SIMPLIFIED CREDIBILITY MATHEMATICS

BY

Joseph L. Tupper, III
CUNA Mutual Insurance Society
Madison, Wisconsin

The conventional formulas for estimating risk means and variances are derived using techniques no more advanced than in Part 2 of the Society of Actuaries syllabus. The main tool is linear regression and a slightly simpler formula for variance credibility is given.

Credibility is often described in terms of how much observations can be trusted. In insurance, we really want to know a true (but unknowable) value for a given parameter (such as claims) for one risk in a class of risks. Sometimes we have experience data for the risk to help us. The essence of credibility theory is to use the experience data to make a "best estimate" of the true risk parameter.

Since the term "best estimate" is not clearly defined, credibility practitioners have all appealed to fitting "what we want" and "what we have" to a preassigned model. While it has not been stated in exactly this fashion anywhere I have looked, all credibility really does is construct the least squares line of best fit relating the "true risk parameter given the experience" and the calculated observed value based on risk experience". In other words, credibility theory is an application of linear regression. Moreover, "the credibility" of a risk is just the slope of this regression line.

The development is handled in four sections as follows:

1. A review of linear regression;
2. A review of results on conditional expectations;
3. Credibility in general (estimation of an unknown parameter);
4. Estimation of Means and Variances.

Sections 1 and 3 are quite straightforward once Section 2 is grasped. There is some complexity in Section 2 which seems to be unavoidable. The exposition attempts to warn the reader before it gets out of hand as it may be preferable to look at Sections 3 and 4 before spending too much time in Section 2.

Section 3 does have one involved calculation set forth in full gory detail. This has been included because those of us who have not had much prior experience with conditional distributions seem to get lost somewhere in the middle.

"And now a word from our sponsor..." (M.Spivak, Differential Geometry)

This paper was written to enable the practicing actuary to wade through the somewhat opaque mathematics behind credibility theory as painlessly as possible. The only real sacrifice has been to use probability density functions rather than measures. Densities permit writing a mean as " $\int x f(x) dx$ " where a measure theoretic approach would write " $\int x\, dF(x)$ ", and the usual relationship between f and F occurs. There is no such density for a coin toss (fair or unfair); but there is a measure. Since claims processes involve discrete as well as continuous distributions, measure theory has played a significant role in simplifying the notation of theoretical works and in providing stronger results more reflective of real world processes.

Credibility theory is no exception. The beauty of measure theory is that with what seems to be a slight notational change, essentially from " $f(x)dx$ " to " $dF(x)$ ", every result in this paper is true in a measure theoretic context. The expression of Bayes' Theorem is the only casualty. (Bayes' Theorem still works without a hitch, but the notation I've used would be inconsistent with a "measured" approach.) There is a great deal more to measure theory than a "notational change." In fact, careful readers of this paper will object to the implicit assumption in section 3 that the risks A have a continuous distribution. That implicit assumption is necessary unless we require all readers to be familiar with measures.

Hans Bühlmann is credited with developing the notion of linking a linear approximation (i.e., a credibility estimate) with a least squares estimate. He has presented a excellent development in his risk theory text [1]. Several actuaries have faced difficulties with this text due to the level of mathematics employed: real analysis is not part of the SOA syllabus and Bühlmann's text is very difficult to appreciate without an understanding of real analysis. This paper recreates some of his results with mathematical tools of perhaps a more popular appeal. In addition, the form of the credibility estimate for $E[\sigma^2(A) \mid S_1, \dots, S_n]$ in Section 4 is simpler than in Bühlmann [1, pg. 105].

I believe that understanding the results of credibility theory does not require real analysis. However, applying credibility relies on using results that depend on the stronger definition of integration found with measure theory. For those readers interested in learning more on measure theory, the texts by Royden[2] and Rudin[3] are widely used. (Roydan is somewhat simpler; Rudin is more general).

1. A review of linear regression.

For linear regression, we have observations of pairs (x,y) with y viewed as depending on x. To estimate y as a linear function of x, written $y \sim ax+b$ we have the problem of choosing "best" values of a and b. As is well known, the method of least squares produces values of a and b which minimize the expression

$$\sum_{i=1}^{n} [y_i - ax_i - b]^2$$

where the pairs (x_i, y_i), for i=1 to n, are the observations.

Finding a and b from this statement is a fairly simple minimum problem from calculus. One standard presentation of the final equation for the fitted line is as follows:

$$y - E[Y] = \frac{Cov[X,Y]}{Var[X]}(x - E[X])$$

(Notation: $E[\cdot]$ = expected value, $Var[\cdot]$ = variance, $Cov[\cdot,\cdot]$ = covariance)

For credibility, we will need the "continuous case" presented below. Since the result is the same, this can be skipped.

If the independent variable, x, is distributed with density f(x), then y in (x,y) has a distribution conditioned on x, g(y|x) and (x,y) has joint distribution. h(x,y)=g(y|x)f(x).

The regression problem now is to find a and b to minimize:

$$S(a,b) = \iint [y - ax - b]^2 g(y|x) f(x)\, dy\, dx$$

by calculating partial derivatives

$$\frac{\partial S}{\partial a} = -2E[XY] + 2a\,E[X^2] + 2b\,E[X]$$

and

$$\frac{\partial S}{\partial b} = -2\,E[Y] + 2a\,E[X] + 2b$$

Setting the partials equal to zero and solving, we find

$$a = \frac{E[XY] - E[X]E[Y]}{E[X^2] - E[X]^2} = \frac{Cov[X,Y]}{Var[X]}$$

and

$$b = E[Y] - a\,E[X]$$

so,

$$y = a\,x + E[Y] - a\,E[X]$$

regrouping,

$$y - E[Y] = a\,(x - E[X])$$

and

$$y - E[Y] = \frac{Cov[X,Y]}{Var[X]}\,(x - E[X])$$

as before.

2. A review of results on conditional expectations.

We will need three things about conditional expectation and densities in Section 3. Two are computational rules and the third is Bayes' theorem (otherwise known as changing the order of integration).

A conditional expectation arises whenever a set of individuals divides naturally into subsets. An example is all claims incurred under group contracts. These divide naturally into subsets of claims incurred within each insured group.

The rules are the following two equations:

Rule 1. $E[S] = E[E[S|A]]$

Rule 2. $Var[S] = E[Var[S|A]] + Var[E[S|A]]$

Before we can prove these rules, we will need to set the stage for Bayes' theorem. This will show that the first rule means: if we take all observations and divide them into categories A_1, A_2, etc., then over each category A_i, we have an average $E[S|A_i]$. $E[S|A_i]$ is to be thought of as a function of the risk A_i alone. Also, each category A_i has a weight and the weighted average of the $E[S|A_i]$ is just the mean overall observations or $E[S]$. An example to keep in mind is average claim size. Then S is claim amounts and E[S] is the average claim size. $E[S|A_i]$ is the average claim size for risk A_i.

S is something we are observing.
S has a density f(s), and

$$E[S] = \int s f(s)\, ds$$

Somehow, the observations of S are divided into subsets A (we are not concerned with the method).

For each category, A, there is a density of S written $f(s|A)$. Also there is a probability distribution $u(A)$ which links the subsets together: $u(A)$ is the probability that a random observation of S belongs to the subset A. The joint density satisfies $g(s, A) = f(s|A)u(A)$.

These notions are connected as follows:

$$f(s) = \int f(s|A)\, u(A)\, dA = \int g(s, A)\, dA$$

and

$$E[S|A] = \int s f(s|A)\, ds$$

We prove Rule 1 by direct calculation:

$$\begin{aligned}
E[E[S|A]] &= \int E[S|A]\, u(A)\, dA = \int \left[\int s f(s|A)\, ds\right] u(A)\, dA \\
&= \iint s f(s|A)\, u(A)\, ds\, dA = \iint s\, g(s, A)\, ds\, dA \\
&= \int s \left[\int g(s, A)\, dA\right] ds = \int s f(s)\, ds \\
&= E[S]
\end{aligned}$$

Rule 2 is also a calculation:

$$\begin{aligned}
Var[S] &= E[S^2] - E[S]^2 \\
&= E[S^2] - E[E[S|A]^2] + E[E[S|A]^2] - E[S]^2 \\
&= E[E[S^2|A]] - E[E[S|A]^2] + E[E[S|A]^2] - E[E[S|A]]^2 \\
&= E[E[S^2|A] - E[S|A]^2] + E[E[S|A]^2] - E[E[S|A]]^2 \\
&= E[Var[S|A]] + Var[E[S|A]]
\end{aligned}$$

While this algebra may seem a bit dense, line 2 introduces a convenient representation of zero, line 3 uses rule 1 on the first and last terms, line 4 just aggregates the first and second terms which are both expectations over A's, and line 5 is just the definition of variance (twice).

Rules 1 and 2 were derived using the one half of Bayes' theorem:

$$g(s,A) = f(s|A)\,u(A)$$

The other half of Bayes' theorem is the part we will need in Section 3. We need a new probability distribution for A given a value for S, denoted $u(A|s)$. This satisfies:

$$g(s,A) = u(A|s)\,f(s)$$

We will use:

$$u(A|s)\,f(s) = f(s|A)\,u(A)$$

and its generalization:

$$u(A|s_1, \ldots, s_n)\,f(s_1)\,f(s_2)\cdots f(s_n) = f(s_1|A)\,f(s_2|A)\cdots f(s_n|A)\,u(A)$$

3. Credibility in general (estimation of an unknown parameter).

To begin the general analysis, let us take inventory. We have available the following:

- A collection of risks, A, and a probability density function, u(A), reflecting the weight assigned to each risk.

- A set of observations $S_1, \ldots, S_n$ of some parameter S (e.g., average claim size or variance of claim size). These observations are attached to a particular risk. The parameter S has a density $f(s)$ over all risks or $f(s|A)$ if restricted to the risk A.

- An unbiased estimator of the parameter S: $\bar{S} = \bar{S}(S_1, \ldots, S_n)$ (i.e., $\bar{S}$ is a function of $S_1, \ldots, S_n$).

In practice, u(A) is the weight to assign each risk so that rule 1 of Section 2 will work for observations of the parameter 3.

We want a best estimate of the "true mean" of S at a risk, A. That is, we want $E[S|A]$ denoted $\mu(A)$. We will also use $\sigma^2(A) = Var[S|A]$. Since $\mu(A)$ is unavailable, the next best estimate answers the question, "What do we expect $\mu(A)$ to be given $S_1, \ldots, S_n$?" Translated to symbols, this is $E[\mu(A) | S_1, \ldots, S_n]$.

In the real world, f(s) and u(A) will not really be available, but experience data for the collection of risks can be used to approximate what we need.

$\bar{S}$ must be unbiased. In particular, $\bar{S}$ must provide an unbiased estimator of $\mu(A)$. That is, $E[\bar{S}|A] = \mu(A)$. Since we will be most concerned with means and variances, it is enough to know that such unbiased estimators exist.

Look again: We have $\bar{S}$, we want $E[\mu(A)|S_1, \ldots, S_n]$. Let us try linear regression with $x = \bar{S}$ and $y = E[\mu(A)|S_1, \ldots, S_n]$.

From Section 1, we can write the equation of the regression line immediately:

$$y - E\left[E[\mu(A)|S_1, \ldots, S_n]\right] = \frac{Cov[\bar{S}, E[\mu(A)|S_1, \ldots, S_n]]}{Var[\bar{S}]} \left(x - E[\bar{S}]\right)$$

The rest of this section transforms this to

$$E[\mu(A) | S_1, \ldots, S_n] \sim z \cdot \bar{S} + (1 - z)\ E[\mu(A)]$$

with

$$z = \frac{Var[\mu(A)]}{Var[\mu(A)] + E[Var[\bar{S}|A]]}$$

Section 4 will analyze z further when estimating means and variances.

First let us transform the regression line equation to

$$E[\mu(A)|S_1,\ldots,S_n] \sim z\cdot\bar{S} + (1-z)\cdot E[\mu(A)]$$

Since the expectations are over all possible $S_1, \cdots, S_n$,

(1) $E[\bar{S}] = E[\mu(A)]$ by the assumption that $\bar{S}$ is an unbiased estimator of $\mu(A)$. $\left(E[S] = E[E[S|A]] = E[\mu(A)] \right)$

(2) $E\left[E[\mu(A)|S_1,\ldots,S_n]\right] = E[\mu(A)]$ by Rule 1 of Section 2.

The convention is to write $E[\mu(A)]$ rather than use $\mu = E[\mu(A)]$ .This highlights the importance of the risks.

Thus our regression line equation is

$$E[\mu(A)|S_1,\cdots,S_n] - E[\mu(A)] \sim z\left(\bar{S} - E[\mu(A)]\right)$$

with

$$z = \frac{Cov[\bar{S}, E[\mu(A)|S_1,\ldots,S_n]]}{Var[\bar{S}]}$$

which transforms into the familiar $E[\mu(A)|S_1,\cdots,S_n] \sim z\cdot\bar{S} + (1-z)\cdot E[\mu(A)]$

We still need to analyze z.

Using Rule 2 of Section 2, expand the denominator of z :

$$Var[\bar{S}] = E[Var[\bar{S}|A]] + Var[E[\bar{S}|A]] = E[Var[\bar{S}|A]] + Var[\mu(A)]$$

The numerator of z must be handled explicitly. This is where the comments on Bayes' theorem apply. The important idea is that we are really just switching the order of integration and shuffling symbols.

Now,

$$Cov[\bar{S}, E[\mu(A)|S_1,\ldots,S_n]] = E[\bar{S}\cdot E[\mu(A)|S_1,\ldots,S_n]] - E[\bar{S}]\cdot E[E[\mu(A)|S_1,\ldots,S_n]] \quad (1)$$

$$= \int\cdots\int \bar{S}\, E[\mu(A)|S_1,\ldots,S_n]\, f(S_1)\cdots f(S_n)\, dS_1\cdots dS_n - E[\mu(A)]\cdot E[\mu(A)] \quad (2)$$

$$= \int\cdots\int \bar{S}\left[\int \mu(A)\, u(A|S_1,\ldots,S_n)\, dA\right] f(S_1)\cdots f(S_n)\, dS_1\cdots dS_n - E[\mu(A)]^2 \quad (3)$$

$$= \int\cdots\iint \bar{S}\mu(A)\left[u(A|S_1,\ldots,S_n) f(S_1)\cdots f(S_n)\right] dA\, dS_1\cdots dS_n - E[\mu(A)]^2 \quad (4)$$

[Note use of Bayes' Theorem]

$$= \iint\cdots\int \bar{S}\mu(A)\left[f(S_1|A)\cdots f(S_n|A)\, u(A)\right] dS_1\cdots dS_n\, dA - E[\mu(A)]^2 \quad (5)$$

$$= \int \mu(A)\, E[\bar{S}|A]\, u(A)\, dA - E[\mu(A)]^2 \quad (6)$$

$$= \int \mu(A)^2\, u(A)\, dA - E[\mu(A)]^2 \quad (7)$$

$$= E[\mu(A)^2] - E[\mu(A)]^2 \quad (8)$$

$$= Var[\mu(A)] \quad (9)$$

The step from line 6 to line 7 again makes use of $\bar{S}$ as an unbiased estimator of $\mu(A)$. The rest is definitions and Bayes' theorem.

Thus, $$Z = \frac{Cov[\bar{S}, E[\mu(A)|S_1,\ldots,S_n]]}{Var[\bar{S}]} = \frac{Var[\mu(A)]}{Var[\bar{S}]}$$

$$= \frac{Var[\mu(A)]}{Var[\mu(A)] + E[Var[\bar{S}|A]]}$$

This concludes the general developement. The rest of credibility theory has to do with particular choices of estimators $\bar{S}=\bar{S}(S_1,\ldots,S_n)$ of the parameters we wish to study.

4. Estimation of Means and Variances

As stated at the end of Section 3, the main trick in credibility theory is finding an unbiased estimator for the parameter under observation. In this section the credibility levels, z, for means and variances will be discussed.

In rules 1 and 2 from section there are three terms needed to calculate total means and variances. Two of the three have a reasonable interpretation for the risk: $E[\mu(A)|s_1,\ldots,s_n]$ and $E[\sigma^2(A)|s_1,\ldots,s_n]$. Comparing these with rules 1 and 2, a third item is suggested: $Var[\mu(A)|s_1,\ldots,s_n]$. Bühlmann [1,pg. 99] relates this component to the "fluctuation part" of his credibility premium concept.

For means, our unbiased estimator is $\bar{S} = \frac{1}{n}\sum_{i=1}^{n} s_i$

$$E[\bar{S}|A] = \mu(A)$$

$$Var[\bar{S}|A] = \frac{1}{n}\sigma^2(A)$$

Now

$$z = \frac{Var[\mu(A)]}{Var[\mu(A)] + E[Var[\bar{S}|A]]} = \frac{Var[\mu(A)]}{Var[\mu(A)] + E[\frac{1}{n}\sigma^2(A)]}$$

$$= \frac{Var[\mu(A)]}{Var[\mu(A)] + \frac{1}{n}E[\sigma^2(A)]}$$

Rearranging,

$$z = \frac{n}{n + \frac{E[\sigma^2(A)]}{Var[\mu(A)]}}$$

Writing

$$K = \frac{E[\sigma^2(A)]}{Var[\mu(A)]}$$

We have

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

For variances, we are first looking for a credible estimate of $\sigma^2(A)$ given $s_1,\ldots,s_n$. To keep the notation straight, $\mu(A)$ and $\bar{S}$ of section 3 are replaced by the analogous symbols for variance:

$\bar{S}$ becomes $\Sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}[s_i - \bar{s}]^2$ and $\mu(A)$ becomes $\sigma^2(A)$

Σ^2 is our unbiased estimator of $\sigma^2(A)$. $E[\Sigma^2|A] = \sigma^2(A)$

This time $E[\sigma^2(A)|s_1,\ldots,s_n] \sim z\cdot\Sigma^2 + (1-z)\,E[\sigma^2(A)]$

with

$$z = \frac{Var[\sigma^2(A)]}{Var[\sigma^2(A)] + E[Var[\Sigma^2|A]]}$$

Calculating $Var[\Sigma^2|A]$ is an exercise for the reader. The formula below is presented in Bühlmann [1, pg. 104] without complete proof. Since my derivation of it takes several pages and sheds little extra light, it is not included. The following is an explanation of the results.

Let $\gamma_2(A) = \frac{E[[S-\mu(A)]^4|A]}{\sigma^4(A)} - 3$ This is called the "excess." (Excess kurtosis over a normal distribution.)

Then $Var[\Sigma^2|A] = \left(\frac{n\sigma^2(A)}{n-1}\right)^2 \left[\frac{\gamma_2(A)+2}{n} - \frac{2(\gamma_2(A)+1)}{n^2} + \frac{\gamma_2(A)}{n^3}\right]$

This can be rewritten $Var[\Sigma^2|A] = \frac{\sigma^4(A)\gamma_2(A)}{n} + \frac{2\sigma^4(A)}{n-1}$

Substituting in Z we have

$$Z = \frac{Var[\sigma^2(A)]}{Var[\sigma^2(A)] + E[Var[\Sigma^2|A]]}$$

$$Z = \frac{Var[\sigma^2(A)]}{Var[\sigma^2(A)] + E\left[\frac{\sigma^4(A)\gamma_2(A)}{n} + \frac{2\sigma^4(A)}{n-1}\right]}$$

Rearranging:

$$Z = \frac{n(n-1)}{n(n-1) + (n-1)K_1 + nK_2}$$

Where

$$K_1 = \frac{E[\sigma^4(A)\gamma_2(A)]}{Var[\sigma^2(A)]}$$

and

$$K_2 = \frac{2E[\sigma^4(A)]}{Var[\sigma^2(A)]}$$

The numbers K, K_1 and K_2 I have taken to calling "50%" credibility levels.

For $Var[\mu(A)|S_1,\ldots,S_n]$ there is no obvious candidate for an unbiased estimator. However, if there is no way to differentiate between risks we always have:

$$Var[\mu(A)|S_1,\ldots,S_n] \sim E[Var[\mu(A)|S_1,\ldots,S_n]]$$

Moreover, as Bühlmann observes, we have an estimate of $\mu(A)$ given $S_1,\ldots,S_n$. Following this through, as on the next page, we find

$$Var[\mu(A)|S_1,\ldots,S_n] \sim \left(1 - \frac{n}{n+K}\right) Var[\mu(A)]$$

with K defined as for the mean.

By definition $Var[\mu(A)|s_1,\dots,s_n] = E\{[\mu(A) - E[\mu(A)|s_1,\dots,s_n]]^2 | s_1,\dots,s_n\}$

$$\sim E\{[\mu(A) - z\bar{s} - (1-z)E[\mu(A)]]^2 \mid s_1,\dots,s_n\}$$

Then

$$E[Var[\mu(A)|s_1,\dots,s_n]] \sim E[E\{[\mu(A) - z\bar{s} - (1-z)E[\mu(A)]]^2 | s_1,\dots,s_n\}]$$

$$= E[\{\mu(A) - z\bar{s} - (1-z)E[\mu(A)]\}^2]$$

$$= E[\{[\mu(A) - E[\mu(A)]] - z[\bar{s} - E[\mu(A)]]\}^2]$$

$$= E[\{\mu(A) - E[\mu(A)]\}^2] - 2z\,E[\{\mu(A) - E[\mu(A)]\}\{\bar{s} - E[\mu(A)]\}] + z^2 \cdot E[\{\bar{s} - E[\mu(A)]\}^2]$$

$$= Var[\mu(A)] - 2z\,E[E[\{\mu(A) - E[\mu(A)]\}\{\bar{s} - E[\mu(A)]\} \mid A]] + z^2\,Var[\bar{s}]$$

$$= Var[\mu(A)] - 2z\,E[\{\mu(A) - E[\mu(A)]\}\{\mu(A) - E[\mu(A)]\}] + z^2\,Var[\bar{s}]$$

$$= Var[\mu(A)] - 2z\,E[\{\mu(A) - E[\mu(A)]\}^2] + z^2\,Var[\bar{s}]$$

$$= Var[\mu(A)] - 2z\,Var[\mu(A)] + z^2\,Var[\bar{s}]$$

$$= Var[\mu(A)](1-2z) + \left(\frac{Var[\mu(A)]}{Var[\bar{s}]}\right)^2 \cdot Var[\bar{s}]$$

$$= Var[\mu(A)](1-2z) + \left(\frac{Var[\mu(A)]}{Var[\bar{s}]}\right) Var[\mu(A)]$$

$$= Var[\mu(A)](1-2z) + z \cdot Var[\mu(A)]$$

$$= (1-z)\,Var[\mu(A)]$$

$$= \left(1 - \frac{n}{n+K}\right) Var[\mu(A)]$$

Summary and Applications

The building blocks for credibility estimates are:

$$E[\mu(A)]$$
$$Var[\mu(A)]$$
$$E[\sigma^2(A)]$$
$$E[\sigma^4(A)]$$
$$E[\sigma^4(A)\,\gamma_2(A)]$$

(and $Var[\sigma^2(A)] = E[\sigma^4(A)] - E[\sigma^2(A)]^2$)

These five numbers must be estimated from historical data. To do this, there are many different approaches, but two very different approaches should at least be considered. The first is to assume particular distributions (Poisson, normal, binomial, etc.) and use data to estimate a fit. Bühlmann presents an example of this [1, pp.106-109]. The second approach is to use the experience data to directly estimate the building blocks, at a risk level, and then to use the risks which are "more credible" [higher u(A) is close enough] as an approximation to the true distribution. This bypasses the need to construct (numerically) the explicit distributions.

References

[1] Bühlmann, Hans. Mathematical Methods in Risk Theory, Springer-Verlag, 1970. Chapter 4, pp. 85-106.

[2] Royden, H.L. Real Analysis, MacMillan Company, 1968.

[3] Rudin, Walter. Real and Complex Analysis, McGraw-Hill, 1966.