ACTUARIAL RESEARCH CLEARING HOUSE
1988 VOL. 2

Bayesian Credibility -- A Summary of AERF Sponsored Research

Stuart Klugman
The University of Iowa
and
Drake University

Paper Presented at the 22nd Actuarial Research Conference

Abstract

This paper presents a brief description of recent work on Bayesian Credibility through the hierarchical normal linear model. This work was funded during the summer of 1986 by the Actuarial Education and Research Fund. Details of the research can be found in the papers Klugman (1987a) and Klugman (1987b).

0. Introduction

The term credibility has been used to mean a variety of things and most often it implies a restriction to linear estimators. This is usually done to avoid making a distributional assumption. Least squares estimates of means will then depend on the underlying distribution only through the first two moments. However, if too much emphasis is placed on linearity one may lose sight of the original objective, the estimation of a set of parameters. In the insurance setting this is usually the estimation of expected losses or loss ratios. The goal, then, is the same as in any statistical estimation problem: to find point estimates, to find measures of the accuracy of the point estimate (standard errors), and to verify the assumptions made (or the model used). It is not possible to perform the last two tasks unless some assumption is made with regard to the underlying distribution of the observations.

In this research the normal distribution was selected as the model to use. The reasons for this choice are given in Section 1. The next step is to select an estimation method. Arguments for a hierarchical Bayesian model are presented in Section 2. In Section 3 the most general solution for linear models is presented. In Section 4 the specific solution for the simplest model is given. In Section 5 a number of models are presented that can be solved with the approach given in Section 3. Finally, in Section 6 some topics for future research are presented.

1. The Normal Model

There are two reasons for choosing the normal distribution. The first is computational convenience. Quite simply, it is (relatively) easy to obtain answers. As our computational ability increases, this advantage is likely to disappear. The second reason is a more practical one. Those who make no distributional assumptions solve the problem by restricting attention to estimators that are linear in the data. It turns out that these same linear estimators follow from the normal model. Furthermore, most practitioners estimate the second moments by using sums of squares. While their estimators are unbiased regardless of the underlying distribution, any optimality properties relate to the normal model.

2. The estimation method

Several methods are currently in use for performing the estimation. All of them agree on the point estimator to use in the case where the population variances are known. The divergence comes in the estimation of these variances. Most of these methods can be categorized as empirical Bayes methods if one is willing to adopt a loose definition. By an empirical Bayes method I mean any method that first finds the Bayes solution in the case where all second moments are known, and then uses any reasonable technique for estimating the variances. The usual choice is to find unbiased estimators based on carefully selected sums of squares. This is the method used by Bühlmann and Straub (1972) in their classic paper that introduced actuaries to this approach and has continued to be advocated. In a more recent paper Venter (1985) extends this procedure to more complicated models.

There are three major drawbacks to this approach. The first is that the estimates can easily fall outside the parameter space. This is because they are basically method of moments estimates. A second is that standard errors are not available. There is no way to account for the variability introduced via the estimation of the variances. Finally, because the method itself provides no guidelines for divining the variance estimates, extension to complicated models becomes very difficult.

I advocate a true Bayesian approach. That is, prior distributions for all parameters must be given and must themselves be free of unknown parameters. Once this is established it is merely a matter of applying the rules of probability to obtain the posterior distribution of any parameter of interest. If the

posterior mean is used as the point estimate and the posterior variance used as a measure of error, the first two objections in the preceding paragraph are met. The difficulty associated with complex models will depend on the computational difficulty of obtaining the posterior distribution. But the difficulty is always computational and never conceptual.

3. The Hierarchical Normal Linear Model

This model, introduced by Lindley and Smith (1972), provides considerable flexibility, yet is not too difficult to manipulate. The model begins by assuming that the observations are recorded in a column vector $\mathbf{Y}$. It is further assumed that given a vector $\boldsymbol{\theta}$ and a matrix G

$$\mathbf{Y} \mid \boldsymbol{\theta}, G \sim N(A\boldsymbol{\theta}, G)$$

where A is a matrix of known constants. Our goal is to estimate the elements of the vector $\boldsymbol{\theta}$. All credibility type estimation procedures are based on an assumption that there is some relationship between the parameters in $\boldsymbol{\theta}$. Here it is modeled by assuming

Bayesian Credibility

Bayesian approach a prior free of parameters must be specified, that is

$$p(\boldsymbol{\mu},G,H).$$

Obtaining the solution is a matter of performing the probability manipulations. The results are given below.

Let

$$\Lambda = (A'G^{-1}A)^{-1}$$
$$\Xi = (B'H^{-1}B)^{-1}$$
$$\Psi = \Lambda + H.$$

The least squares estimates are

$$\hat{\boldsymbol{\theta}} = \Lambda A'G^{-1}\mathbf{Y} \quad \text{and}$$

$$\hat{\boldsymbol{\mu}} = (B'\Psi^{-1}B)^{-1}B'\Psi^{-1}\hat{\boldsymbol{\theta}}.$$

The conditional distribution of $\boldsymbol{\theta}|\mathbf{Y},G,H$ is multivariate normal with mean vector

$$\tilde{\boldsymbol{\theta}} = (H^{-1} + \Lambda^{-1})^{-1}(\Lambda^{-1}\hat{\boldsymbol{\theta}} + H^{-1}B\hat{\boldsymbol{\mu}})$$

and covariance matrix

$$V = (\Lambda^{-1} - H^{-1}B\Xi B'H^{-1} + H^{-1})^{-1}.$$

The posterior density of $G,H|\mathbf{Y}$ is proportional to

$$p(G,H)(|\Lambda|/|G|)^{1/2}(|(B'\Psi^{-1}B)^{-1}|/|\Psi|)^{1/2}$$
$$\times \exp[-(\mathbf{Y}-A\hat{\boldsymbol{\theta}})'G^{-1}(\mathbf{Y}-A\hat{\boldsymbol{\theta}})/2 - (\hat{\boldsymbol{\theta}}-B\hat{\boldsymbol{\mu}})'\Psi^{-1}(\hat{\boldsymbol{\theta}}-B\hat{\boldsymbol{\mu}})/2].$$

Bayesian Credibility

The posterior mean can be obtained by integration:

$$E(\theta_i | Y) = \int \tilde{\theta}_i p(G,H|Y)dGdH$$

where $\tilde{\theta}_i$ is the ith element of $\tilde{\boldsymbol{\theta}}$.

The posterior variance is

$$Var(\theta_i | Y) = \int (\tilde{\theta}_i)^2 p(G,H|Y)dGdH - [E(\theta_i | Y)]^2 + \int v_{ii} p(G,H|Y)dGdH$$

where v_{ii} is the ith diagonal element of V.

A final item of interest is the predictive density of the next observation from the ith class. In particular, if $X_i | \theta_i \sim N(\theta_i, s)$ then

$$E(X_i | Y) = E(\theta_i | Y)$$

and

$$Var(X_i | Y) = Var(\theta_i | Y) + E(s | Y)$$

where s is a function of the elements of G and the final expectation is with respect to p(G,H|Y).

4. The simplest case

The most elementary case is one in which there are k classes of insureds and n_i observations have been taken from each class. The objective is to estimate θ_i, the mean loss (or loss ratio) from the ith class. The model becomes

$$Y_{ij} | \theta_i, \sigma^2 \sim N(\theta_i, \sigma^2 / P_{ij})$$

$$\theta_i | \mu, \tau^2 \sim N(\mu, \tau^2)$$

where P_{ij} is a measure of the volume of business for the jth observation from class i. Applying the above formulas reveals that given σ^2, τ^2, and the data, $\boldsymbol{\theta}$ is multivariate normal with means

$$\tilde{\theta}_i = w_i\hat{\theta}_i + (1-w_i)\hat{\mu}$$

where

$$\hat{\theta}_i = \Sigma_j P_{ij} Y_{ij} / \Sigma_j P_{ij} \quad \text{and}$$

$$\hat{\mu} = \Sigma_i w_i \hat{\theta}_i / \Sigma_r w_r .$$

The variance is $\tau^2(1-w_i)[1 + (1-w_i)/\Sigma_r w_r]$ while the covariance of θ_i and θ_s is $\tau^2(1-w_i)(1-w_s)/\Sigma_r w_r$.

The other important quantity is the posterior density of the variances. It is

$$p(\sigma^2,\tau^2|Y) \propto p(\sigma^2,\tau^2)(\sigma^2)^{-(N-k)/2}(\tau^2)^{-(k-1)/2}[\Pi_i w_i/\Sigma_i w_i]^{1/2} \times \exp[-\Sigma_{ij}(y_{ij}-\hat{\theta}_i)^2/2\sigma^2 - \Sigma_i w_i(\hat{\theta}_i-\hat{\mu}_i)^2/2\tau^2]$$

where $N = \Sigma_i n_i$ is the total number of observations. The prior distribution can be selected to provide virtually no information or an inverse gamma distribution can be used if the investigator has a good idea about these variances. Obtaining the estimates and their variances is now a matter of performing a two-dimensional integral. One dimension can be accomplished analytically, leaving a single integration that must be performed numerically. The functions are sufficiently well behaved to make this a relatively simple process.

In Klugman (1987a) two data sets are analyzed using the above formulas. One set included the results of the following year and

5. Other models

A variety of interesting models that are relevant to estimation of insurance losses fit the framework of the hierarchical normal linear model. A few examples are given in this section. To simplify the writing of the models, the elements being conditioned upon are omitted from the left hand side. It should be obvious what they are.

A. Unequal variance

A model with the variance varying from class to class is given by

$$Y_{ij} \sim N(\theta_i, \sigma_i^2/P_{ij})$$
$$\theta_i \sim N(\mu, \tau^2)$$
$$\sigma_i^2 \sim \text{Inverse Gamma}(\nu, \lambda)$$
$$p(\mu, \tau^2, \nu, \lambda).$$

It turns out that the individual variances are estimated by a weighted average of the sample variance from the ith group and the harmonic mean of the k sample variances.

B. Uncertainty

This model allows the group mean to fluctuate over time. The simplest version treats it is a purely random process. More complex models could introduce autoregressive or linear trend relationships.

$$Y_{ij} \sim N(\alpha_{ij}, \sigma^2/P_{ij})$$
$$\alpha_{ij} \sim N(\theta_i, \gamma^2$$
$$\theta_i \sim N(\mu, \tau^2)$$
$$p(\mu, \sigma^2, \tau^2, \gamma^2)$$

This model is examined in Klugman(1987b).

C. Time series, state space

The Kalman filter has become popular as a general model for insurance losses. It is similar to the uncertainty model but is more structured. There are two processes at work. One governs the year to year changes in the parameters. The other describes how the parameters affect the observations. A model describing a linear trend with noise is

$$
\begin{aligned}
&Y_{ij} \sim N(\theta_{ij}, \sigma^2/P_{ij}) \\
&\theta_{ij} \sim N(\beta_i + \theta_{i,j-1}, \tau^2) \\
&p(\beta_i, \sigma^2, \tau^2).
\end{aligned}
$$

The second level indicates that the mean loss in year j is the previous year's loss plus a constant (the slope) that may vary from group to group but remains fixed over time plus some random movement.

D. Whittaker graduation

This example shows that there are other problems of interest to actuaries that can also be solved with this model. Let the θ_i be a sequence of true mortality rates and let the Y_i be estimates based on experience at age i. Consider the following model:

$$
\begin{aligned}
&Y_i \sim N(\theta_i, \sigma^2/P_i) \\
&(\theta_1, \ldots, \theta_n)' \sim N[(0, \ldots, 0)', H/\lambda] \\
&p(\sigma^2, \lambda).
\end{aligned}
$$

The posterior mean of $\boldsymbol{\theta}$ given the structural parameters σ^2 and λ is

$$(P/\sigma^2 + \lambda H^{-1})^{-1} PY/\sigma^2$$

where P is a diagonal matrix with the P_i on the diagonal. But this is just the formula for a Whittaker graduation if the matrix H is selected in the appropriate manner. What is interesting is that the techniques of Bayesian analysis make it possible to estimate the smoothing parameter λ and also to estimate the standard error in the results after the graduation has been performed.

6. Future work

There are a number of problems that remain to be investigated. A few of them are listed below:

Working out the details for the models presented in the previous section as well as other models not given there.

Investigating, probably through simulation studies, the sensitivity of the results to the normality assumption.

Development of methods for non-normal models.

References

Bühlmann, H. and Straub, E. (1972), "Credibility for Loss Ratios," English translation appearing in *ARCH*, **1972.2**.

Klugman, S. (1987a), "Credibility for Classification Ratemaking via the Hierarchical Normal Linear Model," *Proceedings of the Casualty Actuarial Society*, **74**, to appear.

Klugman, S. (1987b), "Credibility Under Parameter Uncertainty: An Illustration of the Hierarchical Normal Linear Model," Submitted for publication.

Lindley, D. and Smith, A. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society, Series B*, **34**, 1-41.

Venter, G. (1985), "Structured Credibility in Applications -- Hierarchical, Multidimensional, and Multivariate Models," *ARCH*, **1985.2**, 267-308.