ACTUARIAL RESEARCH CLEARING HOUSE
1996 VOL. 1

"Credibility Using a Loss Function from Spline Theory"

Virginia R. Young
University of Wisconsin - Madison

Current formulas in credibility theory often calculate net premium as a weighted sum of the average experience of the policyholder and the average experience of the entire collection of policyholders. Because these formulas are linear, they are easy to use. Another advantage of linear formulas is that the estimate changes a fixed amount per change in claim experience, if an insurer uses such a formal, then the policyholder can predict the change in premium. On the other hand, Venter (1990) points out that in some cases, the loss of accuracy makes a linear formula undesirable.

We apply decision theory to develop a credibility formula that minimizes a loss function that is a linear combination of a squared-error term and a second-derivative term. The squared-error term measures the accuracy of the estimator, while the second-derivative term constrains the estimator to be close to linear. An actuary may balance the sometimes conflicting goals of accuracy and linearity by changing a single parameter in the loss function. Our loss function is similar to one used in spline theory, although in a different context.

Venter, G. (1990), Credibility, *Foundations of Casualty Actuarial Science*, Casualty Actuarial Society, New York, New York.

Credibility using a Loss Function from Spline Theory

Virginia R. Young
University of Wisconsin-Madison

Actuarial Research Conference
Pennsylvania State University
17-19 August 1995

- Most credibility formulas calculate expected claims as a linear convex combination of the average of the policyholder's experience and the average of the book of business
- Advantages of a linear formula:

    easy to use

    changes a fixed amount per unit change in claim experience

    predictable from the standpoint of the ph
- Disadvantage of a linear formula

    may not be as "accurate" as the predictive mean (Venter, 1990)

Possible Solution:

- minimize the expected value of a loss function that is a linear combination of a squared-error term and a second-derivative term
  - squared-error term measures the accuracy of the estimator
  - second-derivative term constrains the estimator to be close to linear
- balance the (sometimes) conflicting goals of accuracy and linearity

Assumptions:

- claims of a policyholder is a random variable $X|(\Theta = \theta)$
- random variables $X|\theta$ are independent for different values θ of Θ
- write $X_i|\theta$ for the claims in year i, $i = 1,\ldots,n$
- random variables $X_i|\theta$, $i = 1,\ldots,n$, are independently and identically distributed according to a probability density function $f(x|\theta)$
- θ is fixed for a given ph, although it is generally unknown
- the support of the marginal probability density function of X, $f(x)$, is the interval $[a, b]$, a finite interval
- f is non-zero on $[a, b]$

Loss Function from Spline Theory

- the insurer has n years of claim experience for a policyholder: $\boldsymbol{x} = <x_1, \ldots, x_n>$ in $[a, b]^n$
- denote a credibility estimator by $d(\boldsymbol{x})$, in which d is a real-valued function on $[a, b]^n$
- find d to minimize
- $$E[(d(\mathbf{x}) - X_{n+1})^2] + hE[Lin(d)],$$
- in which $$Lin(d) = \sum_{i=1}^{n} \left(\frac{\partial^2 d}{\partial x_i^2} \right)^2 + c \sum_{i<j} \left(\frac{\partial^2 d}{\partial x_i \, \partial x_j} \right)^2$$
- *Lin* is a linearity penalty with null space equal to the set of functions that are linear in the data, with no interactions-- c is a positive constant that controls the relative weight we place on the penalty for interaction terms
- h is a positive constant that controls the degree to which d is close to linear

Simplification

- We find d:[a, b] to R to minimize the following:
- $$E(d) = E[(d(\bar{x}) - \mu\,(\bar{x}))^2] + hE[(d''(\bar{x}))^2],$$
- in which $\mu(\bar{x})$ is the predictive mean of X_{n+1} given $\bar{X}=\bar{x}$ and we take the expectation with respect to the marginal distribution of $\bar{X}$
- If we let h approach 0, then the optimal d is the predictive mean.
- If we let h approach infinity, then the optimal d converges to the Bühlmann linear credibility estimator.

Equivalent Boundary-Value Problem

- The minimizing d is the solution of the following fourth-order boundary-value problem:
- $$h\,[f(\bar{x})d''(\bar{x})]'' + f(\bar{x})d(\bar{x}) = f(\bar{x})\mu(\bar{x}),$$
- with $\quad d''(a) = d''(b) = d'''(a) = d'''(b) = 0.$
- Recall that we assume that the marginal density f is defined on a finite interval $[a, b]$ and is non-zero on that interval.
- *Theorem:* If the predictive mean is square-summable with respect to the probability measure defined by f, $E[\mu^2] < \infty$, then the solution d of the boundary-value problem has the form

$$d(\bar{x}) = (1-Z)E[X] + Z\bar{x} + \sum_{n=1}^{\infty} \frac{d_n(\bar{x})}{1 + h\,r_n} E[\mu\, d_n].$$

Theorem (cont'd)

- $$d(\bar{x}) = (1 - Z)E[X] + Z\bar{x} + \sum_{n=1}^{\infty} \frac{d_n(\bar{x})}{1 + h r_n} E[\mu\, d_n],$$
- in which $\{0\}$ plus $\{r_n\colon n > 0\}$ are the eigenvalues of the self-adjoint linear operator $(\mathcal{L}_f - 1)/h$, in which
- $$\mathcal{L}_f\, z = \frac{h}{f}[f\, z'']'' + z,$$
-
- and $\{1, \frac{\bar{x} - E[X]}{\sqrt{Var[\bar{X}]}}\}$ plus $\{d_n\colon n > 0\}$ are the corresponding orthonormal eigenfunctions.
- Also, $$Z = \frac{Var_{\Theta}E[X|\theta]}{Var\bar{X}} = \frac{Var_{\Theta}E[X|\theta]}{Var_{\Theta}E[X|\theta] + \frac{1}{n}E_{\Theta}Var[X|\theta]}.$$

Theorem (cont'd)

- $$d(\bar{x}) = (1-Z)E[X] + Z\bar{x} + \sum_{n=1}^{\infty} \frac{d_n(\bar{x})}{1+hr_n} E[\mu\, d_n]$$
- Intuitively our result makes sense: The null space of the second-derivative penalty is the space of linear functions; thus, it is indifferent to any linear portion of our solution d. The squared-error, then, forces the linear portion of d to be the Bühlmann linear estimator.
- *Aside:* The complex number λ is an *eigenvalue* of a linear differential operator $\mathcal{L}$ on $\mathcal{S}$ if there exists a non-zero function φ in $\mathcal{S}$ such that $\mathcal{L}\varphi = \lambda\varphi$. The function φ is called an *eigenfunction* of the eigenvalue λ.

Corollaries

- If the predictive mean is linear, then *d* equals the predictive mean.
- As the linearization parameter *h* approaches 0, *d* approaches the predictive mean.
- As the linearization parameter *h* approaches infinity, *d* approaches the Bühlmann linear estimator.
- The estimator *d* is unbiased and its variance is

$$\frac{(Var_{\Theta} E[X|\theta])^2}{Var[\bar{X}]} + \sum_{n=1}^{\infty} \frac{(E[\mu\, d_n])^2}{(1+h\, r_n)^2}.$$

Example

- Let $f = 1$ on the interval [0, 1]; that is, X is marginally uniformly distributed between 0 and 1. The eigenvalues r_n, for $n > 0$, are the solutions of $\cos s_n = \operatorname{sech} s_n$, in which $s_n = (r_n)^{1/4}$.
- The orthonormal eigenfunctions corresponding to the eigenvalue $r_0 = 0$ are 1 and $\sqrt{3}(2x - 1)$.
- For $\sin s_n > 0$, the eigenfunction corresponding to r_n is

$$d_n(x) = \frac{1}{\exp(s_n) - 1}[\exp(s_n x) - \exp(s_n(1-x)) - (\exp(s_n) - 1)\cos(s_n x) + (\exp(s_n) + 1)\sin(s_n x)].$$

- For $\sin s_n < 0$, the eigenfunction corresponding to r_n is

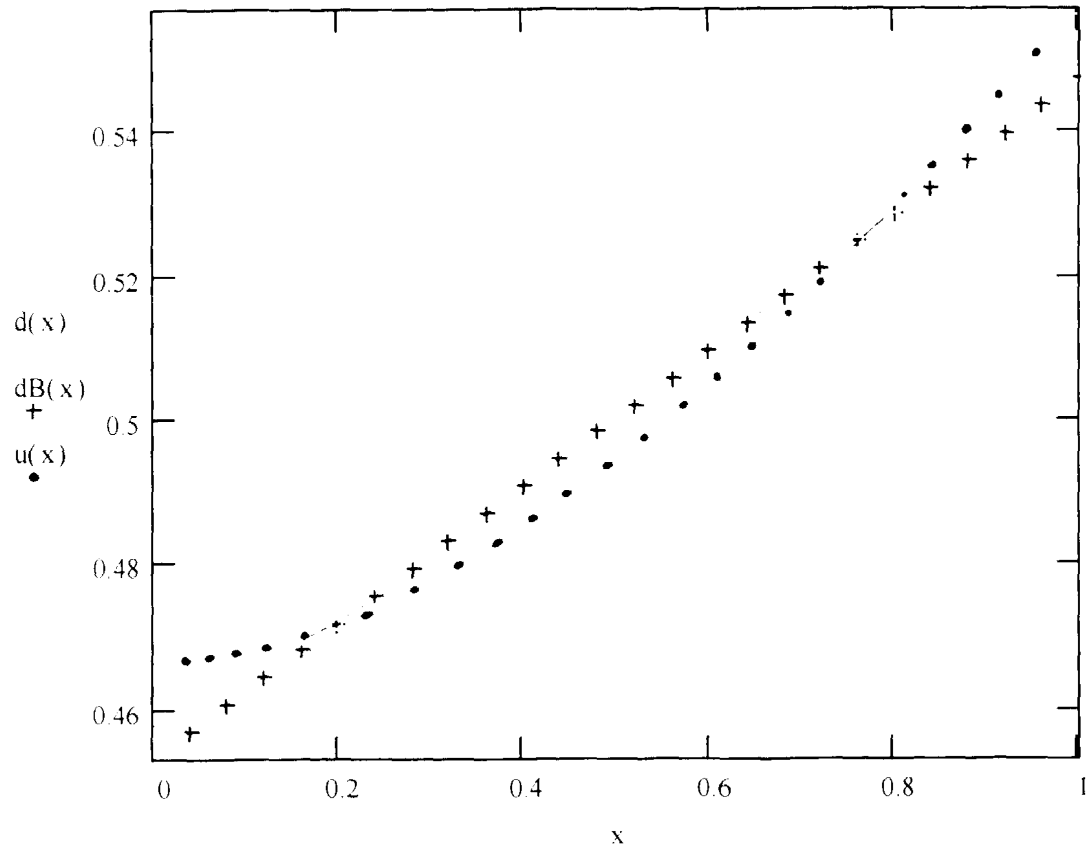
$$d_n(x) = \frac{1}{\exp(s_n) + 1}[\exp(s_n x) + \exp(s_n(1-x)) + (\exp(s_n) + 1)\cos(s_n x) - (\exp(s_n) - 1)\sin(s_n x)].$$

Example (cont'd)

- Consider
- $$f(x,\theta) = \frac{2}{x+2},\ 0 < \theta < 0.5x + 1,\ 0 < x < 1.$$
- The conditional distribution of $X|\theta$ is given by
- $$f(x|\theta) = \begin{cases} \dfrac{1}{(x+2)\ln 1.5}, & 0 \le \theta < 1; \\ \dfrac{1}{(x+2)\cdot(\ln 1.5 - \ln\theta)}, & 1 \le \theta \le 1.5. \end{cases}$$
-
-
- The predictive mean is

$$\mu(x) = \frac{2}{(x+2)\ln 1.5} - 2 + \frac{2}{x+2}\int_1^{0.5x+1} \frac{3-2\theta}{\ln 1.5 - \ln\theta}\, d\theta.$$

Example (cont'd)



Simplification (background)

- *Proposition:* An orthonormal basis of the null space of *Lin* is

$$d_{00}(\mathbf{x}) = 1$$

$$d_{01}(\mathbf{x}) = \frac{\sum_{i=1}^{n}(x_i - E[X])}{\sqrt{n(Var[X] + (n-1)Var_\Theta E[X|\theta])}}$$

$$d_{02}(\mathbf{x}) = \frac{\sum_{i=1}^{n-1} x_i - (n-1)x_n}{\sqrt{n(n-1)(Var[X] - Var_\Theta E[X|\theta])}}$$

$$\cdots$$

$$d_{0(k+1)}(\mathbf{x}) = \frac{\sum_{i=1}^{n-k} x_i - (n-k)x_{n-k+1}}{\sqrt{(n-k+1)(n-k)(Var[X] - Var_\Theta E[X|\theta])}}$$

$$\cdots$$

$$d_{0n}(\mathbf{x}) = \frac{x_1 - x_2}{\sqrt{2(Var[X] - Var_\Theta E[X|\theta])}}.$$

Simplification (background, cont'd)

- *Proposition:* The predictive mean is orthogonal to the basis functions $d_{0(k+1)}$, $k = 1, ..., n$-1, and the projection of the predictive mean onto this null space is,
- $$(1 - Z)E[X] + Z\overline{x},$$
- in which the weight Z is the Bühlmann credibility weight.