Credibility Theory

Southeastern Actuarial Conference
June 18, 2009
Don Behan, Consulting Actuary

Statistical Credibility in Actuarial Science

- Blending parameters from a subset and the whole group to project future outcomes
- Combines statistical methods and actuarial judgment
- Several methods at different levels of base statistics

Why is Credibility Important?

- Observed results for a small data set help to estimate future expectations, but have a random element.
- Results for a larger data set have greater statistical significance, but the population may have different characteristics than the subset.
- We need to combine these to make the best use of all available data.

This Presentation is an Introduction

- Sources of follow-up information
  - American Academy of Actuaries Credibility Practice Note (including references)
  - Actuarial Standard of Practice No. 25
  - Bühlman & Gisler, *A Course in Credibility Theory*
  - Klugman, *Bayesian Statistics in Actuarial Science with Emphasis on Credibility*

Be Careful with Terminology

- When communicating with non-actuaries, be careful about confusion with the dictionary meaning of "credibility."
- When we use the term credibility, we are not questioning the accuracy of the data, but noting its partial statistical weight as a projection of future expectations.

Usually not Pure Statistics

- Because of the uncertainty about the underlying probability distributions, significant actuarial judgment is needed.
- ASOP 25: "[t]he use of credibility procedures is not always a precise mathematical process."

Procedures Are Not All Well Based

- Some credibility procedures go back to a time when computations were more difficult, and may have been oversimplified.
- Desire to use a linear combination of overall and subset data may create too much restriction on the method.

What is the Goal of Credibility?

- Make the most accurate projection, taking cost-benefit into consideration
- Balance of stability and responsiveness is a natural outcome
- Concern about stability is a marketing consideration, not a basis for making the most accurate projection

Standard Credibility Methods

- Bayesian
  - Least Squares
  - Empirical Bayesian
  - Bühlman, Bühlman-Straub
- Frequentist
  - Greatest Accuracy Credibility (Bayesian?)
  - Limited Fluctuation Credibility (issues with this)

Two Examples of Applications

- Coin toss example
  - narrow population distribution implies very limited value of specific experience
- Climate example
  - uniform population distribution implies specific experience is the best predictor
- Conclusion
  - credibility method depends strongly on the situation

Bühlmann-Straub Method

- Population with mean μ_0 and variance v
- Subsets with mean μ_i, variance τ, and weight w_i
- The variance of subset i is τ/w_i

Bühlmann-Straub Method

- Calculate credibility factor K = τ/v
- Let $Z = w_i / (w_i + K)$
- Then the credibility estimate of the mean for class i is

$$Z\mu_i + (1 - Z)\mu_0$$

A Specific Example

- More details in accompanying text document.
- A study of charitable gift annuities found 744 deaths when 783 were expected on the basis of the US 2000 Basic Annuity Table.
- Commercial annuity results by amount may not apply to charitable gifts.
  - (continued)

Specific Example (continued)

- The observed result was 95% actual to expected.
- The distribution by group had mean of 100% and a standard deviation of 6%.
- A weight of 7 was assigned to the observed results, one fourth of the weight obtained from the square root of the number of deaths.
  - (continued)

Specific Example (continued)

- The Bühlmann-Straub credibility factor is 3.6% / 6% = 0.6.
- Z = 7 / (7 + 0.6) = 0.92.
- The computed projection of future mortality is $0.92 \times 95\% + 0.08 \times 100\% = 95.4\%$.

The Issue of Longley-Cook’s 1,082 Claims

- A famous paper of Longley-Cook is the basis for using 1,082 claims as the standard for 100% credibility.
- This paper has numerous assumptions that are often not mentioned.
- The number of claims is based on 90% confidence of an error less than 5%.
- Don’t use this without reading his paper.

Issue of Zero Claims

- If the number of claims is zero or very small, the central limit theorem does not help.
- Do not use normal distribution methods in this case.
- In the Poisson distribution, zero claims gives 95% confidence that the expected value is not more than three claims.

Regulatory Credibility Standards

- Several state insurance departments require specific credibility methods for certain types of claim estimates.
- The methods that I have reviewed are rather arbitrary, and have weak (if any) theoretical basis.
- The above comment does not affect the requirement to apply these standards.

Necessary to Check Total of Estimates

- Because of the fact that the multipliers for different subsets will vary, the total of the credibility estimates for subsets may not be the best estimate for the population as a whole.
- Adjustments need to be made to get the best population estimate.

Statistical Credibility Theory

Donald F. Behan
Presented to the Southeastern Actuarial Conference
June 18, 2009

Introduction

Statistical credibility theory provides methods for estimating parameters of a subset of a given population by combining results for that particular subset with results for the population as a whole, which is larger, and more statistically stable. Actuaries typically use credibility to project future experience for a subset of a population by blending the historical results of the population and the particular subset. In traditional actuarial practice, credibility theory provided a method for producing a weighted average of results for the population of interest and corresponding results for a larger population. Today the basic concepts are similar, but the methods are more general. This presentation will provide some basic examples of the application of credibility theory, and references to information that you can use to get further information about credibility theory.

The purely statistical approach to credibility theory provides a way to estimate a parameter for a subset of a population given certain probability elements. These probability elements are the observed mean and standard deviation of the data for the small data set, and the probability distribution of these parameters for subsets of the population as a whole. This would provide the basis for a credibility approach of a purely statistical nature.

Actuaries rarely use credibility theory in a purely statistical manner. The subsets of an overall population have characteristics that are not entirely known, and actuarial judgment is needed to determine the use of these characteristics. In most cases the effects of population characteristics are known to some degree, but there is not enough data to eliminate the need for some judgment in the choice of probability distributions and parameters. In addition, in a practical situation, cost benefit criteria are also appropriate to consider in determining the degree of statistical analysis that should be used to project future results. As stated in Actuarial Standard of Practice No. 25, "[t]he use of credibility procedures is not always a precise mathematical process."

Some people consider a balance of responsiveness and stability of projected results to be the goal of applying credibility theory, but I don't agree with this point of view. I believe that the actual goal is to get the most accurate answer, taking cost benefit considerations into account. The balance between responsiveness and stability results from the best approaches to reaching this goal, rather than actually being the goal. In other words, a projection that takes into account actual experience (responsiveness), but recognizes that taking into account the overall population results produces a better estimate (stability) produces an accurate result when credibility is between zero and 100%. Marketing considerations may lead a company to modify premium rates more gradually than would be indicated on the basis of statistical estimates, but this should not be confused with the actuarial objective of making the most accurate projection of future results.

Two Basic Examples

I took 100 coins and flipped them. I then selected the ones that came up heads, and flipped again. After six flips, only one of the coins had come up heads every time, and I wanted to project the number of times it would come up heads if I flipped that coin ten more times.

Without doing a scientific study, it seems clear that the range of probabilities of a coin coming up heads is very narrow. The coin that I selected was in the 98th percentile of the number of heads, based on my test. But if the probability of a random coin coming up heads is normally distributed with a mean of ½ and a standard deviation (based on a guess on my part) of 0.001, a coin in the 98th percentile would have a probability of 0.502 of coming up heads. In the actual example I flipped the coin ten more times, and it came up heads four times. It would take thousands of flips to come up with a result that would be significantly different from 0.5 if the coin were biased by two tenths of one percent. In this example, judgment would cause the projected results to be based only on population averages unless the number of observations were extreme.

As a second example, I obtained the temperatures in the 20 largest cities in the world on January 15th and 16th, 2009. The idea of this example is to estimate the expected January 16th temperature for a particular city based on the temperature in that city on January 15th and the average temperature for the 20 cities at that time of year. It turns out that the temperatures for the 20 cities on these dates are approximately uniformly distributed between 10° F and 90° F. The change in temperature for a given city between the two days is approximately normally distributed, with a standard deviation of 3° F. This means that if the temperature of the particular city chosen is not within about 10° of the end of the uniform distribution, virtually no information can be derived from the results for the group of 20 cities, so that the entire projection will be based on the results for the individual city. In this example, population results have little, if any, effect on the projection for a specific selection.

The point of these two examples is that the correct application of credibility is not possible through a simple formula that can be applied in all cases. The approach used should take into account the probability distribution of the entire population and the distributions of the subsets for which the estimates are to be made.

Fundamental Methods of Credibility Theory

The methods used to apply credibility theory are sometimes divided into two categories, frequentist methods and Bayesian methods. This distinction is not very clear, as credibility theory essentially uses the concepts of Bayesian statistics, even when the method is considered "frequentist."

The methods traditionally used by actuaries are the frequentist methods. These are known as the greatest accuracy method and the limited fluctuation method. I will not be discussing the limited fluctuation method, except to say that it has some serious flaws, and is based on an attempt to combine statistical methods and marketing methods. There are enough valid methods that it seems to me that there is no reason to continue dealing with the limited fluctuation method.

An example of the greatest accuracy method is the following:

Assume that the claims of individuals in a given underwriting class are independently normally distributed with variance τ. Also assume that the means of the classes are normally distributed with mean μ and variance ν. Then the projected mean of an individual class is given by a formula involving the credibility parameter

$$Z = \frac{n}{n + \tau/\nu}$$

where *n* is the number of claims for the class. Given the mean for the class equal to $\bar{x}$, the projected value for the class taking credibility into account is $Z\bar{x} + (1 - Z)\mu$. This approach to credibility has limited effectiveness, because the assumptions about the distributions are rarely met in practice.

The Bühlmann-Straub method (also called empirical Bayesian) is very similar, but applies in a much more general situation. The basic approach is to require that the formula be based on a linear combination of the results of the total population and the given subset. More details of this method can be found in Chapter 4 of the book by Bühlmann and Gisler cited in the references.

In the Bühlmann-Straub method it is assumed that each class has a claim ratio X_i and a weight w_i. The class has an expected value μ_i and variance τ/w_i, and the classes are assumed to be independent of each other. The means of the different classes are assumed to be distributed with mean μ_0 and variance ν.

Then the Bühlmann-Straub method uses a credibility factor $K = \tau/\nu$. The projected mean for class *i* is $Z\mu_i + (1 - Z)\mu_0$, where

$$Z = \frac{w_i}{w_i + K}.$$

The historical approach to credibility theory, in which the degree of credibility is based on a fixed formula involving the number of claims, is appropriate only if a series of specific assumptions is met. For example, the famous 1962 paper of L. H. Longley-Cook has been used to create a standard formula for credibility, but aside from the statistical assumptions underlying Mr. Longley-Cook's results (which are frequently not met in practice), the results that are used, such as the well known number 1,082 for full credibility, are based on specific standards for confidence and error margins. The 1,082 number is based on 90% confidence that the value will be within 5% of the true value. These standards may or may not be appropriate in a given situation. Of course, if the subset of the population is large enough you may decide that blending with other data is unnecessary, which is the result of 100% credibility. Rather than trying to use a standard number of claims as the basis for 100% credibility, it is more appropriate to base this decision on the specific characteristics of the statistical situation under analysis, and the effects of potential inaccuracies.

A Specific Example

In a recent study of mortality for donors of charitable gift annuities the number of deaths was 744 compared to 783 expected on the basis of the US 2000 Basic Annuity Table. This represented an actual to expected ratio of 95%. The calculated standard deviation of this result was 3.6%. Charities wanted to determine the expected future mortality rate, taking into account their experience, but recognizing that it did not have 100% credibility.

The recent Society of Actuaries annuity mortality study[1] showed a strong relationship between annuity amount and mortality, with lower mortality for annuities of higher amount. Charitable annuities tend to be of smaller amount than commercial annuities, but the donors are typically of a high socioeconomic status. Therefore, it was assumed that the distribution by amount might provide a helpful basis for projecting future mortality, but that, rather than using the actual amounts, a credibility method based on the mortality distribution could be used. Taking mortality improvement into account, and the fact that the charitable annuity study was based on data four years earlier than the Society of Actuaries study, the distribution of mortality by group appeared to be approximately a normal distribution with an expected value of 100% and a standard deviation of 6%.

Using the Bühlmann-Straub method it is necessary to have a weight for this subset. A purely statistical weight for this purpose would be the square root of the claim count, but it was felt that more weight should be applied to the Society of Actuaries study data, which was based on a very large data set. Therefore, a judgment decision was made to use approximately one-fourth of this weight, or a weight of 7 for the charitable annuity experience. The credibility factor for the Bühlmann-Straub method is then $K = 3.6\% / 6\% = 0.6$. The credibility multiple is

$$Z = \frac{w}{w+K} = \frac{7}{7+0.6} = 0.92$$

The projected future mortality ratio is then $0.92 \times 95\% + 0.08 \times 100\% = 95.4\%$.

Note that this is not a purely statistical result, because there was no overall study that had groups corresponding to charitable gift annuities, so judgment had to be used to select the weight to be applied.

---

[1] Society of Actuaries 2000-2004 Individual Payout Annuity Report, April, 2009.

Adjustments to Assure Overall Adequacy

Since the credibility parameters vary on the basis of the sizes of the various subsets and their different statistical parameters, the sum of the credibility results for the subsets may not equal the best estimate for the population as a whole. Therefore, it is necessary to review the overall results, and possibly make adjustments, to ensure the adequacy of the total of the subset results.

A Specific Issue Regarding Zero Observed Claims

Actuaries frequently use continuous distributions to approximate discrete distributions, such as using the normal distribution to approximate the number of claims when the actual or expected number of claims is large. The central limit theorem says that the sum of results from a continuous distribution (with easily satisfied assumptions) approaches the normal distribution as the number of observations goes to infinity. This type of approximation breaks down when there are zero or a small number of claims. One of the problems with the application of credibility methods occurs in this situation, but rather than actually representing a credibility problem, the real issue is using an approximation to the actual claim distribution in a situation where the approximation is no longer appropriate.

A reasonable and practical approach in the case of zero or a small number of claims is to use the Poisson distribution. For example, the probability of a zero outcome with the Poisson distribution is 5% when the expected number of claims is three. This means that a zero claim count provides 95% confidence that the expected number of claims is three or less. The normal distribution does not provide a reasonable way to get to this result.

Use Caution when Using the Term Credibility with Non-Actuaries

We need to be careful in discussing the credibility of data to be clear that we are talking about statistics. Of course the dictionary meaning of credible is “believable.” As actuaries, when we say that data has low credibility, we do not mean that it is not believable, but that we would assign less than 100% statistical weight in a weighted average. People who are unfamiliar with the statistical meaning of the word may incorrectly interpret our statement to mean that we don’t believe that the data is accurate. The term “statistical credibility” tends to make our meaning more clear.

References

Actuarial Standards Board, *Actuarial Standard of Practice No. 25, Credibility Procedures Applicable to Accident and Health, Group Term Life, and Property/Casualty Coverages,* October, 1996.

American Academy of Actuaries, *Credibility Practice Note,* July, 2008.

Bühlmann, Hans and Alois Gisler, *A Course in Credibility Theory,* Springer-Verlag, New York, 2005.

Klugman, Stuart A., *Bayesian Statistics in Actuarial Science with Emphasis on Credibility,* Kluwer Academic Publishers, Boston, 1991.