Credibility for Hiawatha

by Oakley E. Van Slyke

Hiawatha Designs An Experiment

by Maurice G. Kendall

1. Hiawatha, mighty hunter
He could shoot ten arrows upwards
Shoot them with such strength and swiftness
That the last had left the bowstring
Ere the first to earth descended.
This was commonly regarded
As a feat of skill and cunning.

2. One or two sarcastic spirits
Pointed out to him, however,
That it might be much more useful
If he sometimes hit the target.
Why not shoot a little straighter
And employ a smaller sample?

3. Hiawatha, who at college
Majored in applied statistics
Consequently felt entitled
To instruct his fellow men on
Any subject whatsoever
Waxed exceedingly indignant
Talked about the law of error,
Talked about truncated normals,
Talked about loss of information,
Talked about his lack of bias
Pointed out that in the long run
Independent observations
Even though they missed the target
Had an average point of impact
Very near the spot he aimed at
(With the possible exception
Of a set of measure zero.)

4. This, they said, was rather doubtful.
Anyway, it didn't matter
What resulted in the long run;
Either he must hit the target
Much more often than at present
Or himself would have to pay for
All the arrows that he wasted.

5. Hiawatha, in a temper
Quoted parts of R.A. Fisher
Quoted Yates and quoted Finney
Quoted yards of Oscar Kempthorne
Quoted reams of Cox and Cochran
Quoted Anderson and Bancroft
Practically in extenso
Trying to impress upon them
That what really mattered
Was to estimate the error.

6. One or two of them admitted
Such a thing might have its uses
Still, they said, he might do better
If he shot a little straighter.

7. Hiawatha, to convince them,
Organized a shooting contest
Laid out in proper manner
Of designs experimental
Recommended in the textbooks
(Mainly used for tasting tea, but
Sometimes used in other cases)
Randomized his shooting order
In factoral arrangements
Used in the theory of Galois
Fields of ideal polynomials
Got a nicely balanced layout
And successfully confounded
Second-order interaction.

8. All the other tribal marksmen
Ignorant, benighted creatures,
Of experimental set-ups
Spent their time of preparation
Putting in a lot of practice
Merely shooting at a target.

9. Thus it happened in the contest
That the scores were most impressive
With one solitary exception.
This (I hate to have to say it)
Was the score of Hiawatha,
Who, as usual, shot his arrows
Shot them with great strength and swiftness
Managing to be unbiased
Not, however, with his salvo
Managing to hit the target.

10. There, they said to Hiawatha,
That is what we all expected.

11. Hiawatha, nothing daunted,
Called for pen and called for paper
Did analyses of variance
Finally produced the figures
Showing beyond preadventure
Everybody else was biased
And the variance components
Did not differ from each other
Or from Hiawatha's
(This last point, one should acknowledge
Might have been much more convincing
If he hadn't been compelled to
Estimate his own component
From experimental plots in
Which the values all were missing.
Still, they didn't understand it
So they couldn't raise objections.
This is what often happens
With analyses of variance.)

12. All the same, his fellow tribesmen
Ignorant, benighted heathens,
Took away his bow and arrows,
Said that though my Hiawatha
Was a brilliant statistician
He was useless as a bowman,
As for variance components
Several of the more outspoken
Made primeval observations
Hurtful to the finer feelings
Even of a statistician.

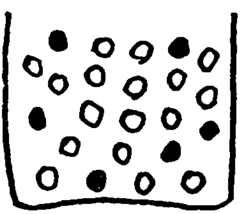
13. In a corner of the forest
Dwells alone my Hiawatha
Permanently cogitating
On the normal law of error
Wondering in idle moments
Whether an increased precision
Might perhaps be rather better
Even at the risk of bias
If thereby one, now and then, could
Register upon the target.

Introduction

This paper is called credibility for Hiawatha because it is about *expected value* ratemaking. Like Hiawatha, and unlike users of "classical" credibility, we are concerned today with making good estimates, that is, minimum variance unbiased estimates of the expected value of the outcome of a stochastic process. Our first point is that Bayesian credibility is *always* better than classical credibility if the goal is to estimate future loss costs.

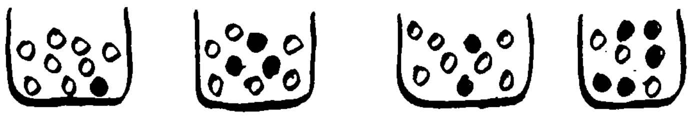
Nonetheless, like Hiawatha, we must consider that: "... an increased precision / Might perhaps be rather better / Even at the risk of bias / If thereby one, now and then, could / Register upon the target." Our second point is that there are tricks you can use to make Bayesian credibility computations easily. Each trick introduces a little bias, but the tricks improve the precision of your estimates as well as making them easy to calculate.

Definition of the Problem



Consider a big urn with an unknown number of red and white balls. Each red ball has a number on it (called a "loss amount").

A sprite has drawn balls at random from the big urn and put them into four small urns. The sprite may have put a greater proportion of red balls into some urns, and a lesser proportion into others.



The small urns correspond to various classes, various territories, various years, or *any other way* the universe of risks is divided into experience groups.

Problem 1:

Examine the entire contents of each of the four small urns.

Estimate $\overline{R}_i$, the expected rate of loss per draw, for each small urn.

	Urn 1	Urn 2	Urn 3	Urn 4
Number of balls N_i	10	20	30	40
Total losses L_i	10	40	90	160
$\overline{R}_i = \frac{L_i}{N_i}$	1	2	3	4

Problem 2:

Examine the entire contents of each of the four small urns. Estimate $\bar{R}_{\bullet}$, the expected rate of draw for the big urn.

		Big Urn
Number of balls	ΣN_i	100
Total losses	ΣL_i	300
$\bar{R}_{\bullet}=\frac{\Sigma L_i}{\Sigma N_i}$		3.0

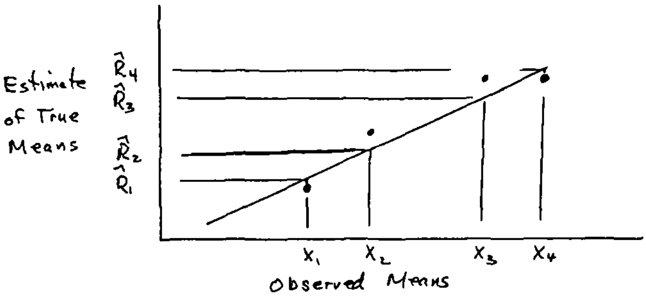
This is equivalent to $\frac{\Sigma \bar{R}_i N_i}{\Sigma N_i}$. That is, the various R_i's are weighted according to their number of balls (their sample sizes).

Problem 3:

Sample the contents of each small urn N_i times with replacement. Estimate $\bar{R}_i$, the expected rate of loss per draw from each small urn, as a linear combination of the observed rates of loss.

This is the Bayesian formulation of the insurance rating problem. It represents the insurance rating problem because the observed accidents are a random sample of the accidents that might have happened.

We use a linear combination of the observed rates of loss because this is the simplest way to reflect the information about the means of the various classes. Graphically, the process is as follows:



Historically, the general results have been in use since they were first published in 1914.

Hans Bühlmann showed that the best linear unbiased estimate of $\bar{R}_i$ is a weighted average of:

1. The observed average, X_i, with weight:

$$\frac{e_i}{var[S]}$$

(the reciprocal of the process variance for the urn)
where:
e_i = number of units of exposure in the observed average
$var[S]$ = variance of the claims process for one unit of exposure

2. The estimated grand average, $\hat{R}_{\bullet}$, with weight

$$\frac{1}{var[\mu]}$$

(the reciprocal of the variance of the distribution of means among the urns).

These weights correspond to the "number of balls" in Problem 2 because the number of balls is proportional to the reciprocal of the estimated variance of the estimate of $\hat{R}_i$.

Bühlmann showed this result in terms of credibility:

BAYESIAN ESTIMATES OF LOSS RATES

$$E[\overline{R}_i|X_i] = X_i\left[\frac{e_i}{e_i+K}\right]+\hat{R}_{\bullet}\left[1-\frac{e_i}{e_i+K}\right]$$

$$\hat{R}_{\bullet}=\frac{\sum \frac{e_i}{e_i+K}X_i}{\sum \frac{e_i}{e_i+K}}$$

The constant K is the expected value of the process variance for one draw divided by the variance of the means among the small urns.

The quantity $\frac{e_i}{e_i+K}$ is called the credibility of X_i. It is often denoted Z_i.

The expected loss rate for the average of all classes for Problem 3 is the expected loss rate for the big urn. The formula says that when we only observe a sample from each small urn, the best estimate of the loss rate for the big urn is a weighted average of the observed averages of the small urns, with weights equal to the credibility of each urn's average loss rate. The complement of the credibility goes to the credibility-weighted average of the observed average loss rates.

Bayesian, or expected-value, credibility says that K depends on the expected value of the process variance for one unit of exposure and the variance of unknown class means.

"Classical" credibility says that K is a function of the process variance, the choice of a tolerable percentage error, such as ± 5%, and the choice of a tolerable probability of unacceptable error.

Therefore classical credibility theory will only be correct when the percentage error and probability of error are *chosen* to yield the same credibility value as expected-value credibility. In all other cases, classical credibility theory will give the wrong credibility weight, if the objective is to estimate the expected loss rates.

Implicit Assumptions:

1. All classes have some process variance per unit of exposure. That is, all classes have measures of process variance and exposure, and process variance decreases as exposure increases.
2. The underlying mean for any particular class is a random variable from a certain process, and that process is applicable to all classes. (I.e., don't credibility-weight malpractice loss rates with homeowners loss rates.)

A Practical Understanding of *K*

1. The Three Components of *K*

The "credibility constant", K, is the number of times we must sample from a small urn to have enough experience to give the observed average, X_i, a credibility of 50%. That is, when

$$E[\overline{R}_i|X_i] = \frac{1}{2}X_i + \frac{1}{2}\hat{R}_{\bullet} \ ,$$

then:

$$\frac{e_i}{e_i+K} = \frac{1}{2}$$

and:

$$K = e_i$$

The purpose of this presentation is to show that K is the product of the average exposure per claim and two dimensionless quantities that reflect the predictability of claim sizes and the relevance of the grand mean to the prediction of individual means.

$$K = \frac{var[N]\,(E[Y])^2 + E[N]\,var[Y]}{var[\mu]}$$

$$= \frac{1}{E[N]} \cdot \frac{\dfrac{var[N]}{E[N]} + \dfrac{var[Y]}{(E[Y])^2}}{\dfrac{var[\mu]}{(E[N]\,E[Y])^2}}$$

$$= \frac{1}{E[N]} \cdot \frac{1+\beta+CV_y^2}{CV_\mu^2}$$

$$= Avg\ Exposure\ per\ Claim \cdot \frac{Dispersion\ of\ the\ Loss\ Process}{Dispersion\ of\ the\ Unknown\ Means}$$

For a Poisson frequency distribution:

$$K = \frac{1}{E[N]} \cdot \frac{1+CV_y^2}{CV_\mu^2}$$

The credibility of frequency relativities is:

$$K_N = \frac{1}{E[N]} \cdot \frac{1+\beta}{CV_\mu^2}$$

Notation:

$E[N]$, $var[N]$: Claim frequency process

$E[Y]$, $var[Y]$: Claim severity process

$E[N]E[Y]$ = Expected value of class means

$\frac{1}{E[N]}$ = Average exposure per claim

β = Dispersion of the claim frequencies among risks within the classes. (The coefficient of variation, or CV, of a probability distribution is the ratio of its standard deviation to its mean.)

CV_Y^2 = Dispersion of claim sizes

CV_μ' = Dispersion of mean loss rates

This result is exact only if frequency and severity are independent. In practice, a lack of independence is usually a negative correlation between frequency and severity and can be reflected by increasing K slightly.

2. Average Exposure per Claim

The starting point for determining K is the average exposure per claim. This could be:

Auto Insurance	60 car-years in a given class
Worker's Compensation	$120,000 of payroll in a given class.
Property, wind exposure, claims over $10 million	100 billion dollar-years of insured value.

The average exposure per claim is defined by the problem. It is easily determined from loss experience, and it is known with considerable accuracy, even if the expected claim frequency is small.

One quality of a good choice of exposure unit is that both expected loss costs and expected process variance increase in proportion to the number of units of exposure. Alternatively, the average loss rate is unaffected by the volume of exposure, and the variance of the observed loss rate decreases in proportion to the exposure.

3. Estimating the Dispersion of Claim Sizes

The next point is the determination of the dimensionless quality reflecting the volatility of the claims process. The dispersion of frequencies, β, is usually small and can usually be ignored. The quantity $(1 + CV_Y^2)$ can be computed from claim size data. For a group of claims valued at Y_i, $i = 1 \ldots n$, and n sufficiently large, this can be estimated from:

$$1+CV_y^2 \doteq \frac{\sum\left(\frac{Y_i}{\overline{Y}}\right)^2}{n}$$

$$= \frac{\sum Y_i^2}{n(\sum Y_i)^2}$$

Another estimate, more stable for claim size distributions that are highly skewed (including but not limited to the lognormal) is e^{σ^2}, where σ^2 is the variance of the logs of the claim sizes. The value of $(1 + CV_Y^2)$ may be any number greater than 1, but it is usually between 5 (for claims that are not widely dispersed) and 35 (for claims that are very widely dispersed).

The following table shows values from my experience. It also shows the effect on K of truncating various claim sizes. Truncating really unusual claims sizes reduces $(1 + CV_Y^2)$ and K, but truncating more common values, such as worker's compensation claims between $25,000 and $100,000, has little effect.

Examples of Dispersion of Claim Sizes

Line of Business	Approximate Value $(1 + CV_Y^2)$
California School District	
Worker's Compensation	
Unlimited	40
Limited to $100,000	15
Limited to $25,000	10
Private Passenger Automobile Collision	3
Commercial Truck Liability	
Limited to $250,000	15
Limited to $600,000	25
Limited to $1,000,000	35
Hospital Professional Liability	
Unlimited	45
California Municipal Liability	
24 mm. Excess of 1 mm.	5
Physician Medical Malpractice	
Limited to $250,000	3
Limited to $2,500,000	10
Automobile Products Liability	
Unlimited	80

4. Estimating CV_μ^2

Estimating CV_μ^2 from Data

One way to estimate CV_μ^2 is to use the data from the problem at hand. If there are enough "urns" (classes, regions, etc.), and a sufficient number of them have a credible volume of experience, then a value of CV_μ^2 can be found by trial and error which gives estimates of z_i, $\hat{R}_\bullet$, and

$$CV_\mu^2 = \frac{1}{\sum Z_i^2}\left[\sum Z_i^2\left(\frac{\overline{X}_i-\hat{R}_\bullet}{\hat{R}_\bullet}\right)^2 - \frac{(1+\beta+CV_Y^2)}{e_i E[N]}\right]$$

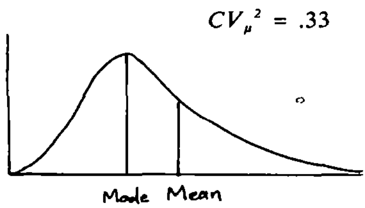
When these simultaneous equations are solved, the credibilities are underestimated because of the dependence upon data for estimates of unkown intermediate quantities. For k classes, the unbiased estimates of the z_i are:

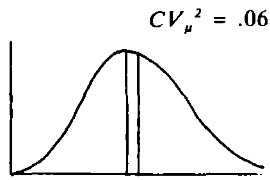
$$Z_i \leftarrow 1-\frac{k-3}{k}(1-Z_i)$$

Often in practice the X_i are by chance close to $\hat{R}_\bullet$ and this formula gives an unreliable estimate of CV_μ^2. This is particularly a problem if there are few classes or the classes have low credibility.

Estimating CV_μ^2 from the Relationship of Mean to Mode

In most cases of actuarial interest, the various class means must be greater than zero. As a result, the mean class mean, or grand average, is greater than the modal class mean, or most common class average. The greater the variance of the distribution, the greater the ratio of the mean to the mode.





Although your understanding of the classification process and the resulting means might be sketchy, you might be able to make your estimate of the CV_μ^2 consistent with your understanding of the extent by which the mean class mean exceeds the mode.

For example, for a gamma distribution of unknown class means, the results are as follows. A gamma is a reasonable choice because of its genesis as mixture of exponentials.

Gamma Distribution
of Unknown Means

Ratio of Mean to Mode	CV_μ^2	$\frac{1}{CV_\mu^2}$
2.00	.50	2
1.50	.33	3
1.25	.2	5
1.11	.1	10
1.06	.06	18
1.03	.03	34

The reciprocal of CV_μ^2 is usually between 2 and 25. A tabulation of data that divides good risks from bad will lead to a high ratio of the mean to the mode and a CV_μ^2 of as much as 0.50. A tabulation that does not meaningfully distinguish one group from another (such as tabulation by accident year of on-level premiums and losses) will lead to a low ratio of mean to mode and a low CV_μ^2 of 0.03 or less.

5. Introduction of Bias

These estimates of loss rates are biased (in statistical terms) because they rely on outside data. This is unimportant. In practice, the gain in accuracy more than makes up for the bias that is introduced. Like Hiawatha's tribesmen, we are introducing some bias in order to hit the target more often. Even more important, we are aiming our arrows at the target of expected-value estimation.

Credibility of Claim-Free Experience

A simple example of the usefulness of Bayesian credibility is the calculation of the credit for claim-free experience for a particular risk. One such credit is offered by reinsurers whose risks present seven years of claim-free experience. Another such credit is offered by auto insurers who give lower rates to claim-free drivers.

$$E[loss\ rate, 0\ claims] = \hat{R}_\bullet\left(1 - \frac{e_i}{e_i + K}\right)$$

where $\hat{R}_\bullet$ = a priori estimate of expected rate

= 1.0, for determining a credit for claim-free experience.

Charge = $\left(1-\frac{e_i}{e_i+K}\right)$

Credit = $\frac{e_i}{e_i+K}$

We define our exposure period to be 1.0 units. Then:

Credit = $\frac{1}{1+K}$

$$= \frac{E[N]\cdot CV_\mu^2}{1+\beta+E[N]\cdot CV_\mu^2}$$

Where:

$E[N]$ = The expected number of claims in the exposure period.

$CV_\mu{}^2$ = The dispersion of means of claim frequencies from risk to risk.

Examples

The credit is:

	ß:	0		0.3	
$E[N]$	$CV_\mu{}^2$:	.05	.50	.05	.50
0.10		.0050	.048	.0038	.037
1.00		.048	.33	.037	.28
10.0		.33	.83	.28	.79

If the risks are believed to be different from one another the credit is more than if the risks are believed to be similar to one another. The greater the number of expected claims, the greater the credit for claim-free experience.