The Rev. Thomas Bayes and Credibility Theory

IN HIS ***FOUNDATIONS OF CASUALTY ACTUARIAL SCIENCE***, Matthew Rodermund wrote, “It is the concept of credibility that has been the casualty actuaries' most important contribution to actuarial science.” I would expand that assessment a bit further by adding that the best approach to credibility is via a theorem of the Rev. Thomas Bayes (1702-1761), whose work has had a tremendous influence on modern actuarial and statistical analyses, especially credibility theory.

Born in London, Bayes served as a nonconformist minister in Turnbridge Wells, about 35 miles southeast of London. Although he was elected a fellow of the Royal Society in 1742, his membership was something of a mystery since his major work was not published until 1764, three years after his death. However, recently discovered letters indicate that Bayes corresponded privately with many of the leading intellectuals of his era. Bayes' work was showcased posthumously by his fellow clergyman Richard Price and championed and extended by the French astronomer, probabilist, and mathematician Pierre Simon Laplace.

Like Bayes, Price (1723-1791) was the son of a dissenting minister. Interestingly, Price made his reputation as a financial writer and political commentator. In a 1769 letter to Benjamin Franklin, Price discussed life expectancy and population increase (his observations were subsequently published in the *Philosophical Transactions* of that year), and a year later he presented a paper to the Royal Society on the proper method of calculating the values of contingent reserves. His *Appeal to the Public on the Subject of the National Debt,* first published in 1771, is credited with pushing William Pitt the Elder to re-establish a sinking fund for the extinction of the national debt. Laplace (1749-1827) produced seminal work in astronomy, mathematics, and probability. Mathematicians are all familiar with the Laplace transform in functional analysis. In a philosophical essay on probabilities, Laplace described a mathematical framework for conducting statistical inference that extended Bayes' work and constituted the essence of Bayesian statistical inference.

Statistical Paradigms

There are currently two main paradigms to statistical inference: sampling theory, also known as the frequentist approach, and the Bayesian approach.

Under the frequentist approach, the probability of an event is its relative frequency. Random variables, such as the aggregate claim amount in a period of observation, are assumed to have probability distributions whose parameters are constants. Prior information enters statistical model building only in the selection of a class of models. With this paradigm, the main tools of statistical inference are confidence intervals and tests of hypotheses.

Under the Bayesian approach, probability is treated as a rational measure of belief. Random variables, such as the aggregate claim amount in a period of observation, are assumed to have probability distributions whose parameters may also have probability distributions. Both recent information and prior opinions (and/or information available before the recent information) are used to construct the probability distributions of the statistical models or parameters of interest. Thus, the Bayesian approach is based on personal or subjective probabilities and utilizes a theorem developed by Bayes that states that if A and B are events, then P[B]>0, then

$$P[A \mid B] = \frac{P[B \mid A] \cdot P[A]}{P[B]}$$

Besides this theorem, the main tools of Bayesian statistical inference are predictive distributions, posterior distributions, and (posterior) odds ratios.

Bayes' Theorem is important to actuaries because it enables them to perform statistical inference by computing inverse probabilities, but it has practical applications in many other fields as well (Michael Kanellos, for instance, has written in CNET News.com about the application of Bayes' theorem to computer data searches). What exactly do actuaries mean by inverse probabilities? We use the term inverse because we are inferring backwards from results (or effects) to causes. Let's look at some simple examples to examine this further.

A typical probability problem might be stated as follows: I have a standard die with six sides numbered from 1 through 6 and throw the die three times. What is the probability that I will roll a 6 each time I toss the die? Now, I might have a second (non-standard) die with three sides numbered 1 and three sides numbered 6. Again, I can ask the same question: What is the probability that the result of each of three tosses of this die will be a 6?

THOMAS HERZOG is chief actuary for the Federal Housing Administration in Washington.

The idea behind inverse probabilities is to turn the question around. Here, I might observe that the results of three throws of a die were all 6. I would then ask the question: What is the probability that I threw the standard die (as opposed to the non-standard die), given these results?

Predictive Distributions

In insurance work, we typically experience a number of claims or an aggregate amount of losses in one or more prior observation periods. The questions we want to answer are:

(1) Given such results, how many claims will we experience during the next observation period?

(2) Given such results, what will be the aggregate loss amount during the next observation period?

Using Bayes' Theorem, we can construct an entire probability distribution for such future claim frequencies or loss amounts. Probability distributions of this type, usually called predictive distributions, give an actuary much more information than is available in an average or other summary statistic (such as the expected aggregate amount of losses in the next period). It shows us, for instance, a complete profile of the tail of the probability distribution of aggregate losses that are of note in a value-at-risk analysis. Predictive distributions are an important tool with which to make business decisions under uncertainty.

Basic Credibility Concepts

Under some approaches to credibility, a compromise estimator, C, is calculated from the relationship

$$C = ZR + (1 - Z)H$$

where R is the mean of the current observations (for example, the data), H is the prior mean (for example, an estimate based on the actuary's prior data and/or opinion), and Z is the credibility factor, satisfying the condition The credibility factor, Z, is the weight assigned to the (current) data; (1-Z) is the weight assigned to the prior data.

As an insurance example, a new premium rate, C, is derived as a weighted average of an old premium rate, H, and a premium rate, R, whose calculation is based solely on observations from a recent period. An alternative interpretation is to let C be the premium rate for a particular class of business, to let R be the premium rate whose calculation is based solely on the recent experience of that class, and to let H be the insurance rate whose computation takes into account the recent experience of all classes combined.

The credibility factor, Z, is typically of the form

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

where k > 0 and n is a measure of exposure size. Under such a formulation, the problem becomes how to determine k.

Limited Fluctuation Approach to Credibility

The goal of the limited fluctuation approach is to find a compromise estimate, C, of the form

$$C = ZR + (1 - Z)H.$$

The limited fluctuation approach is one of the oldest, going back at least as far as A.H. Mowbray's 1914 paper and F.S. Perryman's 1932 paper. More modern treatments, including complete mathematical derivations, are found in a 1962 article by L.H. Longley-Cook and in Chapter 5 of my 1999 textbook, *Introduction to Credibility Theory.*

The main advantage of the limited fluctuation approach is that it is relatively simple to use. But a number of researchers and practitioners have raised questions about the method.

The Swiss mathematician Hans Bühlmann reportedly felt that the mathematical reasoning behind the limited fluctuation approach as presented by Longley-Cook wasn't convincing. Bühlmann noted that:

- The reasoning was under the frequentist paradigm of statistics and so ignored prior data;
- The approach began by deriving the minimum volume of risks required for full credibility;
- The derivation was performed using a confidence interval;
- The approach used a standard deviation argument to go from the full credibility requirement to partial credibility for a smaller volume of risks.

Bühlmann then raised the following concern with Longley-Cook's derivation: Why should a confidence interval that, by definition, includes the true value with a probability of less than 1, guarantee full credibility?

Others have asked why the same weight, 1-Z, is given to the prior mean, H, regardless of the analyst's view of the accuracy of H. Also, in the case of full credibility, no weight is given to the prior mean, H, as all of the weight is given to the observed data, R. This raises the philosophical question of whether it makes sense to talk about full (i.e., 100 percent) credibility because more data can generally be obtained. Some analysts believe that no data are entitled to full credibility, so that the credibility curve should approach 1 asymptotically, without ever reaching it.

Finally, according to B. Sundt, there is an internal inconsistency in the limited fluctuation approach to full credibility. He notes that the criterion for replacing the old premium rate is based on the assumption that the old premium rate is correct, leading to the following conundrum: If the old premium is correct, then why replace it?

Bayesian Approach

In 1918, A. Whitney stated that the credibility factor, Z, needed to be of the form

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

where n represents earned premiums and k is a constant to be determined. The problem was how to determine k. Whitney wrote that "(i)n practice k must be determined by judgment" and added that "(t)he detailed solution to this problem depends upon the use of inverse probabilities via Bayes' Theorem." Whitney was right!

The credibility estimates produced by Bühlmann's approach are the best linear approximations to the corresponding Bayesian point estimates. Because it typically requires less mathematical skill and fewer computational resources and doesn't require the selection of a prior distribution (which can sometimes require difficult judgments), Bühlmann's approach to producing point estimates is often much more computationally tractable than a complete Bayesian approach.

Bühlmann expressed strong support for the Bayesian paradigm but felt that whenever possible the prior should be based on experience data rather than on subjective judgment. To quote Bühlmann, as reported by J.C. Hickman and L. Heacock in an article that appeared in *NAAJ* in 1999,

> *Whereas early Bayesian statisticians used the prior distribution of risk parameters as a means to express judgment (which in insurance we would call underwriting judgment), [I] think of the probability distribution of the risk parameters as having an objective meaning. It hence needs to be extracted from the data gathered about the collective. (Only in the case of a lack of such data might one accept the subjective view faute de mieux.) For this reason, I have always insisted on speaking about the structural distribution of risk parameters, avoiding the standard Bayesian terminology, prior distribution.*

Writing in 2006, Stephen Fienberg referred to "a long line of efforts to discover the statistical holy grail: [objective] prior distributions reflecting ignorance." Here he uses the term "ignorance" as the opposite of highly subjective personal beliefs that many actuaries and statisticians view as being unscientific.

With the increased power of 21st-century computing equipment, advances in statistical algorithms (e.g., the EM algorithm and Markov chain Monte Carlo methods) used to implement the Bayesian approach, and widely available software that performs Bayesian inference (i.e., WinBUGS[1]), a wider class of problems is becoming susceptible to solution via the Bayesian approach. ●

References

Fienberg, S.E., "When Did Bayesian Inference Become 'Bayesian'?" *Bayesian Analysis*, Vol. 1., No. 1, 2006, pp 1-40

Herzog, T.N., *Introduction to Credibility Theory* (Third Edition). Winsted: ACTEX Publications, 1999

Hickman, J.C. and L. Heacox, "Credibility Theory: The Cornerstone of Actuarial Science," *NAAJ*, Vol. 3 No. 2, April 1999

Kanellos, M., *18th-century theory is new force in computing*, CNET News.com, Feb. 18, 2003 (available at http://news.com.com/Old-school+theory+is+a+new+force/2009-1001_3-984695.html)

Laplace, P.S., *Philosophical Essay on Probabilities*, translated from the fifth French edition of 1825 by Andrew I. Dale. New York: Springer, 1995

Longley-Cook, L.H., "An Introduction to Credibility Theory," *PCAS*, 49 (1962), 194-221

Mowbray, A.H., "How Extensive a Payroll Is Necessary to Give a Dependable Pure Premium?" *PCAS*, 1 (1914), 24-30

Perryman, F.S., "Some Notes on Credibility," *PCAS*, 19 (1932), 65-84

Roddermund, M., "Introduction," Chapter 1 of *Foundations of Casualty Actuarial Science.* New York: Casualty Actuarial Society, 1989

Stigler, S.M., *The History of Statistics; the Measurement of Uncertainty Before 1900.* Cambridge: Harvard University Press, 1986

Sundt, B., "On Greatest Accuracy Credibility with Limited Fluctuation," *Scandinavian Actuarial Journal*, 2 (1992), 109-119

Venter, G., "Credibility," Chapter 7 of *Foundations of Casualty Actuarial Science.* New York: Casualty Actuarial Society, 1989

Whitney, A., "The Theory of Experience Rating," *PCAS*, 5 (1918), 274-292

Note

1. The BUGS (**B**ayesian inference **U**sing **G**ibbs **S**ampling) Project (begun by the MRC Biostatistics unit at Imperial College, London) is concerned with the development of flexible software for Bayesian analysis of complex statistical models using Markov chain Monte Carlo methods. The "**Win**" prefix refers to Microsoft's Windows operating system.