Credibility Theory for Dummies

Gary G Venter

Guy Carpenter Instrat

Least squares credibility is usually derived from some fairly complicated looking assumptions about risk across a collective. It turns out, however, that the basic results can be developed from some standard statistical operations with weighted regression. This is outlined, and some more advanced models are tied to the same approach, in this note.

Credibility Theory for Dummies

Credibility theory is usually presented as a mathematically dense body of formulas. Here is something a little different: a short, simple approach. "Dummies" is of course a relative term. Algebra, differential calculus, and some background in statistics are all assumed.

What is credibility?

Credibility theory is all about weighted averages. Different estimates of a quantity are to be weighted together. The more credible estimates get more weight.

In the context of estimating expected losses for a member of a class, there are two natural estimates: the experience of the member itself, and the average of the entire class. The former is more relevant but also more volatile than the latter. Two general approaches have been taken to calculating weights in this case. The limited fluctuation approach is willing to accept the member experience at face value if it meets a pre-defined standard of stability (full credibility) and if not reduces the weight enough for the weighted average to meet the stability requirement. The greatest accuracy approach measures relevance as well as stability and looks for the weights that will minimize an error measure. The average of the entire class could be a very stable quantity, but if the members of the class tend to be quite different from each other, it could be of less relevance for any particular class. So the relevance of a wider class average to any member's mean is inversely related to the variability among the members of the class.

The error measure used in the greatest accuracy approach is almost always expected squared error, so this method is often called "least squares credibility." In Europe it is sometimes called "classical credibility." The limited fluctuation approach is called classical in North America. Thus "classical" is a term worth avoiding, not only because of its geographic ambiguity, but also because it is a historical rather than a methodological description.

Least squares credibility

Suppose you have two independent estimates x and y of a quantity, with respective expected squared errors u and v. Take a weighted average $a = zx + (1–z)y$. The expected squared error of

a is $w = z^2u + (1–z)^2v$. What z minimizes w? Here is where the calculus comes in. The derivative dw/dz is $2zu + 2(z–1)v$. If you set that to zero you get: $zu+zv = v$, or $z = v/(u+v)$. Then $1–z$ is $u/(u+v)$. This makes it look like each estimate gets a weight proportional to the expected squared error of the other. To express the weights as properties of the estimates themselves, note that $(1/u)/[1/u + 1/v] = 1/[1+u/v] = v/(u+v) = z$. This shows that each estimate gets a weight proportional to the reciprocal of its expected squared error [1]. Least squares credibility is an application of this principle.

As an example, consider a class of risks. Suppose the losses L_{ij} in year j for the ith member of the class are randomly distributed as follows:

$$L_{ij} = C + M_i + \varepsilon_{ij} \qquad (1)$$

where C is the class mean loss, $C+ M_i$ is the mean loss for the ith member, and ε_{ij} is the random component for the jth period for this member. It is not much of a restriction to assume that the M_i's average to zero as do the ε_{ij}'s. Suppose the variance of the M_i's is t^2 and the variance of the random components ε_{ij} all are s_i^2. Denote their average $E(s_i^2)$ by s^2.

Sometimes t^2 is called the variance of the hypothetical means and s^2 the expected process variance. "Hypothetical" refers to the fact that the means $C + M_i$ are not observed.

With this setup, consider two estimates of member i mean losses: x, the average losses of the member for n periods, and y, the class mean loss C, which for now we will assume to know or at least be able to estimate well enough to ignore the error. To apply the inverse variance weightings, we needed to know the expected squared errors of x and y from the true value of $C + M_i$. By the definitions, y's expected squared error is just t^2. The expected squared error of x is the expected value of its variance s_i^2/n, i.e., s^2/n. Then applying the inverse expected squared error principle gives a weight to x of $z = (n/s^2)/[n/ s^2 + 1/t^2] = n/[n + s^2/t^2]$. This is the original Bühlmann credibility formula.

---

[1] This assumes the expected squared error is minimized rather than maximized at this z. The second derivative of w is $2u + 2v$ which is positive, so this assumption is valid.

The above would be an appropriate set of assumptions for a class where all members had roughly the same exposure, such as single cars. If the exposure varies much across members, like in territory ratemaking or commercial experience rating, the variances of the random components could not reasonably be assumed to be constant over time. To address this case, introduce an exposure measure P_{ij} for the ith member in period j, and assume that the variance of its random loss component is $P_{ij}s_i^2$, so each unit of exposure has a variance of s_i^2. In this case it would not be right to assume that M_i has mean zero, in that different members of the class would depart from the class mean loss in differing amounts depending on exposure. However, if in equation (1) L is reinterpreted as losses per unit of exposure, i.e., pure premium, this assumption could be reasonable. In that case, the variance of ε_{ij} would be s_i^2/P_{ij}. So here, x is the average loss per exposure for the ith member for n periods, and y is the mean pure premium for the class.

Thus the expected squared error of y from $C + M_i$ would still be t^2. Assume further that x is calculated as the sum of the n period losses divided by the sum of the exposures. Use a "~" in a subscript to denote summation, so the total exposures for the ith class over the n periods is $P_{i\sim}$. Then the variance of x is just $P_{i\sim}s_i^2/P_{i\sim}^{\ 2} = s_i^2/P_{i\sim}$, with expected value $s^2/P_{i\sim}$. So what is the credibility of the pure premium? The inverse expected squared error weighting gives $z_i = (P_{i\sim}/s^2)/[\ P_{i\sim}/s^2 + t^{-2}] = P_{i\sim}/[\ P_{i\sim} + s^2/t^2]$. This is often expressed more simply as $z = P/[P+K]$, which is the Bühlmann-Straub credibility formula.

C can be estimated by a weighted average of the x's, the member means. The expected squared error of x from C is $t^2 + s^2/P_{i\sim}$, so x should get a weight inversely proportional to that, so proportional to $t^{-2}P_{i\sim}/[\ P_{i\sim} + s^2/t^2]$, which is proportional to z_i . Thus C can be estimated as a weighted average of the x's where the weights for each member are proportional to the member's credibility.

What has been lost by the simplified approach? First, instead of (1), L_{ij} is often considered to be a conditional process with a parameter, say q_i .and a conditional mean and variance given the parameter. The conditional means are assumed to average to the class mean C with a variance t^2 and the conditional variances average to s^2. Then defining M_i as the conditional mean less C is equivalent to the additive formulation (1). However the full usual derivation gets an additional

result: a weighted average of member means is the best linear combination of any sort of the individual member observations by time period. However, this is a fairly general statement itself, and it might be true in general separate from the credibility formulation.

Both this formulation and the usual credibility derivation ignore the estimation error for C in the credibility formula. Empirical Bayes theory addresses this issue, which does make a difference in small samples. It might be possible to get the empirical Bayes results from the inverse squared error principle as well.

Beyond Bühlmann-Straub: Large vs. small risk differences

The assumption that each unit of exposure generates the same amount of loss variance is sometimes described as assuming that a large risk behaves like an independent combination of small risks. Hewitt in his 1967 paper presented some data showing this was not the case[2]. Actually large risks have more variance than would be expected from treating them as independent combinations of smaller risks. One thing that contributes to this is that risk conditions change over time. Size of exposure does not provide much stability against changing economic and business sector changes. A way to model this would be to assume that the variance of the observed loss for each risk for each period has the usual component that increases with risk size plus another component that increases with risk size squared, i.e., assume that the loss variance is $P_{ij}^2u^2 + P_{ij}s^2$. Then the variance of the pure premium would be $u^2 + s^2/P_{ij}$.

The credibility formula now gets more complicated, but is not too bad in the special case where there is just one time period. With the inverse expected squared error formula, $z = [P_{i\sim}/(P_{i\sim}u^2+s^2)]/[P_{i\sim}/(P_{i\sim}u^2+s^2)+ t^{-2}] = P_{i\sim}/[P_{i\sim} + P_{i\sim}u^2/t^2 + s^2/t^2]$. This could be written as $z = P/[P + AP + K]$. For larger values of P this makes the denominator larger, so decreases the credibility compared to P/[P+K].

In this case risk stability is a more complicated function of exposure than in the original model. In experience rating workers compensation another phenomenon has sometimes been observed:

---

[2] *Loss Ratio Distributions – A Model*, PCAS LIV.

the large risks' mean loss exposures are less different from the overall mean than are the small risks'. This could be a matter of regulation, where large risks must follow more safety precautions, but other reasons are possible. Whatever causes this phenomenon, the result is that the variance of M_i (i.e., the variance among risk means) also becomes a function of exposure. Since it is the smaller risks that have more potential for large departures from the overall average pure premium, this average becomes less relevant for the small risks, which increases the credibility of their own experience. A reasonable formula for the variance among risk means in this situation might be $t^2+v^2/P_{i\sim}$ in the single time period case for member i. Suppressing the subscripts on P, z becomes $z = [P/(Pu^2+s^2)]/[P/(Pu^2+s^2)+ P/(Pt^2+v^2)] = (Pt^2+v^2)/[Pu^2+s^2+ Pt^2+v^2]$. This can be simplified to $z = (P + B)/(P + AP + K + B)$. The extra B in the numerator and denominator increases z, especially for smaller risks where P is smaller, which is what was anticipated.

When linear estimates don't work

So far this discussion has been non-parametric. That is, the forms of the distributions have not entered in. That is the advantage of linear estimates with squared error penalties. If you have some information about the type of distribution available, you can give up the restriction to linear functions. In a Bayesian framework the class experience becomes the prior distribution for the member experience, and then the Bayesian conditional expected value of the member mean given the data is the least squares estimator of the member mean of any sort, linear or not. In some cases the conditional mean is a linear function of the data (e.g., normal and gamma distributions) so the linear restriction of credibility theory does not reduce the accuracy. However in highly skewed distributions, like some lognormal cases, the Bayes estimate is highly non-linear, and credibility weighting can give large errors for classes with small means.

If the distribution type is fairly well understood, Bayesian methods would be preferable in such cases. However, an alternative when the member means can be very different from each other is to do the usual credibility estimation in the logs of the data, then exponentiate the results. This introduces a downward bias, however, which has to be adjusted multiplicatively to balance to the overall data.