Univariate Credibility as a Boundary-Value Problem, A Symbolic Green's Function Method (Regular Case)

Windiani Erliana¹, AgahDrajat Garnadi¹, Sri Nurdianti¹ & I GustiPutu Purnaba¹

¹Department of Mathematics, Faculty of Mathematics and Natural Science, Bogor
Agricultural University, Bogor, Indonesia
Email: windi.math@gmail.com

Presented at ICREM 2017

Abstract. 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 which a formal, then the policyholder can predict the change in premium. In a series of writing, Young(1997,1998,2000) apply decision theory to develop a credibility formula that minimizes a loss function that is linear combination of a squared-error term and a second-derivative term or first order term. This loss function as a variational forms, is equivalent to fourth order or second order linear differential equation, respectively. This allows us for evaluation to Green's function computation via symbolic calculation to compute details of Green's function to obtain the solution.

Keywords: *actuarial credibility, boundary value problem, calculation, Green's function, symbolic, univariate.*

1 Introduction

Credibility models are actuarial tools to distribute premiums fairly among a heterogeneous group of policyholders. More generally, they can be seen as prediction methods applicable in any setting where repeated measures are made for subjects with different risk levels.

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. Because these formulas are linear, they

36 are easy to use. Another advantage of linear formulas is that the estimate
37 changes a fixed amount per change in claim experience. If an insurer uses
38 which a formal, then the policyholder can predict the change in premium.
39 On the other hand, Venter (1990) points out that in some cases, the loss
40 of accuracy makes a linear formula undesirable.

41 We apply decision theory to develop a credibility formula that minimizes
42 a loss function that is linear combination of a squared-error term and a
43 second-derivative term. The squared-term measures the accuracy of the
44 estimator, while the second-derivative term constrains the estimator close
45 to be linear. Young (1997) has completed analytical solution by using
46 Green's function or Sturm's-Liouville eigenvalues/eigenvector. We
47 demonstrate in this note, solving the bvp's using symbolic computing
48 based on Cabada *et al.* (2012), inspired by Drekić *et al.* (2004) works of
49 which solves ruin problems by means of symbolic computation.

50 **2 Buhlmann Model**

51 We consider a model-based approach to the solution of the credibility
52 problem. This approach, referred to as greatest accuracy credibility
53 theory, is the outgrowth of a classic 1967 paper by Buhlmann. Suppose
54 that a policyholder has experienced X_i claims or losses in past experience
55 period i , where $i \in \{1, 2, 3, \dots, n\}$. Another view is that X_i is the experience
56 from the i^{th} policy in a group or from the i^{th} member of a particular class
57 in rating a scheme. Assume that the claims of a risk, or policyholder, per
58 policy period (usually one year), is a random variable $X|\Theta=\theta$, or more
59 simply, $X|\theta$. A value θ of Θ completely specifies the distribution of the

60 claims of the corresponding risk. Assume that the value θ is fixed for a
61 given risk, although it is generally unknown. For a fixed value $\Theta=\theta$,
62 assume that the random variables $X_i|\theta$, $i=1,\dots,n$, are independently and
63 identically distributed according to a probability density function $f(x|\theta)$.
64 Also assume that the random variables $X|\theta$ are independent for different
65 values θ of Θ .

66 Suppose the insurer has n years of claim experience for a policyholder:
67 $x = \langle x_1, \dots, x_n \rangle \in [a, b]^n$. In this work, we consider credibility formulas
68 that are not necessarily linear, denoted by $d(x)$, in which d is a real-
69 valued function on $[a, b]^n$. We refer to $d(x)$ as the premium for year $n+1$,
70 and we are not loading for administrative expenses.

71 To estimate the future claims of a risk, $X_{n+1}|\theta$, with unknown θ ,
72 Buhlmann (1967, 1970) minimizes the expected value of the squared-
73 error loss function $(d(x)-X_{n+1})^2$. One takes the expectation with respect to
74 the joint distribution of $X_1, X_2, \dots, X_n$ and Θ . The resulting premium $d(x)$ is
75 the expected value of $X_{n+1}|x$, called the predictive mean. For simplicity,
76 we write $\mu(x)=E[X_{n+1}|\theta]$ to represent the predictive mean. By restricting
77 the form of the renewal premium $d(x)$ to be a linear combination of the
78 claim experience, $\{x_1, x_2, \dots, x_n\}$, and by using the same squared-error loss
79 function, Buhlmann (1967, 1970) obtains the following credibility
80 formula $d(x) = (1 - Z)E(X) + Z\bar{x}$. For details see Klugman (2012).

81 **3 Loss Function from Spline Theory**

82 Young (1997) develop a credibility formula that allows an actuary to
83 balance the (sometimes) conflicting goals of accuracy (as measured by a

squared-error term) and of linearity (as measured by a second-derivative term). We want a linearity penalty with null space equal to the set of functions that are linear in the data, with no interactions. One such penalty function is

$$Lin(d) = \sum_{i=1}^n \left(\frac{\partial^2 d}{\partial x_i^2} \right)^2 + c \sum_{i < j} \left(\frac{\partial^2 d}{\partial x_j \partial x_i} \right)^2$$

in which c is a positive constant that controls the relative weight we place on the penalty for interaction terms. Therefore, propose finding a function $d: [a, b]^n \rightarrow \mathbf{R}$ to minimize the combination

$$E[(d(x) - X_{n+1})^2] + hE[Lin(d)] \quad (1)$$

in which h is a positive constant that controls the degree to which d is close to linear. By simplification, finding d to minimize expression (1) is equivalent to finding d to minimize:

$$E[(d(x) - \mu(x))^2] + hE[Lin(d)] \quad (2)$$

in which we take the expected value with respect to the joint distribution of $X_1, X_2, \dots, X_n$.

In this work, we focus on the case for which $n=1$. In this instance, finding d to minimize expression (2) is equivalent to finding $d: [a, b] \rightarrow \mathbf{R}$ to minimize the following:

$$E[(d(x) - \mu(x))^2] + hE[(d''(x))^2] \quad (3)$$

in which $\mu(x)$ is the predictive mean of X_2 given $X_1=x$, and we take the expectation with respect to the marginal distribution of X_1 . If we let h approach 0, then the optimal d is the predictive mean. At the other extreme, if we let h approach ∞ , then the optimal d converges to the

105 Buhlmann linear credibility estimator (Young 1997). Alternatively,
106 Young (2000) proposes the following, finding d to minimize expression
107 the following:

$$E \left[(d(x) - \mu(x))^2 \right] + hE \left[(d'(x))^2 \right] \quad (4)$$

108 which she claimed have better approximation at the end of interval.
109 Furthermore, taking advantages of loss functions (3) and (4), in we
110 propose a new loss function defined by the following:

$$E \left[(d(x) - \mu(x))^2 \right] + h_1 E \left[(d''(x))^2 \right] + h_2 E \left[(d'(x))^2 \right] \quad (5)$$

111 where $h_1 = (1 - \alpha)h$, $h_2 = \alpha h$, $1 \geq \alpha \geq 0$, a convex combination of
112 penalty function. Unfortunately, the proposed loss function (5) do not
113 have equivalent ode's such as in (3) and (4) cases.

114 **4 Equivalent Boundary-Value Problem**

115
116 As we state in Section 3, our goal is to find a function d that minimizes
117 (3), which for $n=1$ is the following expression:

$$\int_a^b \left[(d(x) - \mu(x))^2 + h(d''(x))^2 \right] f(x) dx$$

118 in which f is the marginal distribution of the claim random variable X .
119 Recall that we assume that the support off is the interval $[a,b]$, a possibly
120 infinite interval. Also assume that f is non-zero on $[a,b]$, except possibly
121 at $x=a$ or b , and that all written derivatives and expected values exist. The
122 minimizer d is the solution of the following fourth-order boundary-value
123 problem (Young 1997):

$$h[f(x)d''(x)]'' + f(x)d(x) = f(x)\mu(x) \quad (6)$$

124 with $d''(a)=d''(b)=d'''(a)=d'''(b)=0$. As for the expression (4), the
125 minimizer d is the solution of the following second-order boundary-value
126 problem (Young 2000):

$$h[d''(x)] + f(x)d(x) = f(x)\mu(x) \quad (7)$$

127 with boundary values

$$128 \quad d'(a) = d'(b) = 0.$$

129 It is unfortunate that expression (5) do not have an equivalent minimize
130 in the forms of boundary value problem. However, in variational
131 formulation the expression (5) can be solved using other means such as
132 using mixed finite element methods which will be pursued in future
133 works.

134 **5 Symbolic Calculation**

135 Young (1997) propose to solve the ordinary differential equation using
136 Green's function. Cabada *et al.* (2012) develop symbolic calculation of
137 Green's function of boundary value problems of ode's in *Mathematica*.

138 **5.1 Regular Case**

139
140 In the following boundary value problem, derived from credibility
141 formula which allows actuaries to balance two things: accuracy and
142 linearity. Here is a case where the boundary value problem for the
143 distribution of marginal (f) of the amount of the claim (X) is not zero in a
144 finite interval $[a,b]$, see expression(6) and (7).

145 We calculating the solution d for the following joint distribution of X and
146 Θ :

$$f(x, \theta) = \frac{2}{(x+2)}, 0 < \theta < 0.5x + 1, 0 < x < 1.$$

147 The marginal distribution of X is uniform on $[0,1]$ and the conditional
148 distribution of $X|\theta$ is given by

$$f(x|\theta) = \begin{cases} \frac{1}{(x+2)\ln 1.5}, & 0 \leq \theta \leq 1 \\ \frac{1}{(x+2)(\ln 1.5 - \ln \theta)}, & 1 \leq \theta \leq 1.5. \end{cases}$$

149 The expected value of $X|\theta$ is, thus,

$$E(X|\theta) = \begin{cases} \frac{1 - 2\ln 1.5}{(x+2)\ln 1.5}, & 0 \leq \theta \leq 1 \\ \frac{3 - 2\theta - 2(\ln 1.5 - \ln \theta)}{(x+2)(\ln 1.5 - \ln \theta)}, & 1 \leq \theta \leq 1.5 \end{cases}$$

150 and the predictive mean is

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

151 The boundary value problem in (6) and (7) for this example can be
152 written respectively as

$$\begin{cases} d''''(x) = \frac{1}{0.0025} [\mu(x) - d(x)] \\ d''(0) = d''(1) = d'''(0) = d'''(1) = 0 \end{cases}, \quad (8)$$

153 and

$$\begin{cases} d''(x) = \frac{1}{0.0025} [\mu(x) - d(x)] \\ d'(0) = d'(1) = 0. \end{cases} \quad (9)$$

154 The boundary value problem in (8) and (9) is solved using Green's
155 function calculated symbolically using *Mathematica* package (Cabada *et*
156 *al.* 2012, Kurnia *et al.* 2017).

157 The Green's function for the second-order boundary value problem (9) is
158 given by

$$159 \quad G_1(s, t) = \begin{cases} \frac{1}{20} \csc(20) \cos(20s) \cos(20(t-1)) & 0 \leq s \leq t \leq 1 \\ \frac{1}{20} \csc(20) \cos(20(s-1)) \cos(20t) & 0 < t < s \leq 1 \end{cases}.$$

160 So the solution of the previous boundary-value problem is

$$161 \quad d(x) = \int_0^1 G_1(t, s) \frac{\mu(x)}{0.0025} ds.$$

162 Graph of $\mu(x)$ and $d(x)$ for the second-order problem is shown below.

163
164 Figure 1. Graph of $\mu(x)$ and $d(x)$ for the second-order problem

165 The Green's function for the fourth-order boundary value problem (9) is
166 given by

$$167 \quad G_2(t, s) = \begin{cases} h(t, s) & 0 \leq s \leq t \leq 1 \\ h(s, t) & 0 < t < s \leq 1 \end{cases}$$

168 where

$$\begin{aligned}
 h(t, s) = & \frac{1}{80\sqrt{10}} \left(e^{\sqrt{10}(s-t)} \sin(\sqrt{10}(t-s)) + e^{\sqrt{10}(t-s)} \sin(\sqrt{10}(t-s)) + e^{\sqrt{10}(s-t)} \cos(\sqrt{10}(t-s)) - \right. \\
 & e^{\sqrt{10}(t-s)} \cos(\sqrt{10}(t-s)) + \frac{1}{-2 + \cos(2\sqrt{10}) + \cosh(2\sqrt{10})} \left((1-i)e^{-\sqrt{10}(s-1)} \left(\sin((1+i)\sqrt{10}t) + \right. \right. \\
 & \sinh((1+i)\sqrt{10}t) \left((e^{2\sqrt{10}(s-1)} + 1) \sin(\sqrt{10}(s-1)) \left(\cos(\sqrt{10}) \sinh(\sqrt{10}) + \sin(\sqrt{10}) \cosh(\sqrt{10}) \right) + \right. \\
 & \cos(\sqrt{10}(s-1)) \left(e^{2\sqrt{10}(s-1)} \left(\sin(\sqrt{10}) \cosh(\sqrt{10}) + \sinh(\sqrt{10}) (2 \sin(\sqrt{10}) + \cos(\sqrt{10})) \right) + \right. \\
 & \sin(\sqrt{10}) (-\cosh(\sqrt{10})) - \sinh(\sqrt{10}) (\cos(\sqrt{10}) - 2 \sin(\sqrt{10})) \left. \right) \left. \right) \left. \right) + \\
 & \frac{1}{-2 + \cos(2\sqrt{10}) + \cosh(2\sqrt{10})} \left(4e^{-\sqrt{10}(s-1)} \sin(\sqrt{10}) \sinh(\sqrt{10}) \cos(\sqrt{10}t) \cosh(\sqrt{10}t) \right. \\
 & \left((e^{2\sqrt{10}(s-1)} + 1) \sin(\sqrt{10}(1-s)) + \cos(\sqrt{10}(s-1)) \left(e^{2\sqrt{10}(s-1)} (-1 + \cot(\sqrt{10}) - \coth(\sqrt{10})) + \right. \right. \\
 & \left. \left. 1 + \cot(\sqrt{10}) - \coth(\sqrt{10}) \right) \right) \left. \right).
 \end{aligned}$$

169

170 So the solution of the previous boundary-value problem is

$$d(x) = \int_0^1 G_2(t, s) \frac{\mu(x)}{0.0025} ds.$$

171

172 Graph of $\mu(x)$ and $d(x)$ for the fourth-order problem is shown below.

173

174 Figure 2. Graph of $\mu(x)$ and $d(x)$ for the fourth-order problem

175 Furthermore, we offer the conjectured solution of a linear convex
176 combination of second-order and fourth-order problem (see expression
177 (5)). It is shown by the graph below.

Figure 3. Graph of $\mu(x)$ and $d(x)$ for the linear convex combination of second-order and fourth-order problem

6 Summary

In this paper, we present a way to solve symbolically 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. The equivalent boundary value problems of the full nonlinear credibility formula is proposed and solved by Green's function computation via symbolic calculation.

7 Acknowledgment

This work funded by PUPT-IPB under contract no: 079/SP2H/LT/DRPM/II/2016.

8 References

- [1] Buhlmann H. 1967. Experience Rating and Credibility. ASTIN Bulletin. 4: 199-207.
- [2] Cabada A, Cid JA, Maquez-Villamarin B. 2012. Computation of Green's functions for boundary value problems with Mathematica. Applied Mathematics and Computation. 219(4): 1919-1936.

- 198 [3] Drekić S, Stafford JE, Willmot GE. 2004. Symbolic calculation of the
199 moments of the time of ruin. *Insurance: Mathematics and Economics*.
200 34(1): 109-120.
- 201 [4] Klugman SA, Panjer HH, Willmot GE. 2012. *Loss Models from Data*
202 *to Decisions*. 4th Ed. New Jersey: John Wiley and Son, Inc.
- 203 [5] Kurnia, R., A.D.Garnadi, S. Nurdianti, 2017, Symbolic Green's
204 Function Methods to Solve A Boundary Value Problem., *Jurnal*
205 *Matematika dan Aplikasinya*, v 16 (2), ??-?? (In Indonesian)
- 206 [6] Venter G. 1990. *Credibility. Foundations of casualty actuarial*
207 *science*. Casualty Actuarial Society, New York.
- 208 [7] Young VR. 1997. Credibility Using a Loss Function from Spline
209 Theory. *Scand. Actuarial J.* 2: 160-185.
- 210 [8] Young VR. 1998. Credibility using a loss function from spline
211 theory: Parametric models with a one dimensional sufficient statistic.
212 *North American Actuarial Journal*. 2(1): 101-111.
- 213 [9] Young VR. 2000. Credibility using semiparametric models and a loss
214 function with constancy penalty. *Insurance: Mathematics and*
215 *Economics*. 26(2): 151-156.