value of the process variance is

$$(1/2)(5/36) + (1/2)(1/4) = 7/36.$$

The calculations required to compute the process variance for each risk characteristic of the die-spinner example are summarized in Table 8 and described in the following discussion.

The mean frequencies of column (3) are simply column (2) of Table 2. Since the variance of a single trial of a Bernoulli random variable is pq, the variances of column (4) are:

$$\text{Var}[I \,|A\,] = (1/6)(5/6) = 5/36$$

TABLE 8

PROCESS VARIANCE OF DIE-SPINNER EXAMPLE

(1)	(2)	Frequency		Severity		(7)
State A_iB_j	State Probability $P[A_iB_j]$	(3) Mean $E[I_1\|A_i]$	(4) Variance $Var[I_1\|A_i]$	(5) Variance $Var[S_1\|B_j]$	(6) Mean Squared $(E[S_1\|B_j])^2$	Process Variance
A_1B_1	1/4	1/6	5/36	20	16	50/9
A_1B_2	1/4	1/6	5/36	36	64	134/9
A_2B_1	1/4	1/2	1/4	20	16	14
A_2B_2	1/4	1/2	1/4	36	64	34

and

$$Var[I_1|A_2] = (3/6)(3/6) = 1/4.$$

The variances of the severity, $Var[S_1|B_j]$, which constitute column (5), are:

$$Var[S_1|B_j] = (2 - E[S_1|B_j])^2 P[S_1 = 2] + (14 - E[S_1|B_j])^2 P[S_1 = 14].$$

So, for $j = 1$,

$$Var[S_1|B_1] = (2 - 4)^2(5/6) + (14 - 4)^2(1/6) = 20$$

and for $j = 2$,

$$Var[S_1|B_2] = (2 - 8)^2(3/6) + (14 - 8)^2(3/6) = 36$$

where the mean severities, $E[S_1|B_j]$, employed to calculate the variances, $Var[S_1|B_j]$, as well as the entries of column (6) are taken from column (3) of Table 2.

Finally, using Equation (2) of Theorem 4.1, we calculate the process variances of column (7) according to:

$$E[I_1|A_i] \, Var[S_1|B_j] + Var[I_1|A_i] \, (E[S_1|B_j])^2.$$

The expected value of the process variance is obtained by multiplying the process variance of each state, column (7), by the corresponding state probability, column (2), and summing the results. This yields:

$$\text{Expected value of the process variance} = 154/9. \qquad (4.4)$$

4.3.3 Credibility Estimates

So, using Equations (4.2) – (4.4), we get for the die-spinner example:

$$k = \frac{154/9}{14/9} = 11.$$

For one observation $(n = 1)$, the credibility factor is:

$$Z = \frac{n}{n + k} = \frac{1}{1 + 11} = \frac{1}{12}.$$

Using Equation (1.1) with $Z = 1/12$ and $H = E[X_1] = 2$, we can compute C, the Bühlmann estimate given the result, R, of the first trial.

TABLE 9

BÜHLMANN CREDIBILITY ESTIMATES

Outcome R	Bühlmann Estimate
0	11/6
2	2
14	3

where for $R = 14$, for example,

$$ZR + (1 - Z)H = (1/12)14 + (11/12)2 = 3.$$

4.3.4 Characteristics of the Credibility Factor

We note the following important characteristics of the credibility factor:

$$Z = \frac{n}{n + k}.$$

(1) Z is an increasing function of n. In the limit as n approaches infinity, Z approaches 1.
(2) Since Z is a decreasing function of k, Z is also a decreasing function of the expected value of the process variance, with a lower limit of 0. Thus, the larger the variation associated with the individual combinations of risk characteristics, the less weight given to the current observations, R, and the more weight given to the prior mean, H.
(3) Finally, Z is an increasing function of the variance of the hypothetical means, with an upper limit of 1. Thus, Z increases with the variation in the expected values of the various combinations of risk characteristics.

4.3.5 Further Examples of Bühlmann Estimates

The starting point for the calculation of k is the individual combination of risk characteristics. For each combination of risk characteristics, appropriate expected values (that is, means) and variances can be calculated. The variance across the population of risk characteristics of these expected values (that is, means) is referred to as the variance of the hypothetical means. The expected value over this population of the individual variances is referred to as the expected value of the process variance. Thus, if the distributions of the individual means and variances are known, k can be calculated directly from the moments of these distributions. Two additional examples follow.

Example 4.1:

Suppose that for a group of n risks, the aggregate claim amount, X_i, of the i-th risk has mean, μ_i, and variance, σ_i^2, for $i = 1, \ldots, n$. Further, suppose that:

(1)   each of the μ_i is, in turn, selected at random from a distribution with mean, r/a, and variance, r/a^2, and
(2)   each of the σ_i^2 is chosen at random from a distribution with mean of q/b and variance, q/b^2.

The pure premium of the i-th risk is:

$$E_{\mu_i} \{E_{X_i} [X_i|\mu_i]\} = E_{\mu_i} [\mu_i] = r/a,$$

the expected value of the mean aggregate claim amount, μ_i. The variance of the expected value of the i-th individual risk is:

$$\mathrm{Var}_{\mu_i} \{E_{X_i} [X_i|\mu_i]\} = \mathrm{Var}_{\mu_i} [\mu_i] = r/a^2.$$

The expected value of the i-th individual variance is:

$$E[\sigma_i^2] = q/b,$$

where the expectation here is with respect to σ_i^2. Hence,

$$k = \frac{(q/b)}{(r/a^2)} = \frac{qa^2}{br}.$$

Example 4.2:

Each risk has a Poisson frequency distribution, $P(\lambda)$, where the parameter (that is, mean) λ is uniformly distributed over the interval $(0.07, 0.13)$. The severity component is assumed to follow distribution B_1 40 percent of the

time and B_2 60 percent of the time where the distributions B_1 and B_2 are defined by:

Severity	Severity Probabilities	
	B_1	B_2
2	5/6	1/2
14	1/6	1/2

We calculate k and Z for three periods of observation (for example, for three years).

The risk characteristics are denoted by the ordered pair (λ, B_i), where $0.07 < \lambda < 0.13$ and $i = 1,2$. (The reader will recognize B_1 and B_2 as the spinner severity distributions of Example 2.2.) Table 10, similar to Table 8, can be employed for this example. Since $E[\lambda] = 0.1$, the expected value of the process variance is equal to:

$$(0.4)(36)E[\lambda] + (0.6)(100)E[\lambda] = (74.4)E[\lambda] = 7.44.$$

TABLE 10

State	Frequency		Severity		Process Variance
	Mean	Variance	Variance	Mean Squared	
(λ,B_1)	λ	λ	20	16	36λ
(λ,B_2)	λ	λ	36	64	100λ

For state (λ,B_1) the conditional (or hypothetical) mean is 4λ, and for state (λ,B_2) it is 8λ. Thus, the expected value of the hypothetical means is:

$$(0.4)E[4\lambda] + (0.6)E[8\lambda] = (6.4)E[\lambda] = 0.64.$$

The squares of the hypothetical means are $16\lambda^2$ and $64\lambda^2$ for (λ,B_1) and (λ,B_2), respectively. The expected value of the square of the hypothetical means is:

$$
\begin{aligned}
(0.4)E[16\lambda^2] + (0.6)E[64\lambda^2] &= (44.8)E[\lambda^2] \\
&= (44.8)\{[(0.13)^3 - (0.07)^3]/(3)(0.06)\} \\
&= 0.46144.
\end{aligned}
$$

Hence, the variance of the hypothetical mean is:

$$0.46144 - (0.64)^2 = 0.05184.$$

Thus,

$$k = \frac{7.44}{0.05184} = 143.52.$$

Finally, for three periods of observation (for example, three years):

$$Z = \frac{n}{n + k} = \frac{3}{3 + 143.52} = 0.02.$$

4.3.6 Calculation of k in Practice

In practical applications k may be calculated in the following fashion:

(1) A separate sample variance is calculated for the available data on each combination of risk characteristics. The average of these sample variances is then calculated and used as the estimate of the expected value of the process variance.
(2) The available data on each individual risk are then used to calculate the total sample variance, which is then used as the estimate of the total variance.
(3) Since

$$\frac{\text{total}}{\text{variance}} = \frac{\text{expected value of the}}{\text{process variance}} + \frac{\text{variance of the}}{\text{hypothetical means}} ,$$

the variance of the hypothetical means can be obtained by subtracting the result of (1) from that of (2).
(4) Finally, k is obtained by dividing the result of (1) by that of (3).

While this is a direct method of calculating k, it is not necessarily optimal. In fact, the estimated variance of the hypothetical means produced by this method may be negative. Moreover, the estimator of the variance of the hypothetical means in (3) above is *not* an unbiased estimator. Several refinements of this procedure have been developed for its practical application.

4.4   Credibility and Least-Squares

Suppose that the actuary has decided to use $E[X_2|X_1 = k]$, the mean of the predictive distribution (see Table 7), as an updated estimate of the pure premium. Further, suppose that the actuary is unwilling to assume a particular family of distributions for (1) the process generating the claims and (2) the prior distribution of the parameters of the claims process. Such distributions are required in order to do a Bayesian analysis and thereby produce a predictive distribution. However, the actuary is willing to use a linear approximation of the mean of the predictive distribution as the (estimate of the) pure premium.

For the above situation, Bühlmann [2] and [3] has shown that the credibility estimates of Section 4.2 are the "best" linear approximations to the Bayesian estimates of the pure premium. By "best," we mean that the weighted sum (that is, expected value) of the squared differences between the linear approximations and the Bayesian estimates is minimized. The result, whose proof is given in the Appendix, can be illustrated as follows.

4.4.1 Linear Approximation Example

Let the possible outcomes of the result of a single trial be denoted by $\mathbf{R}$ = (R_1, R_2, R_3) = $(0, 2, 14)$. Let the corresponding Bayesian premium estimates $\mathbf{P}$ = (P_1, P_2, P_3) = $(7/4, 55/24, 35/12)$ as in Table 7, and let the initial or prior probabilities of the results be denoted by $\mathbf{W}$ = (W_1, W_2, W_3) = $(2/3, 2/9, 1/9)$ as in Table 1. Then, for our example, we can specify the definition of "best" linear approximation as: The "best" linear approximation to $\mathbf{P}$ is obtained by determining the values of a and b that minimize:

$$\sum_{i=1}^{3} W_i(a + bR_i - P_i)^2.$$

This is the usual weighted least-squares approach. (See, for example, Draper and Smith [5], especially pages 108–116.) The desired estimates $\hat{a}$ and $\hat{b}$ of a and b are 11/6 and 1/12, respectively, and each component of the least-squares linear approximation of the premium estimate

$$\hat{a} + \hat{b}\mathbf{R} = (11/6) + (1/12)\mathbf{R} = (11/6, 2, 3),$$

is identical to the corresponding Bühlmann credibility estimate of Table 9.

We summarize the results of this example by presenting the revised pure premium estimates, given the result of a single trial, for both the Bayesian and Bühlmann credibility models in Table 11.

TABLE 11

PURE PREMIUMS

Outcome of First Trial $\mathbf{R}$	Bühlmann Estimate $\hat{a} + \hat{b}\,\mathbf{R}$	Bayesian Estimate $\mathbf{P}$
0	11/6	7/4
2	2	55/24
14	3	35/12

4.5   *Credibility and Bayesian Inference*

In Section 4.4 we showed that the Bühlmann credibility estimate is the best linear approximation to the Bayesian estimate of the pure premium. In this section, following Jewell [16], we show that the Bühlmann credibility estimate is equal to the Bayesian estimate of the pure premium for a large class of problems. We begin with some definitions.

4.5.1 Conjugate Prior Distributions

Suppose that H is some hypothesis and B is an event. Then we can use $P[H]$ to denote one's initial or prior degree of belief in H. We call $P[H]$ the *prior probability of H*. According to the definition of conditional probability, we can consider $P[B|H]$ to represent the conditional probability of event B given hypothesis H. However, when considered as a function of H, the expression $P[B|H]$ is called the *likelihood of H on B* and is a crucial element of Bayesian analysis. Finally, $P[H|B]$ is called the *posterior probability of H*.

It is useful at this point to reconsider Bayes' Theorem:

$$P[H|B] = \frac{P[B|H]P[H]}{P[B]}. \qquad (4.5)$$

So ignoring $P[B]$, which may be considered a constant since it does not depend on H, we can interpret Equation (4.5) as a function of H as follows:

The posterior probability of H is proportional to the product of (1) the prior probability of H, $P[H]$, and (2) the likelihood of H on B, $P[B|H]$.

In Section 3 we constructed various prior and posterior probability distributions. So, we can now introduce several formal definitions.

Definition 4.1:

A distribution of prior probabilities (that is, a *prior distribution*) is said to be a *conjugate prior distribution* if the prior distribution is so related to the likelihood, that is, the conditional distribution of the current observations (or data), that the posterior distribution is the same type of distribution as the prior.

One major advantage of the use of conjugate prior distributions is that the posterior distribution for one year can be used as the prior distribution for the next year. Since no new functional forms are required, this reduces the complexity of the procedure considerably.

Definition 4.2:

The density function corresponding to the prior distribution is called the *prior density function*.

Definition 4.3:

The density function corresponding to the posterior distribution is called the *posterior density function*.

We illustrate the concept of conjugate prior distributions with two examples.

4.5.2 Binomial Distribution Example

We assume that we have a sequence of n independent Bernoulli trials with constant probability, θ, of success. We discuss the inferences that can be made about θ. We first note that

$$f(x) = \frac{(a + b + 1)!}{a!b!} x^a (1 - x)^b \tag{4.6}$$

for $0 \leq x \leq 1$, $a > -1$, and $b > -1$, is the density function of a random variable having a *beta distribution with parameters a and b*. We use Beta(a,b) to denote such a distribution and B(n,θ) to denote a binomial distribution with parameters n and θ. Given B(n,θ), the probability of r successes in n independent Bernoulli trials is:

$$\binom{n}{r} \theta^r (1 - \theta)^{n-r} \qquad \text{for } r = 0, 1, \ldots, n. \tag{4.7}$$

We are now ready to present the main result of this section.

Theorem 4.2:

If (1) n independent Bernoulli trials are performed with constrant probability, θ, of success and (2) the prior distribution of θ is Beta(a,b), then the posterior distribution of θ is Beta($a+r,b+n-r$), where r is the number of successes observed in the n trials.

Proof:

From Equation (4.6), we note that the prior density of θ is:

$$\frac{(a+b+1)!}{a!b!} \theta^a (1 - \theta)^b \qquad \text{for } 0 \leq \theta \leq 1, \tag{4.8}$$

and the likelihood is, from (4.7), proportional to

$$\theta^r (1 - \theta)^{n-r}. \tag{4.9}$$

Since the posterior density is proportional to the product of Expressions (4.8) and (4.9), the posterior density is proportional to

$$\theta^{a+r}(1 - \theta)^{b+n-r},$$

proving the result.

4.5.3. Poisson Distribution Example

We assume that we have a sequence of m independent trials from a Poisson distribution with constant mean, θ. The Poisson density function is:

$$f(n) \begin{array}{l} = \dfrac{\exp(-\theta)\theta^n}{n!} \qquad n = 0, 1, \ldots \\ = 0 \hphantom{\dfrac{\exp(-\theta)\theta^n}{n!}} \qquad \text{otherwise} \end{array} \tag{4.10}$$

We also note that for $\alpha > 0$ and $\beta > 0$:

$$g(x) \begin{array}{l} = \exp(-\beta x)\beta^\alpha x^{\alpha-1} / (\alpha - 1)! \qquad x \geq 0 \\ = 0 \hphantom{\exp(-\beta x)\beta^\alpha x^{\alpha-1} / (\alpha - 1)!} \qquad \text{otherwise} \end{array} \tag{4.11}$$

is the density function of a random variable having a *gamma distribution with parameters α and β*. We use $G(\alpha,\beta)$ to denote such a distribution.

Theorem 4.3:

If (1) m independent Poisson trials are performed with constant mean, θ, and (2) the prior distribution of θ is $G(\alpha,\beta)$, then the posterior distribution of θ is $G(\alpha + m\bar{n}, \beta + m)$, where n_i is the value drawn on the i-th trial and

$$\bar{n} = \sum_{i=1}^{m} n_i/m.$$

For insurance ratemaking, the posterior distribution for one year is the prior distribution for the next year. The use of conjugate prior distributions enables the actuary to continue to employ distributions of the same form and thereby substantially reduce the amount of computation required.

4.5.4 Credibility and Conjugate Distributions

Bailey [1] and Mayerson [20] have shown that for four pairs of likelihoods and their conjugate priors—namely, beta-binomial, gamma-Poisson, gamma-exponential, and normal-normal (with known variance)—the Bühlmann

credibility estimates are equal to the corresponding Bayesian estimate of the pure premium. We illustrate this result for the binomial and Poisson examples considered in Sections 4.5.2 and 4.5.3.

4.5.5 Binomial Distribution Example Revisited

Let I denote the indicator random variable representing the result of a single Bernoulli trial. We assume that I has probability θ of success, where θ has a beta distribution, Beta(a,b). Since the mean and variance of Beta(a,b) are $(a+1)/(a+b+2)$ and $[(a+1)(b+1)]/[(a+b+3)(a+b+2)^2]$, respectively, the value of k is:

$$k = \frac{E_\theta\{\text{Var}_i\,[I|\theta]\}}{\text{Var}_\theta\{E_I\,[I|\theta]\}} = \frac{E_\theta\,[\theta\,(1-\theta)]}{\text{Var}_\theta\,[\theta]}$$

$$= \frac{\dfrac{(a+1)}{(a+b+2)} - \dfrac{(a+1)(a+2)}{(a+b+2)(a+b+3)}}{\dfrac{(a+1)(b+1)}{(a+b+3)(a+b+2)^2}} = a+b+2.$$

So,

$$Z = \frac{n}{n+k} = \frac{n}{n+a+b+2}.$$

For the binomial distribution example of Section 4.5.2, the posterior distribution of θ is, Beta($a+r, b+n-r$). The mean of the posterior distribution of θ is

$$\frac{a+r+1}{a+b+2+n}.$$

We can rewrite the above as

$$\frac{a+1}{a+b+2+n} + \frac{r}{a+b+2+n}$$

$$= \frac{a+1}{a+b+2+n}\left(\frac{a+b+2}{a+b+2}\right) + \frac{n(r/n)}{a+b+2+n}$$

$$= \left(\frac{a+b+2}{a+b+2+n}\right)\left(\frac{a+1}{a+b+2}\right) + \left(\frac{n}{a+b+2+n}\right)(r/n)$$

Since $Z = n/(a+b+2+n)$ and noting that $(1-Z)=(a+b+2)/(a+b+2+n)$, we can rewrite the right-hand side of the above equation as:

$$(1-Z) \left(\frac{a+1}{a+b+2} \right) + Z(r/n);$$

that is, as

$$(1-Z) \text{ (mean of prior)} + Z \text{ (mean of observed data)}$$

since (1) the mean of the prior distribution of Theorem 4.2 is $(a+1)/(a+b+2)$ and (2) (r/n), the proportion of successes observed in n trials, is the observed mean of the data.

Hence, we have shown that if (1) the data can be assumed to have a binomial distribution and (2) a beta distribution is employed as the prior distribution, the mean of the posterior distribution (that is, the Bayesian estimate of the frequency component of the pure premium) is equal to the corresponding Bühlmann credibility estimate. Moreover, the mean of the posterior distribution is equal to the mean of the predictive distribution of the next observation (that is, Bernoulli trial).

4.5.6 Poisson Distribution Example Revisited

Let N denote the random variable representing the result of a single Poisson trial. We assume that N has mean θ, where θ has a gamma distribution, $G(\alpha,\beta)$. Since the mean and variance of $G(\alpha,\beta)$ are α/β and α/β^2, respectively, the value of k is:

$$k = \frac{E_\theta \{\text{Var}_N[N|\theta]\}}{\text{Var}_\theta \{E_N [N|\theta]\}} = \frac{E_\theta\{\theta\}}{\text{Var}_\theta\{\theta\}}$$

$$= \frac{(\alpha/\beta)}{(\alpha/\beta^2)} = \beta.$$

So,

$$Z = \frac{n}{n+k} = \frac{n}{n+\beta}$$

For the Poisson distribution example of Section 4.5.3, the posterior distribution of θ is $G(\alpha + m\bar{n}, \beta + m)$. So, the mean of the posterior distribution of θ is equal to:

$$\frac{\alpha + m\bar{n}}{\beta + m}.$$

We can rewrite the above as

$$\frac{\alpha}{m + \beta} + \frac{m\bar{n}}{m + \beta} = (\alpha/\beta)\frac{\beta}{m + \beta} + (\bar{n})\frac{m}{m + \beta}$$

Since $Z = m/(m + \beta)$ and noting that $1 - Z = \beta/(m + \beta)$, we can rewrite the right-hand side of the above equation as

$$(1 - Z)(\alpha/\beta) + Z(\bar{n});$$

that is, as

$$(1 - Z)(\text{mean or prior}) + Z(\text{mean of observed data})$$

since the mean of the prior distribution of Theorem 4.3 is α/β.

Hence, we have shown that if (1) the data can be assumed to have a Poisson distribution and (2) a gamma distribution is employed as the prior distribution, the mean of the posterior distribution (that is, the Bayesian estimate of the frequency component of the pure premium) is equal to the corresponding Bühlmann credibility estimate. Again, the mean of the posterior distribution is equal to the mean of the predictive distrubtion of the next observation (that is, Poisson trial).

4.5.7 Credibility and Exponential Families

We begin the section by defining the term "exponential families" and presenting a few examples of an exponential family. We then state, without proof, an important general result.

Definition 4.4:

Consider a family $\{f(x;\theta); \theta \in \Omega\}$ of probability density functions where Ω is the interval set $\Omega = \{\theta:c < \theta < d\}$ and c and d are known constants.

(a) A probability density function of the following form:

$$f(x;\theta) \begin{matrix} = \exp[p(\theta)A(x) + B(x) + q(\theta)] & a < x < b \\ = 0 & \text{otherwise} \end{matrix}$$

is said to be a member of the *exponential family* of probability density functions of the *continuous* type.

(b) A probability density function of the form:

$$f(x;\theta) \begin{matrix} = \exp[p(\theta)A(x) + B(x) + q(\theta)] \\ = 0 \end{matrix} \qquad \begin{matrix} x = x_1, x_2, \ldots \\ \text{otherwise} \end{matrix}$$

is said to be a member of the *exponential family* of the *discrete* type.

See Section 10.4 of Hogg and Craig [14] for more details on the definitions of exponential families.

Definition 4.5:

An exponential family for which $A(x) = x$ is said to be a *linear exponential family*.

Example 4.3:

Let the random variable X have a binomial density function of the form of Expression (4.7). For $x = 0, \ldots, n$, we can rewrite

$$\binom{n}{x}\theta^x(1 - \theta)^{n-x} \qquad (4.14)$$

as

$$\binom{n}{x} [\theta/(1 - \theta)]^x(1 - \theta)^n \qquad (4.15)$$

We let $y = [\theta/(1 - \theta)]^x$ and take the natural logarithm of both sides of the equation, yielding

$$\ln y = x \ln [\theta/(1 - \theta)]$$

or

$$y = \exp(\ln y) = \exp\{x \ln[\theta/(1 - \theta)]\}.$$

Hence, we can rewrite Expression (4.15) as

$$\binom{n}{x} \exp\{x \ln[\theta/(1 - \theta)]\}(1 - \theta)^n.$$

Since $\binom{n}{x}$ and $(1 - \theta)^n$ can be written in the form $\exp[B(x)]$ and $\exp[q(\theta)]$, respectively, and since we can let $A(x) = x$ and $p(\theta) = \ln[\theta/(1 - \theta)]$, the

binomial density function is indeed an example of a linear exponential family of the discrete type.

The Poisson, exponential, and normal (with known variance) distributions are also examples of linear exponential families.

Ericson [7] and Jewell [16] have shown that the class of likelihood-prior families for which the Bühlmann credibility estimate of the pure premium is equal to the corresponding Bayesian estimate could be extended beyond the four cases discussed in the first paragraph of Section 4.5.4. Specifically Ericson [7] has shown the following.

Theorem 4.4:

If (1) the likelihood density is a member of a linear exponential family and (2) the conjugate prior distribution is used as the prior distribution, the Bühlmann credibility estimate of the pure premium is equal to the corresponding Bayesian estimate.

4.6   *Frequency-Severity Model with Continuous Severity Component*

In this section we present another two-stage model of an insurance operation. The frequency component is assumed to be based on a Poisson distribution, while the severity component is based on an exponential distribution. In order to describe this two-stage model, we need to return to the concept of a predictive distribution, which we discussed earlier in connection with Table 3. We assume here that the number of claims is independent of the amount of individual claim losses.

4.6.1 Predictive Distributions

We begin by defining X_i as the random variable representing the amount of aggregate claims during the i-th policy year (or, equivalently, during the i-th period of observation). Our goal is to calculate the predictive (or probability) distribution of X_{m+1} given $X_1, \ldots, X_m$. We assume that given a parameter θ, the random variables $X_1, \ldots, X_{m+1}$ are independent and identically distributed with density function p. We use f to denote the density function of θ. So, we write the conditional density of X_{m+1} given $X_1 = x_1, \ldots, X_m = x_m$, as:

$$\frac{\int p(x_{m+1}|\theta) \prod_{i=1}^{m} p(x_i|\theta) f(\theta) d\theta}{\int \prod_{i=1}^{m} p(x_i|\theta) f(\theta) d\theta}$$

where

$$\frac{\prod\limits_{i=1}^{m} p(x_i|\theta)f(\theta)}{\int \prod\limits_{i=1}^{m} p(x_i|\theta)f(\theta)d\theta}$$

is the posterior density of θ given $X_1 = x_1, \ldots, X_m = x_m$.

4.6.2 Frequency Component

For $i = 1, 2, \ldots$, let N_i be the random variable representing the number of claims during the i-th period of observation. We assume that N_i has a Poisson distribution with parameter (for example, mean) λ. Given m observations $n_1, n_2, \ldots, n_m$, we assume that the posterior distribution of λ is $G(\alpha + m\bar{n}, \beta + m)$, as shown in Section 4.5. The parameters α and β determine the prior gamma distribution. The data are summarized by

$$m\bar{n} = \sum_{i=1}^{m} n_i \text{ and } m.$$

We use $g(\lambda)$ to denote the density function of $G(\alpha + m\bar{n}, \beta + m)$. So, we can write the conditional probability of $N_{m+1} = n$ given $N_1 = n_1, \ldots, N_m = n_m$ as:

$$\frac{\int_0^\infty [\exp(-\lambda)\lambda^n/n!]g(\lambda)d\lambda}{\int_0^\infty g(\lambda)d\lambda}$$

$$= \frac{\int_0^\infty [\exp(-\lambda)\lambda^n/n!]\exp[-(\beta + m)\lambda]\lambda^{\alpha+m\bar{n}-1}d\lambda}{\int_0^\infty \exp[-(\beta + m)\lambda]\lambda^{\alpha+m\bar{n}-1}d\lambda}$$

$$= \frac{(1/n!) \int_0^\infty \exp[-(\beta + m + 1)\lambda]\lambda^{\alpha+m\bar{n}-1+n}d\lambda}{\Gamma(\alpha + m\bar{n}) (\beta + m)^{-(\alpha+m\bar{n})}}$$

$$= \frac{(1/n!)\Gamma(\alpha + m\bar{n} + n)(\beta + m + 1)^{-(\alpha+m\bar{n}+n)}}{\Gamma(\alpha + m\bar{n})(\beta + m)^{-(\alpha+m\bar{n})}}$$

$$= \binom{\alpha + m\bar{n} + n - 1}{n}\left(\frac{1}{\beta + m + 1}\right)^n\left(\frac{\beta + m}{\beta + m + 1}\right)^{\alpha+m\bar{n}}$$

$$n = 0, 1, \ldots$$

which is in the form of a negative binomial density function.

The mean of the above negative binomial density function is $(\alpha + m\bar{n})/(\beta + m)$, which was shown in Section 4.5.6 to equal the Bühlmann credibility estimate (of the frequency of claims).

The above negative binomial density represents the predictive density function, given $N_1 = n_1, \ldots, N_m = n_m$, of the number of claims in the $(m + 1) - st$ period of observation. The index $n = n_{m+1}$ takes on the values $0, 1, \ldots$

The above predictive density function provides (estimated) probabilities of each of the possible number of claims. Hence, the predictive density provides much more information than is provided by only its (estimated) mean.

4.6.3 Severity Component

We assume that the amount of each individual claim has an exponential* distribution with probability density function given by:

$$p(y|\delta) = \exp(-y/\delta)/\delta \qquad y > 0, \delta > 0.$$

We have $E_Y[Y|\delta] = \delta$ and $\text{Var}_Y[Y|\delta] = \delta^2$. The mean claim amount, δ, has a conjugate prior distribution whose probability density function, $f(\delta|m', x')$, is proportional to:

$$\exp(-x'/\delta)/\delta^{m'} \qquad x' > 0, m' > 2, \delta > 0.$$

Such a density function is called an *inverted gamma density function* and has mean $x'/(m' - 2)$. The insurance process is observed for m periods of observation with n_i claims incurred in period i. The total aggregate claim amount over the m periods of observation is:

$$x = \sum_{i=1}^{m\bar{n}} y_i.$$

Therefore, the posterior distribution of δ, $f(\delta|m', x', m\bar{n}, x)$, is proportional to:

$$\left[\prod_{i=1}^{m\bar{n}} \exp(-y_i/\delta)/\delta \right] \exp(-x'/\delta)/\delta^{m'} = \exp[-(x' + x)/\delta]/\delta^{m' + m\bar{n}},$$

*The exponential distribution was chosen primarily to keep the level of mathematical sophistication relatively low. Other probability distributions, such as the log-normal distribution, frequently provide a more accurate representation of the actual distribution of losses. See Hewitt [12] for a discussion of a two-stage Bayesian credibility model in which the loss distribution is assumed to be log-normal. Other references on loss distributions are Hewitt [13] and Hogg and Klugman [15].

which is also an inverted gamma distribution. The mean of the posterior distribution of δ is given by:

$$\frac{x' + x}{m' + m\bar{n} - 2} = \left(\frac{m' - 2}{m' + m\bar{n} - 2}\right)\left(\frac{x'}{m' - 2}\right)$$

$$+ \left(\frac{m\bar{n}}{m' + m\bar{n} - 2}\right)\left(\frac{x}{m\bar{n}}\right) = (1 - Z)\,(\text{Prior Mean}) + Z(\text{Data Mean}).$$

The predictive density of Y, which reflects the uncertainty in the estimation of the parameter values as well as in the random nature of the claim amounts, is given by:

$$p(y|m', x', m\bar{n}, x) = \int_0^\infty p(y|\delta)f(\delta|m', x', m\bar{n}, x)d\delta$$

$$= C \int_0^\infty [\exp(-y/\delta)/\delta]\,\{\exp[-(x' + x)/\delta]/\delta^{m' + m\bar{n}}\}d\delta \qquad (4.16)$$

where
$$C = \frac{(x' + x)^{m' + m\bar{n} - 1}}{\Gamma\,(m' + m\bar{n} - 1)}.$$

The expression on the right-hand side of Equation (4.16) can be rewritten as:

$$C \int_0^\infty [\exp(-y + x' + x)/\delta]\delta^{-m' - m\bar{n} - 1}\,d\delta. \qquad (4.17)$$

Making the change of variable $w = (y + x' + x)/\delta$, and noting that $d\delta = -[(y + x' + x)/w^2]dw$, we can rewrite Expression (4.17) as:

$$\left(\frac{(x' + x)^{m' + m\bar{n} - 1}}{\Gamma\,(m' + m\bar{n} - 1)}\right)\left(\frac{\int_0^\infty w^{m' + m\bar{n} - 1}\exp(-w)dw}{(y + x' + x)^{m' + m\bar{n}}}\right)$$

$$= \left(\frac{(x' + x)^{m' + m\bar{n} - 1}}{\Gamma\,(m' + m\bar{n} - 1)}\right)\left(\frac{\Gamma\,(m' + m\bar{n})}{(y + x' + x)^{m' + m\bar{n}}}\right) \qquad (4.18)$$

$$= \frac{(x' + x)^{-1}(m' + m\bar{n} - 1)}{\left(1 + \dfrac{y}{(x' + x)}\right)^{m' + m\bar{n}}}$$

which is a member of the Pareto family of distributions.

In practice, the above predictive distribution should be used for the claims occurring during the $(m + 1) - st$ period of observation. When there is substantial uncertainty about the value of the parameter λ, the relatively tame exponential distribution gets transformed to the heavy-tailed Pareto distribution.

The following example illustrates an important application of the predictive distribution of claim amounts.

Example 4.4:

Suppose that 17 claims are observed during the first four periods of observation and that the total aggregate claim amount is $1,000,000. Using the assumptions of this section, find

$$P[Y_{18} \geq \$100,000|\sum_{i=1}^{17} y_i = \$1,000,000].$$

Solution:

We have

$$m\bar{n} = \sum_{i=1}^{m} n_i = 17$$

and

$$x = \sum_{i=1}^{m\bar{n}} y_i = \$1,000,000.$$

So, using the Pareto density function above, we find the desired probability to be:

$$\int_{100,000}^{\infty} \frac{(x + x')^{-1}(m' + m\bar{n} - 1)}{\left(1 + \dfrac{y}{x + x'}\right)^{m' + m\bar{n}}} \, dy = \left(1 + \frac{y}{x + x'}\right)^{-m' - m\bar{n} + 1}\Bigg|_{100,000}^{\infty}$$

$$= \left(1 + \frac{100,000}{1,000,000 + x'}\right)^{-m' - 16}$$

4.6.4 The Two-Stage Model

In Sections 4.6.2 and 4.6.3 we presented the predictive distributions of the number of claims and the amount of the claim losses, respectively. We

further assumed that the random variables corresponding to the two distributions are independent. Ideally, we would like to use the two predictive distributions to construct a simple, closed-form expression for the predictive distribution of aggregate claim amounts, that is, a predictive distribution of the random sum of random variables

$$X_{n+1} = \sum_{i=1}^{N} L_{i,n+1}$$

where $L_{i,n+1}$ is the amount of the i-th loss during the $(n+1)$-st period of observation. Such a distribution can be used to give the probabilities that losses will exceed various amounts and thereby serve as an excellent management tool. Unfortunately, in most situations it is not possible to calculate a closed-form expression for the predictive distribution. One approach is to simulate the predictive distribution, that is, to use stochastic simulation methods to construct an empirical version of the predictive distribution of aggregate claims.

One illustration of the use of stochastic simulation to construct an empirical distribution is Example 1 in Herzog [10]. A more practical application of the simulation of a two-stage model in a Bayesian context is found in Herzog and Rubin [11].

An alternative procedure is described in Heckman and Meyers [9]. They present an algorithm for determining the aggregate loss distribution by inverting a characteristic function. This entails approximating the claim severity distribution function by a piecewise linear cumulative distribution function. Given this severity distribution, the Heckman-Meyers procedure produces exact results and typically requires less computer time than does a stochastic simulation approach; however, the actuary may find the Heckman-Meyers procedure considerably more complex, at least initially.

4.6.5 The Pure Premium

Theorem 4.1 showed that if X is a random variable representing the total aggregate claims in an epoch and if the claims frequency and claims severity processes may be assumed to be independent, then:

$$E[X] = E[N]E[Y_1].$$

Here, N is the random variable representing the number of claims in an epoch, and Y_1 is the random variable representing the claim amount of an

individual claim. The expected aggregate claim amount, $E[X]$, is also called the *pure premium*, as indicated in Section 3.

If N has a Poisson distribution and Y_1 has an exponential distribution, as above, then the pure premium is given by $\lambda\delta$. If we assume that λ and δ are independent and use the posterior means as given in Sections 4.6.2 and 4.6.3 to estimate λ and δ, we obtain:

$$\text{Pure Premium} = \left[\frac{\alpha + m\bar{n}}{\beta + m}\right]\left[\frac{x' + x}{m' + m\bar{n} - 2}\right].$$

The first term in brackets is the credibility estimate of λ that was derived in Section 4.5.6. The second term in brackets is the credibility estimate of δ.

4.6.6 An Alternative Estimate of the Pure Premium

Alternatively, we could compute an estimate of the pure premium by using a Bühlmann credibility model in which only the aggregate claim amounts of each insured are considered; that is, the frequency and severity components are *not* considered separately. The number of claims, N, is assumed to have a Poisson distribution with mean $E[N] = \lambda$. Moreover, $\text{Var}[N] = \lambda$, and λ is assumed to have the gamma distribution, $G(\alpha, \beta)$, implying that

$$E[\lambda] = \alpha/\beta, \ E[\lambda^2] = \alpha(\alpha + 1)/(\beta)^2, \text{ and } \text{Var}[\lambda] = \alpha/(\beta)^2.$$

The amount of an individual claim, Y_1, is assumed to have an exponential distribution with mean δ and variance δ^2. Finally, δ has the inverted gamma distribution presented in Section 4.6.3, so that

$$E[\delta] = \frac{x'}{m' - 2}, \ E[\delta^2] = \frac{(x')^2}{(m' - 2)(m' - 3)},$$

and

$$\text{Var}[\delta] = \frac{(x')^2}{(m' - 2)(m' - 3)^2}.$$

For state (λ, δ) the hypothetical mean is $\lambda\delta$. So, the variance of the hypothetical means is:

$$\text{Var}[\lambda\delta] = E[\lambda^2\delta^2] - (E[\lambda\delta])^2$$
$$= E[\lambda^2]E[\delta^2] - (E[\lambda])^2(E[\delta])^2$$

$$= \left(\frac{\alpha(\alpha + 1)}{(\beta)^2}\right)\left(\frac{(x')^2}{(m' - 2)(m' - 3)}\right) - \left(\frac{\alpha^2}{\beta^2}\right)\left(\frac{(x')^2}{(m' - 2)^2}\right)$$

$$= \frac{\alpha(x')^2(\alpha + m' - 2)}{(\beta)^2(m' - 2)^2\ (m' - 3)}.$$

The process variance of state (λ,δ) is, using Theorem 4.2:

$$E[N]Var[Y_1] + Var[N]\ (E[Y_1])^2 = \lambda\delta^2 + \lambda\delta^2 = 2\lambda\delta^2.$$

So, the expected process variance is

$$E[2\lambda\delta^2] = 2E[\lambda]E[\delta^2] = \frac{2\alpha(x')^2}{\beta(m' - 2)(m' - 3)}.$$

Hence,

$$k = \frac{2\beta(m' - 2)}{\alpha + m' - 2}.$$

The prior mean is:

$$E[\lambda\delta] = \frac{\alpha x'}{\beta(m' - 2)}.$$

We observe m years of data, so $Z = m/(m + k)$ and x/m is the mean of the observations. Therefore, the credibility estimate is:

$$Z(x/m) + (1 - Z)\ (\text{prior mean}) = \frac{2\alpha x' + \alpha x + xm' - 2x}{m(\alpha + m' - 2) + 2\beta(m' - 2)}.$$

Rewriting the credibility estimate as

$$\frac{2\alpha x' + (\alpha + m' - 2)x}{2\beta(m' - 2) + (\alpha + m' - 2)m}$$

facilitates comparison with the prior mean.

In comparing the alternative estimate of the pure premium to the original estimate, note that $m\bar{n}$, the total number of claims during the m years, is not needed for the Bühlmann estimate. Finally, we note that the pure premium estimator of Section 4.6.5 has a smaller mean squared error than does the Bühlmann estimator of this section.

4.7 *Summary of Section 4*

In Table 12, which is similar to page 13 of the doctoral dissertation of Morgan [22], we summarize much of the material of Section 4. For example, the second column summarizes the results of Sections 4.5.2 and 4.5.5. Here, we assume that the data are from a Bernoulli random variable and that the prior distribution of the probability of success, θ, has a Beta distribution. This implies that the hypothetical mean and process variance are θ and $\theta(1-\theta)$, respectively. These lead to a k value of $a+b+2$ and a predictive density having the form of the Beta density function. The third column summarizes the results of Sections 4.5.3, 4.5.6, and 4.6.2. Finally, the last column summarizes the results of Section 4.6.3.

TABLE 12

CHARACTERISTICS OF SIMPLE EXPONENTIAL FAMILIES

Likelihood	Bernoulli: $B(1,\theta)$	Poisson: $P(\theta)$	Exponential: $\exp(-y/\delta)/\delta$
Prior	Beta: Beta(a,b)	Gamma: $G(\alpha,\beta)$	Inverted Gamma: $\exp(-x'/\delta)/\delta^{m'}$
Hypothetical mean	θ	θ	δ
Process variance	$\theta(1-\theta)$	θ	δ^2
Variance of the hypothetical means	$\dfrac{(a+1)(b+1)}{(a+b+3)(a+b+2)^2}$	α/β^2	$\dfrac{(x')^2}{(m'-2)^2(m'-3)}$
Expected process variance	$\dfrac{(a+1)(b+1)}{(a+b+3)(a+b+2)}$	α/β	$\dfrac{(x')^2}{(m'-2)(m'-3)}$
k	$a+b+2$	β	$m'-2$
Predictive density	Beta	Negative binomial	Pareto

5. REVIEW AND CONCLUDING REMARKS

In Sections 2 and 3, we reviewed the basic concepts of Bayesian analysis, showed how these concepts could be used in the insurance ratemaking process, and illustrated these ideas with an example originally presented by Hewitt [12]. The key ideas were the use of the predictive distribution of aggregate claim amounts (Table 6) and the use of the (Bayesian) conditional mean to estimate pure premium amounts.

In Section 4, we discussed a credibility model proposed by Bühlmann [2]. We showed that this model produced least-squares linear approximations to the Bayesian estimate of the pure premium. This important result was first demonstrated by using the same example employed in Section 3. A detailed proof is presented in the Appendix.

We have also discussed an important, general result of Ericson [7]. It turns out that some specific results described in Mayerson [20] are simply special cases of Ericson's more general result. Using the statistical machinery developed to describe Ericson's result, we presented another two-stage model in which the number of claims had a Poisson distribution and the claim amounts had an exponential distribution.

The emphasis here has been on basic statistical concepts. We have not discussed important practical issues such as how to apply these procedures in dealing with issues likely to be encountered in real-life situations. Moreover, we did not attempt to discuss more sophisticated theoretical concepts such as multivariate extensions of the results presented here. These are all left for a more advanced treatment elsewhere.

We do hope that this brief work has provided the reader with a stimulating and informative introduction to Bayesian credibility.

ACKNOWLEDGMENTS

REFERENCES

1. BAILEY, A.L. "Credibility Procedures—La Place's Generalization of Bayes' Rule and the Combination of Collateral Knowledge with Observed Data," *Proceedings of the Casualty Actuarial Society* 37 (1950):7–23; Discussion, 94–115.
2. BÜHLMANN, H. "Experience Rating and Credibility," *The ASTIN Bulletin* 4 (1967): 119–207.
3. BÜHLMANN, H. *Mathematical Methods in Risk Theory*. New York: Springer-Verlag, 1970.
4. BÜHLMANN, H. and STRAUB, E. "Credibility for Loss Ratios" (Translated by C.E. Brooks), *Actuarial Research Clearinghouse* 1972.2 (1972).
5. DRAPER, N. and SMITH, H. *Applied Regression Analysis*, 2d ed. New York: Wiley, 1981.

6. EDWARDS, W., LINDMAN, H., AND SAVAGE, L.J. "Bayesian Statistical Inference for Psychological Research," *70 Psychological Review* 193 (1963).

7. ERICSON, W.A. "On the Posterior Mean and Variance of a Population Mean," *Journal of the American Statistical Association* 65, No. 330 (1970): 649–52.

8. FUHRER, C. "Some Applications of Credibility Theory to Group Insurance," *TSA* XL, Part I (1989): 387–404.

9. HECKMAN, P.E. AND MEYERS, G.G., "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *Proceedings of the Casualty Actuarial Society* 70 (1983):22–74; 71 (1984):49–66.

10. HERZOG, T.N. *An Introduction to Stochastic Simulation*, Part 3 Study Note. Itasca, Il.: Society of Actuaries, 1983.

11. HERZOG, T.N. AND RUBIN, D.B. "The Use of Multiple Imputations to Handle Nonresponse," Chapter 15 of *Nonresponse in Sample Surveys: The Theory of Current Practice*. New York: Academic Press, 1983.

12. HEWITT, C.C., JR. "Credibility for Severity," *Proceedings of the Casualty Actuarial Society* 57 (1970):148–71; Discussion 58 (1971): 25–31.

13. HEWITT, C.C. "Loss Ratio Distributions—A Model," *Proceedings of the Casualty Actuarial Society* 54 (1967):70–93.

14. HOGG, R.V. AND CRAIG, A.T. *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan, 1978.

15. HOGG, R.V. AND KLUGMAN, S. *Loss Distributions*. New York: Wiley, 1984.

16. JEWELL, W.S. "The Credible Distribution," *The ASTIN Bulletin* 7 (1974):237–69.

17. JEWELL, W.S. "A Survey of Credibility Theory," Operations Research Center, University of California, Berkeley, 1976.

18. KLUGMAN, S.A. "Credibility for Classification Ratemaking via the Hierarchical Normal Linear Model," *Proceedings of the Casualty Actuarial Society* 74 (1987): 272–321.

19. LONGLEY-COOK, L.H. "An Introduction to Credibility Theory," *Proceedings of the Casualty Actuarial Society* 49 (1962): 194–221; available as a separate report from Casualty Actuarial Society, 250 West 34th St., New York, NY 10019.

20. MAYERSON, A.L. "A Bayesian View of Credibility," *Proceedings of the Casualty Actuarial Society*, 51 (1964):85-104; Discussion, 52 (1965):121–27.

21. MOOD, A.M. GRAYBILL, V.A., AND BOES, D.C. *Introduction to the Theory of Statistics*, 3d ed. New York: McGraw-Hill, 1974.

22. MORGAN, I.M. *Credibility Theory under the Collective Risk Model*, Ph.D. diss., University of Wisconsin, 1983.

23. PHILBRICK, S. "An Examination of Credibility Concepts," *Proceedings of the Casualty Actuarial Society* 48 (1981):195–219.

APPENDIX

LINEAR APPROXIMATION—THEORY

Here we present a proof that the Bühlmann credibility estimates are the "best" linear approximations to the Bayesian estimates of the pure premium (that is, to the posterior mean). The proof is based on Bühlmann [2] and [3, pp. 100–103].

We begin by assuming that the random variables $X_1, \ldots, X_n$, which represent aggregate claim losses for years $1, \ldots, n$, respectively, are identically distributed. Further, we assume that their common distribution function, $F(X_i, \theta)$, has mean $\mu(\theta)$ and variance $\sigma^2(\theta)$, where θ represents an unknown parameter value (or possibly a vector) that must be described by means of a prior distribution. Finally, we assume $X_1, X_2, \ldots, X_n$ are mutually independent given θ.

Our problem is to determine the constants a and b, which specify the linear approximation $a + b\bar{X}$ where

$$\bar{X} = \left(\sum_{i=1}^{n} X_i \right) / n.$$

These constants are to be chosen so as to minimize the expression:

$$E_{\mathbf{X}} \{ (E_{\theta|\mathbf{X}} [\mu(\theta)|\mathbf{X}] - a - b\bar{X})^2 \} \qquad (6.1)$$

where

(1) $\mathbf{X} = (X_1, \ldots, X_n)$,
(2) $E_{\theta|\mathbf{X}}$ denotes integration over the probability space of the random variable θ given $\mathbf{X}$, and
(3) $E_{\mathbf{X}}$ denotes integration over the probability space of the random vector $\mathbf{X}$.

We note that by the definition of $\mu(\theta)$

$$E_{\mathbf{X}|\theta}[X] = \mu(\theta)$$

and so

$$E_{\mathbf{X}|\theta}[\bar{X}] = \mu(\theta). \qquad (6.2)$$

Moreover,

$$E_{\mathbf{X}}\{E_{\theta|\mathbf{X}}[\mu(\theta)|X]\} = E_\theta[\mu(\theta)] \qquad (6.3)$$

and

$$E_{\mathbf{X}}[\overline{X}] = E_\theta E_{\mathbf{X}|\theta}[\overline{X}] = E_\theta[\mu(\theta)]. \qquad (6.4)$$

Thus, we seek constants a and b, which minimize the expected squared deviations between the posterior mean, $E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}]$, and the linear approximation, $a + b\overline{X}$. To determine a and b, we first differentiate Expression (6.1) with respect to a, set the result equal to zero, and solve for a. This yields:

$$\hat{a} = E_{\mathbf{X}}\{E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}]\} - bE_{\mathbf{X}}[\overline{X}].$$

Using Equations (6.3) and (6.4), we can rewrite the last expression as:

$$\hat{a} = E_\theta[\mu(\theta)] - bE_\theta[\mu(\theta)] = (1 - b)E_\theta[\mu(\theta)]. \qquad (6.5)$$

In order to determine b, we first replace a in Expression (6.1) by $\hat{a}$ as specified in Equation (6.5), producing

$$E_{\mathbf{X}}\{(E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}] - (1 - b)E_\theta[\mu(\theta)] - b\overline{X})^2\}.$$

Differentiating the above expression with respect to b, we obtain

$$2E_{\mathbf{X}}\{(E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}] - (1 - b)E_\theta[\mu(\theta)] - b\overline{X})(E_\theta[\mu(\theta)] - \overline{X})\}.$$

Setting the last expression equal to zero and solving for b gives us

$$\hat{b} = \frac{E_{\mathbf{X}}\{(E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}] - E_\theta[\mu(\theta)])\ (\overline{X} - E_\theta[\mu(\theta)])\}}{E_{\mathbf{X}}\{\overline{X} - E_\theta[\mu(\theta)]\}^2}. \qquad (6.6)$$

Since we have already established in Equation (6.4) that

$$E_\theta[\mu(\theta)] = E_{\mathbf{X}}[\overline{X}],$$

the denominator of $\hat{b}$ is simply equal to $\mathrm{Var}[\overline{X}]$. We now examine the numerator of (6.6):

$$E_{\mathbf{X}}\{(E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}] - E_\theta[\mu(\theta)])(\overline{X} - E_\theta[\mu(\theta)])\} = E_{\mathbf{X}}\{\overline{X}E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}]\}$$

$$- E_{\mathbf{X}}\{E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}]E_\theta[\mu(\theta)] - E_{\mathbf{X}}\{\overline{X}E_\theta[\mu(\theta)] + E_{\mathbf{X}}\{E_\theta[\mu(\theta)]\}^2$$

$$= E_\theta\{E_{\mathbf{X}|\theta}[\overline{X}\mu(\theta)\} - (E_\theta[\mu(\theta)])^2 - (E_\theta[\mu(\theta)])^2 + (E_\theta[\mu(\theta)])^2$$

$$= E_\theta[(\mu(\theta))^2] - \{E_\theta[\mu(\theta)]\}^2 = \mathrm{Var}[\mu(\theta)].$$

Hence, we can rewrite Equation (6.6) as:

$$\hat{b} = \frac{\mathrm{Var}[\mu(\theta)]}{\mathrm{Var}[\overline{X}]}.$$

So we can approximate $E_{\theta|\mathbf{X}}[\mu(\theta)|\mathbf{X}]$ by:

$$\hat{a} + \hat{b}\overline{X} = (1 - \hat{b})E_{\theta}[\mu(\theta)] + \hat{b}\overline{X} \tag{6.7}$$

Since $\text{Var}_{\overline{X}}[\overline{X}] = E_{\theta}(\text{Var}_{\overline{X}}[\overline{X}|\theta]) + \text{Var}_{\theta}(E_{\overline{X}}[\overline{X}|\theta])$, $\text{Var}_{\overline{X}}[\overline{X}|\theta] = (1/n)\,\text{Var}_X[X|\theta]$, and $E_{\overline{X}}[X|\theta] = \mu(\theta)$,

$$\hat{b} = \frac{\text{Var}[\mu(\theta)]}{\text{Var}[\overline{X}]} = \frac{\text{Var}[\mu(\theta)]}{(1/n)E(\text{Var}[X|\theta]) + \text{Var}[\mu(\theta)]}$$

$$= \frac{n\text{Var}[\mu(\theta)]}{E\{\text{Var}[X|\theta]\} + n\text{Var}[\mu(\theta)]}$$

$$= \frac{n}{n + \dfrac{E\{\text{Var}[X|\theta]\}}{\text{Var}[\mu(\theta)]}}$$

$$= Z$$

So we can rewrite Equation (6.7) as:

$$\hat{a} + \hat{b}\overline{X} = (1 - Z)E_{\theta}[\mu(\theta)] + Z\overline{X} \tag{6.8}$$

where $Z = \dfrac{n}{n + k}$ and $k = \dfrac{E\{\text{Var}[X|\theta]\}}{\text{Var}[\mu(\theta)]}$; that is,

$$k = \frac{\text{expected value of the process variance}}{\text{variance of the hypothetical means}}.$$

DISCUSSION OF PRECEDING PAPER

CHARLES S. FUHRER:

Dr. Herzog is to be commended for an excellent treatment that makes some difficult concepts clearer. We have had very few papers on credibility in the *Transactions*.

Claims Frequency and Severity: An Alternative Method

The last formula in Section 4.6.5 sets the pure premium equal to a credibility estimate of the frequency of claims multiplied by a credibility estimate of the size of claim. Although each of these estimates gives the exact mean of the respective posterior distributions, they are actually linear formulas of the general form:

$$Z \text{ (Data Mean)} + (1 - Z) \text{ (Prior Mean)}.$$

Furthermore, each of these two Z's is of the general form $Z = n/(n + k)$, where n is the volume of exposure and $k =$ (expected value of the process variance)/ (variance of the hypothetical means). In case we do not have conjugate likelihoods and priors (or in general applications an unknown prior), as in the paper, the two estimates are still valid as the best linear least-squares estimates for each posterior mean.

Unfortunately, the product of the two estimates is no longer linear in the number-of-claims random variable. It is no longer clear that the formula in Section 4.6.5 is optimum, in any sense, for estimating the total claims.

Jewell [8] derived an alternative approach. He uses the equivalent of the formula:

[Pure Premium] $\approx Z_1$ (number of claims)

$$+ Z_2 \text{ (total claims)} + \text{(constant)}.$$

The two Z's are then determined by least squares. For a simpler version of this method, under the assumption that the number and size of claims are independent, see Bühlmann [2].

Revision of Proof in Appendix

I found the proof in the Appendix somewhat confusing. For example, it does not make clear which assumptions are necessary to derive which formulas. The following proof is easier to understand. This proof largely follows Gerber [4]. The proof comprises three parts. The first part merely states

83

a result from statistics. The second part uses this result with the minimum necessary assumptions to obtain $Z = n/(n + k)$. The third part shows how these assumptions result from the parameterization that Bühlmann and Straub [3], Bühlmann [1], and others used.

I. First, note that the problem is to approximate the conditional expectation:

$$E(X_{n+1} \mid X_1 \ldots X_n)$$

with the linear expression $a + b\overline{X}$, where:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

in such a way as to minimize

$$E\ [a\ +\ b\overline{X}\ -\ E(X_{n+1}|X_1 \ldots X_n)]^2,$$

the expected squared error. The solution is:

$$b\ =\ \mathrm{cov}(\overline{X}, X_{n+1})/\mathrm{var}(\overline{X}) \tag{1}$$

and

$$a\ =\ E(X_{n+1})\ -\ bE(\overline{X})\ =\ \mu_{n+1}\ -\ b\mu,$$

where $\mu_{n+1} = E(X_{n+1})$ and $\mu = E(\overline{X})$. The b here is not $\hat{b}$; it is a function of the random variables, not an estimate from a sample. The proof of (1) can be found in some statistics texts such as Hoel, Port and Stone [6, pages 43–44], which is included in the examination 110 syllabus. Note that

$$\rho\sigma_2/\sigma_1\ =\ \mathrm{cov}\ (X_1, X_2)/\mathrm{var}(X_1).$$

Surprisingly, many statistics textbooks do not prove this formula. They only prove a similar formula for the slope of the least squares line, fit to a set of sample points. Hogg and Craig [7] do not actually prove it. On page 75, they derive Equation (1), but they use the assumption that the conditional expectation is exactly a linear function.

This is a good spot to point out that this method (called Bühlmann's in Herzog) can be generalized. We could approximate $E(X_{n+1}|X_1 \ldots X_n)$ with the function:

$$a\ +\ \sum_{i=1}^{n} b_i X_i$$

(this is the approach used above). Nonlinear functions also can be used and might fit better. Other loss functions besides squared error could be minimized.

The proof can be to set the partial derivatives with respect to a and b of the expression for the squared error equal to zero. An alternative derivation recognizes that the covariance operator is essentially an inner product on the space of random variables, which is a Hilbert space. We can now use the orthogonality property in the projection theorem to prove Equation (1). This approach is used in Gerber and Jones [5, appendix], also in a credibility setting.

II. Now assume:

$$\text{var}(X_i) = \text{var}(X_j) = A + B \text{ and cov } (X_i, X_j) = A, i \neq j. \tag{2}$$

Substituting the definition of $\overline{X}$ into (1) gives:

$$Z = \frac{\text{cov}\left[\frac{1}{n}\sum_{i=1}^{n} X_i, X_{n+1}\right]}{\text{var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]} = \frac{\frac{1}{n}\sum_{i=1}^{n}\text{cov}\left[X_i, X_{n+1}\right]}{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{cov}\left[X_i, X_j\right]} \tag{3}$$

Because cov is a bilinear operator, we can write (3) as:

$$Z = \frac{\frac{1}{n}nA}{\frac{1}{n^2}\left[(n^2 - n)A + n(A + B)\right]} = \frac{A}{A + \frac{B}{n}} = \frac{n}{n + k}, \tag{4}$$

where $k = B/A$.

III. We can give some meaning to the constants A and B in terms of a parameter as follows: For any random variables X_i and parameter ϑ:

$$\text{var}(X_i) = E[\text{var}(X_i|\vartheta)] + \text{var}[E(X_i|\vartheta)] \tag{5}$$

and for $i \neq j$:

$$\text{cov}(X_i, X_j) = E[\text{cov}(X_i, X_j|\vartheta)] + \text{cov}[E(X_i|\vartheta), E(X|\vartheta)] \tag{6}$$

These follow directly from the definitions of var and cov.

If we assume that the X_i's are independent and identically distributed given ϑ, then:

$$\mathrm{cov}(X_i, X_j|\vartheta) = 0 \text{ and } \mathrm{cov}[E(X_i|\vartheta), E(X_j|\vartheta)] = \mathrm{var}[E(X_i|\vartheta)]. \quad (7)$$

Substituting (7) into (6) gives:

$$\mathrm{cov}(X_i, X_j) = \mathrm{var}[E(X_i|\vartheta)] \quad (8)$$

Now (2), (5), and (8) give the result:

$$A = \mathrm{var}[E(X|\vartheta)],$$

$$B = E[\mathrm{var}(X|\vartheta)],$$

and therefore

$$k = E[\mathrm{var}(X_i|\vartheta)]/\mathrm{var}[E(X_i|\vartheta)].$$

REFERENCES

1. BÜHLMANN, H. *Mathematical Methods in Risk Theory*. New York: Springer-Verlag, 1970.
2. BÜHLMANN, H. "A Comparison of Three Credibility Formulae Using Multidimensional Techniques," *ASTIN Bulletin* VII (1973): 203–207.
3. BÜHLMANN, H., AND STRAUB, E. "Glaubwürdigkeit für Schadensätze," *Mitteilungen der Vereinigung Schweizer Versicherungsmathematiker* LXX (1970): 111–33. Reprinted as "Credibility for Loss Ratios" (Brooks, C.E., trans.), *Actuarial Research Clearing House* 1972.2.
4. GERBER, H.U. *An Introduction to Mathematical Risk Theory*, Huebner Foundation Monograph 8. Homewood, Ill.: Richard D. Irwin Inc., 1979.
5. GERBER, H.U., AND JONES, D.A. "Credibility Formulas of the Updating Type." In *Credibility: Theory and Applications*. New York: Academic Press, 1975; *TSA* XXVII (1975): 31–52.
6. HOEL, P.G., PORT, S.C., AND STONE, J.S. *Introduction to Statistical Theory*. Boston: Houghton Mifflin, 1971.
7. HOGG, R.V. AND CRAIG, A.T., *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan, 1978.
8. JEWELL, W.S. "Multi-dimensional Credibility," Research Report 73-7, Operations Research Center, University of California, Berkeley, 1973; *Actuarial Research Clearing House*, 1973.4.

KRZYSZTOF J. STROIŃSKI[*]:

Dr. Herzog has to be complimented for his contribution on credibility theory. His paper is a revised and abbreviated version of his CAS Study Note on Bayesian credibility. We are indebted to Dr. Herzog for condensing the growing discipline of credibility theory into a manageable note. His paper reviews mostly the earlier works on credibility. Because of the introductory character of the paper, the emphasis was put on basic statistical concepts. I would like to supplement Dr. Herzog's paper with some additional references.

In January 1987, a special note by Dr. Howard R. Waters, "An Introduction to Credibility Theory," was prepared and published as study material for British actuarial students. This note goes a step further than the discussed paper by also describing the Bühlmann-Straub model.

Later, in April 1987, *Credibility Theory* by Marc Goovaerts and Will J. Hoogstad appeared as No. 4 of the *Surveys of Actuarial Studies*. It reviews the classical model of Bühlmann, the Bühlmann-Straub model and seven additional models:

The Hachemeister Regression Model
The De Vylder Non-Linear Regression Model
The De Vylder Semi-Linear Model
The Opitmal Semi-Linear Model
The Hierarchical Model of Jewell
The De Vylder Loss Reserving Model
The Optimal Trimming Model of Gisler
This book contains chapters on special applications of credibility theory and a chapter on credibility for loaded premium. Furthermore, the book comes with an APL workspace that implements calculation of all nine models.

Note that the survey by Goovaerts and Hoogstad gives a different collateral premium estimator than Dr. Waters' note.

Finally, I would like to mention the special issue on credibility theory published by *Insurance: Abstracts and Reviews* in February 1986, which gives a complete list of references up to 1982. Moreover, it contains an excellent introduction to credibility theory written by Dr. Bjørn Sundt. This is one of the best guides through most of the credibility theory literature.

*Krzysztof J. Stroinski, not a member of Society, is an Assistant Professor in the Department of Statistical and Actuarial Sciences, University of Western Ontario.

BIBLIOGRAPHY

1. GOOVAERTS, M.J. AND HOOGSTAD, W.J. *Credibility Theory*, No. 4 of *Surveys of Actuarial Studies*. The Netherlands: Nationale-Nederelanden N.V., 1987.
2. Special Issue on Credibility Theory. *Insurance: Abstracts and Reviews* 2, no. 3 (February 1986): 233–326.
3. WATERS, H.R. *An Introduction to Credibility Theory*. London: Institute of Actuaries, 1987.

W.H. ODELL:

This paper has been sorely needed, and Dr. Herzog is to be congratulated.

For too many years, those aspiring to apply actuarial science to health insurance matters have not had the benefit of applicable material on one of the examination syllabuses. Only a month before the preliminary copies of this paper appeared, I had written James MacGinnitie, then President of the American Academy of Actuaries, asking that, to prepare actuaries taking the Society of Actuaries examinations for health insurance work, some arrangements be made for an educational curriculum that would include the concepts of relativity and credibility. That letter pointed out the material available on the syllabus for those training to apply actuarial science in the field of property and casualty insurance.

This paper garners that material, synthesizes it, and presents it as a subject for study unto itself. This is not only valuable from the point of view of imparting knowledge to those interested in this field, but also important because that knowledge is presented in a usable way. This knowledge is sorely needed. To prepare the next generation (and dare I say, the present one) of actuaries who follow(ed) the Society of Actuaries syllabus, this paper should be immediately incorporated therein to fill a void and help our profession better serve its publics.

We owe our thanks and congratulations to Dr. Herzog.

(AUTHOR'S REVIEW OF DISCUSSION)

THOMAS N. HERZOG:

I would like to thank Mr. Fuhrer for pointing out Professor Gerber's proof. I would also like to thank Professor Stroiński for supplying some recent references that extend the concepts discussed in my introductory paper. Finally, I am grateful to Mr. Odell for his suggestion that my work be placed on the health insurance syllabus.