Credibility Modeling with Applications

by

Tatiana Khapaeva

Thesis presented in a partial fulfilment of the requirements

for the degree of

Master of Science (M.Sc.) in Computational Sciences

School of Graduate Studies

Laurentian University

Sudbury, Ontario

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE

Laurentian Université/Université Laurentienne
School of Graduate Studies/École des études supérieures

Title of Thesis Titre de la thèse	CREDIBILITY MODELING WITH APPLICATIONS		
Name of Candidate Nom du candidat	Khapaeva, Tatiana		
Degree Diplôme	Master of Science		
Department/Program Département/Programme	Computational Sciences	Date of Defence Date de la soutenance	April 22, 2014

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Peter Adamic
(Supervisor/Directeur de thèse)

Prof. Michael Herman
(Committee member/Membre du comité)

Dr. Ratvinder Grewal
(Committee member/Membre du comité)

Dr. Natalia Stepanova
(External Examiner/Examinatrice externe)

Approved for the School of Graduate Studies
Approuvé pour l'École des études supérieures
Dr. David Lesbarrères
M. David Lesbarrères
Director, School of Graduate Studies
Directeur, École des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

The purpose of this thesis is to show how the theory and practice of credibility can benefit statistical modeling. The task was, fundamentally, to derive models that could provide the best estimate of the losses for any given class and also to assess the variability of the losses, both from a class perspective as well as from an aggregate perspective. The model fitting and diagnostic tests will be carried out using standard statistical packages. A case study that predicts the number of deaths due to cancer is considered, utilizing data furnished by the Colorado Department of Public Health and Environment. Several credibility models are used, including Bayesian, Bühlmann and Bühlmann-Straub approaches, which are useful in a wide range of actuarial applications.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The actuary uses observations of events that happened in the past to forecast future events or costs. For example, data that was collected over several years about the average cost to insure a selected risk, sometimes referred to as a policyholder or insured, may be used to estimate the expected cost to insure the same risk in future years. Because insured losses arise from random occurrences, however, the actual costs of paying insurance losses in past years may be a poor estimator of future costs. The insurer is then forced to answer the following questions: how credible is the policyholder's own experience? And what weight should be given to the class rate?

Credibility theory began with papers by Mowbray (1914) and Whitney (1918). In those papers, the emphasis was on deriving a premium which was a balance between the experience of an individual risk and a class of risks. Bühlmann (1967) showed how a credibility formula can be derived in a distribution-free way, using a least-squares criterion. Since then, a number of papers have shown how this approach can be extended. For example, Bühlmann and Straub (1970), Hachemeister (1975), de Vylder(1976,1986), Jewell (1974, 1975), and more recently Klugman (1987, 1998), saw the development of

linear unbiased estimators to the theoretically exact Bayesian estimates. The credibility models espoused in the Klugman paper have a prior or collateral information that can be weighted with current observations. The goal of this approach is the minimization of the square of the error between the estimate and the true expected value of the quantity being estimated.

Chapter 2

The Concept of Credibility

Credibility Theory is a set of quantitative tools which allows an insurer to perform prospective experience rating (adjust future premiums based on past experience) on a risk or group of risks. Credibility provides tools to deal with the randomness of data that is used for predicting future events or costs. For example, an insurance company uses past loss information of an insured or group of insureds to estimate the cost to provide future insurance coverage. But, insurance losses arise from random occurrences. The average annual cost of paying insurance losses in the past few years may be a poor estimate of next years costs. The expected accuracy of this estimate is a function of the variability in the losses. This data by itself may not be acceptable for calculating insurance rates. Rather than relying solely on recent observations, better estimates may be obtained by combining this data with other information. For example, suppose that recent experience indicates that Carpenters should be charged a rate of \$5 (per \$100 of payroll) for workers compensation insurance. Assume that the current rate is \$10. What should the new rate be? Should it be \$5, \$10, or somewhere in between? Credibility is used to weight together these two estimates.

The basic formula for calculating the credibility weighted estimate is:

$$Estimate = Z \times [Observation] + (1 - Z)[Other\ Information], 0 \leq Z \leq 1.$$

Z represents the credibility assigned to the observation. $1 - Z$ is generally referred to as the complement of credibility. If the body of observed data is large and not likely to vary much from one period to another, then Z will be closer to one. On the other hand, if the observation consists of limited data, then Z will be closer to zero and more weight will be given to other information.

The current rate of \$10 in the above example is the "Other Information". It represents an estimate or prior hypothesis of a rate to charge in the absence of the recent experience. As recent experience becomes available, then an updated estimate combining the recent experience and the prior hypothesis can be calculated. Thus, the use of credibility involves a linear estimate of the true expectation derived as a result of a compromise between observation and the prior hypothesis. The Carpenters rate for workers compensation insurance is $Z \cdot \$5 + (1 - Z) \cdot \$10$ under this model.

The following is another example demonstrating how credibility can help produce better estimates.

Example 1:

In a large population of automobile drivers, the average driver has one accident every five years or, equivalently, an annual frequency of .20 accidents per year. A driver selected randomly from the population had three accidents during the last five years for

a frequency of .60 accidents per year. What will be the estimate of the expected future frequency rate for this driver? Is it .20, .60, or something in between?

Solution:

If we had no information about the driver other than that he came from the population, we should go with the .20. However, we know that the driver's observed frequency was .60. Should this be our estimate for his future accident frequency? Probably not. There is a correlation between prior accident frequency and future accident frequency, but they are not perfectly correlated. Accidents occur randomly and even good drivers with low expected accident frequencies will have accidents. On the other hand, bad drivers can go several years without an accident. A better answer than either .20 or .60 is most likely something in between: this driver's Expected Future Accident Frequency is $F = Z \cdot 0.60 + (1 - Z) \cdot 0.20$. ■

Consider a risk that is a member of a particular class of risks. Classes are groupings of risks with similar risk characteristics, and though similar, each risk is still unique and not quite the same as other risks in the class. In class rating, the insurance premium charged to each risk in a class is derived from a rate common to the class. Class rating is often supplemented with experience rating so that the insurance premium for an individual risk is based on both the class rate and actual past loss experience for the risk. The important question in this case is: How much should the class rate be modified by experience rating? That is, how much credibility should be given to the actual experience of the individual risk?

Let the probability function (pf) p_k denote the probability that exactly "k" events occur. Let N be a random variable representing the number of such events. Then

$$p_k = Pr(N = k),\ k = 0, 1, 2, ...$$

When it is unclear, or when the random variable may be continuous, discrete, or a mixture of the two, the term probability function and abbreviation pf will be used. The term probability density function and the abbreviation pdf will be used only when the random variable is known to be continuous.

Suppose that X and Y are two random variables with joint probability function (pf) or probability density function (pdf) $f_{X,Y}(x, y)$ and marginal pfs $f_X(x)$ and $f_Y(y)$, respectively. The conditional pf of X given that $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x|y)}{f_Y(y)}.$$

If X and Y are discrete random variables, then $f_{X|Y}(x|y)$ is the conditional probability of the event $X = x$ under the hypothesis that $Y = y$. If X and Y are continuous, then $f_{X|Y}(x|y)$ may be interpreted as a definition. When X and Y are independent random variables,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

and in this case $f_{X|Y}(x|y) = f_X(x)$, and we observe that the conditional and marginal distributions of X are identical. Also, the main formula may be rewritten as

$$f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y)$$

demonstrating that joint distributions may be constructed from products of conditional

and marginal distributions. Since the marginal distribution of X may be obtained by integrating (or summing) y out of the joint distribution,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

or

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy.$$

Example 2:

Suppose that, given $\Theta = \theta$, X is Poisson distributed with mean θ, that is,

$$f_{X|\Theta}(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, 2, ...,$$

and Θ is gamma distributed with pdf

$$f_\Theta(\theta) = \frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}, \quad \theta > 0,$$

where $\beta > 0$, and $\alpha > 0$ are parameters. Determine the marginal pf of X.

Solution:

The marginal pf of X is

$$f_X(x) = \int_0^\infty \frac{\theta^x e^{-\theta}}{x!}\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}d\theta = \frac{1}{x!\Gamma(\alpha)\beta^\alpha}\int_0^\infty \theta^{\alpha+x-1}e^{-\theta/[\beta/(1+\beta)]}d\theta =$$

$$\frac{\Gamma(\alpha+x)[\beta/(1+\beta)]^{\alpha+x}}{x!\Gamma(\alpha)\beta^\alpha}\int_0^\infty \frac{\theta^{\alpha+x-1}e^{-\theta/[\beta/(1+\beta)]}}{\Gamma(\alpha+x)[\beta/(1+\beta)]^{\alpha+x}}d\theta.$$

The integral in the above expression is that of a gamma pdf with β replaced by $\beta/(1+\beta)$ and α replaced by $\alpha + x$. Hence the integral is 1 and so

$$f_X(x) = \frac{\Gamma(\alpha+x)}{x!\Gamma(\alpha)}\left(\frac{1}{1+\beta}\right)^\alpha\left(\frac{\beta}{1+\beta}\right)^x = \binom{\alpha+x-1}{x}p^x(1-p)^\alpha, \quad x = 0, 1, 2, ...$$

where $p = \beta/(1+\beta)$. This is the pf of the negative binomial distribution. ■

The roles of X and Y can be interchanged, yielding

$$f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$$

since both sides of this equation equal the joint distribution of X and Y. Division by $f_Y(y)$ yields Bayes' theorem, namely

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}.$$

Again, assume that X and Y are two random variables and the conditional pf of X given that $Y = y$ is $f_{X|Y}(x, y)$. Then, this is a valid probability distribution, and it's mean is denoted by

$$E(X|Y = y) = \int x f_{X|Y}(x|y)dx,$$

with the integral replaced by a sum in the discrete case. Clearly, this is a function of y, and it is often of interest to view this conditional expectation as a random variable obtained by replacing y by Y on the right-hand side. Thus we can write $E(X|Y)$ instead on the left-hand side, and so $E(X|Y)$ is itself a random variable since it is a function of the random variable Y. The expectation of $E(X|Y)$ is given by $E[E(X|Y)] = E(X)$. This is because

$$E[E(X|Y)] = \int_{\mathbb{R}} E(X|Y = y)f_Y(y)dy = \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X|Y}(x|y)dx f_Y(y)dy =$$

$$\int_{\mathbb{R}} x \int_{\mathbb{R}} f_{X|Y}(x|y)f_Y(y)dydx = \int_{\mathbb{R}} x f_X(x)dx = E(X).$$

The variance of this conditional distribution is

$$Var(X|Y) = E(X^2|Y) - [E(X|Y)]^2.$$

Thus,

$$E[Var(X|Y)] = E[E(X^2|Y)] - [E(X|Y)]^2 =$$

$$E[E(X^2|Y)] - E[E(X|Y)]^2 = E(X^2) - E[E(X|Y)]^2.$$

We can also obtain,

$$Var[E(X|Y)] = E[E(X|Y)]^2 - E[E(X|Y)]^2 = E[E(X|Y)]^2 - [E(X)]^2.$$

Thus,

$$E[Var(X|Y)] + Var[E(X|Y)] = E(X^2) - E[E(X|Y)]^2 + E[E(X|Y)]^2 - [E(X)]^2 =$$

$$E(X^2) - [E(X)]^2 = Var(X)$$

Thus, we have established the important formula

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)].$$

This formula states that the variance of X is composed of the sum of two parts: the mean of the conditional variance plus the variance of the conditional mean.

Chapter 3

The Bayesian Credibility Model

3.1 Bayesian estimation

The Bayesian approach assumes that only the data count and it is the population that is variable. For parameter estimation the following definitions describe the process and then Bayes' theorem provides the solution.

Definition 3.1 The *prior distribution* is a probability distribution over the space of possible parameter values. It is denoted $\pi(\theta)$ and represents the opinion concerning the relative chances that various values of θ are the true value.

Definition 3.2 An *improper prior distribution* is one for which the probabilities (or pdf) are non-negative, but their sum (or integral) is infinite.

Definition 3.3 The *model distribution* is the probability distribution for the data as collected, given a particular value for the parameter. Its pdf is denoted $f_{X|\Theta}(x|\theta)$, where vector notation for X is used to remind us that all the data appears here. Also note that this is identical to the likelihood function and so that name may also be used at times.

Definition 3.4 The *joint distribution* has pdf

$$f_{X,\Theta}(x,\theta) = f_{X|\Theta}(x|\theta)\pi(\theta).$$

Definition 3.5 The *marginal distribution* of X has pdf

$$f_X(x) = \int_{\mathbb{R}} f_{X|\Theta}(x|\theta)\pi(\theta)d\theta.$$

If the prior distribution is discrete, the integral should be replace by a sum.

Definition 3.6 The *posterior distribution* is the conditional probability distribution of the parameters given the observed data. It is denoted $\pi_{\Theta|X}(\theta|x)$.

Definition 3.7 The *predictive distribution* is the conditional probability distribution of a new observation y given the data x. It is denoted $f_{Y|X}(y|x)$.

These last two items are the key output of a Bayesian analysis. The posterior distribution tells us how our opinion about the parameter has changed once we have observed the data. The predictive distribution tells us what the next observation might look like given the information contained in the data (as well as, implicitly, our prior opinion). Bayes' theorem tells us how to compute the posterior distribution.

Theorem 3.1 The posterior distribution can be computed as

$$\pi_{\Theta|X}(\theta|x) = \frac{f_{X|\theta}(x|\theta)\pi(\theta)}{\int_{\mathbb{R}} f_{X|\theta}(x|\theta)\pi(\theta)d\theta}$$

while the predictive distribution can be computed as

$$f_{Y|X}(y|x) = \int_{\mathbb{R}} f_{Y|\Theta}(y|\theta)\pi_{\Theta|X}(\theta|x)d\theta,$$

where $f_{Y|\Theta}(y|\theta)$ is the pdf of the new observation, given the parameter value. In both formulas the integrals are replaced by sums for a discrete distribution[1].

Example 3:

The following amounts were paid on a hospital liability policy:

125 132 141 107 133 319 126 104 145 223

The amount of a single payment has the single-parameter Pareto distribution with $\theta = 100$ and α unknown. The prior distribution has the gamma distribution with $\alpha = 2$ and $\theta = 1$. Determine all of the relevant Bayesian quantities.

Solution:

The prior density has a gamma distribution and is

$$\pi(\alpha) = \alpha e^{-\alpha}, \quad \alpha > 0,$$

while the model is (evaluated at the data points)

$$f_{X|A}(x|\alpha) = \frac{\alpha^{10}(100)^{10\alpha}}{\prod_{j=1}^{10} x_j^{\alpha+1}} = \alpha^{10} e^{-3.801121\alpha - 49.852823}.$$

The joint density of X and A is (again evaluated at the data points)

$$f_{X|A}(x|\alpha) = \alpha^{11} e^{-4.801121\alpha - 49.852823}.$$

The posterior distribution of α is

$$\pi_{A|X}(\alpha|x) = \frac{\alpha^{11} e^{-4.801121\alpha - 49.852823}}{\int_0^\infty \alpha^{11} e^{-4.801121\alpha - 49.852823} d\alpha} = \frac{\alpha^{11} e^{-4.801121\alpha}}{(11!)(1/4.801121)^{12}}.$$

[1]Klugman, S., et al.(1998), 404.

It is clear that the result must be a probability distribution, then the denominator is just the appropriate normalizing constant. A look at the numerator reveals that we have a gamma distribution with $\alpha = 12$ and $\theta = 1/4.801121$. The predictive distribution is

$$f_{Y|X}(y|x) = \int_0^\infty \frac{\alpha 100^\alpha}{y^{\alpha+1}} \frac{\alpha^{11} e^{-4.801121\alpha}}{(11!)(1/4.801121)^{12}} d\alpha$$

$$= \frac{1}{y(11!)(1/4.801121)^{12}} \int_0^\infty \alpha^{12} e^{-(0.195951+\,log\,y)\alpha} d\alpha$$

$$= \frac{1}{y(11!)(1/4.801121)^{12}} \frac{(12!)}{(0.195951+log\,y)^{13}}$$

$$= \frac{12(4.801121)^{12}}{y(0.195951+log\,y)^{13}}, \quad y > 100. \ \blacksquare$$

In a typical Bayesian analysis the pf of Θ represents the subjective prior opinion about the unknown parameter. In the credibility setting this may still be the case, but Θ may also represent a real, though unobservable, random variable. For example, θ may indicate an automobile driver's propensity to have a claim and $\pi(\theta)$ may describe how that propensity is distributed throughout the population of insured drivers. With no additional information, $\pi(\theta)$ represents our prior opinion about a randomly selected driver's parameter.

The joint distribution of $X_1, ..., X_n$, Θ is obtained by first conditioning on Θ, and it is thus given by the likelihood multiplied by the prior density, that is,

$$f_{X,\Theta}(x, \theta) = [\prod_{j=1}^n f_{X_j|\Theta}(x_j|\theta)]\pi(\theta).$$

The marginal distribution of $X_1, ..., X_n$ is obtained by integrating θ (summing if $\pi(\theta)$ is discrete) out of the joint density of $X_1, ..., X_n, \Theta$, that is

$$f_X(X) = \int_{\mathbb{R}} [\prod_{j=1}^n f_{X_j|\Theta}(x_j|\theta)]\pi(\theta)d\theta.$$

The information about Θ "posterior" to the observation of $X = (X_1, ..., X_n)'$ is summarized in the posterior distribution of Θ. This is simply the conditional density of Θ given that X equals $x = (x_1, ..., x_n)'$, which we express notationally as $\Theta|X = x$ or simply as $\Theta|X$ or $\Theta|x$, and this is also the ratio of the joint density of $X_1, ..., X_n, \Theta$ to the marginal density of $X_1, ..., X_n$. In other words, the posterior density is

$$\pi_{\Theta|X}(\theta|x) = \frac{[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)]\pi(\theta)}{\int_{\mathbb{R}}[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)]\pi(\theta)d\theta}.$$

From a practical viewpoint, the denominator does not depend on θ and simply serves as a normalizing constant. Thus, as a function of $\theta, \pi_{\Theta|X}(\theta|x)$ is proportional to

$$[\prod_{j=1}^{n} f_{X_j|\Theta}(x_j|\theta)]\pi(\theta)$$

and if the form of the terms involving θ is recognized as belonging to a particular distribution, then it is not necessary to evaluate the denominator, but only to identify the appropriate normalizing constant. A point estimate of θ derived from $\pi_{\Theta|X}(\theta|x)$ requires the selection of a loss function, and the choice of squared error loss results in the posterior mean:

$$E(\Theta|X = x) = \int_{\mathbb{R}} \theta\pi_{\Theta|X}(\theta|x)d\theta.$$

3.2 Derivation of the Bayesian Credibility Model

The framework that will be employed to develop the credibility models in this Thesis will be Bayesian. Hence, it will be assumed that there exists some prior opinion regarding the risk characteristics of the population, which will be described via the random variable Θ. The prior probability density function will be denoted as $\pi_{\Theta}(\theta)$, a function with one

parameter, θ. Also, it will be assumed that $X_i|\theta$ and $X_j|\theta$ are independent and identically distributed $\forall\, i \neq j$, where in the present context the subscript designates duration. Other preliminary Bayesian relationships that are needed are given as follows:

- $\mathbf{X}$ will represent a vector of either loss amounts or claim frequencies
- The conditional distribution of $\mathbf{X}$ given $\Theta = \theta$ is defined as,

$$f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) = f(x_1, x_2, ..., x_n|\theta) = \prod_i^n f_{X_i|\Theta}(x_i|\theta)$$

- The joint pdf of $\mathbf{X}$ and Θ is : $f_{\mathbf{X},\Theta}(\mathbf{x}, \theta)$= $f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \cdot \pi(\theta)$
- The marginal distribution of $\mathbf{X}$ is: $f_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}} f_{\mathbf{X},\Theta}(\mathbf{x}, \theta) d\theta$
- The posterior distribution of $\Theta|\mathbf{X}$ is $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{f_{\mathbf{X},\Theta}(\mathbf{x},\theta)}{f_{\mathbf{X}}(\mathbf{x})}$

The significance of the posterior distribution is that it updates the *a priori* probability statements regarding Θ, based on the observed values $(x_1, x_2, ..., x_n)$. From this, a revised predictive distribution can be constructed that incorporates the observed x_i values. The predictive distribution is derived as follows. First, let X_{n+1} denote the losses in the upcoming year, $(n + 1)$, which have not yet been observed. Also, let the expected value of losses for year i given Θ be expressed as $E(X_i|\Theta) = \mu_i(\Theta)$. In particular, $E(X_{n+1}|\Theta) = \mu_{n+1}(\Theta)$, where $\mu_{n+1}(\Theta)$ represents the hypothetical mean for next year. Now, the conditional distribution of X_{n+1} given $\mathbf{X} = [x_1, x_2, ..., x_n]$ may be expressed as,

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \frac{f(x_1,...,x_n,x_{n+1})}{f(x_1,...,x_n)} = \frac{f(x_1,...,x_n,x_{n+1})}{f_{\mathbf{X}}(\mathbf{x})}.$$

Furthermore,

$$\left[\prod_i^{n+1} f_{X_i|\Theta}(x_i|\theta)\right] \pi(\theta) = f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \cdot \left[\prod_i^n f_{X_i|\Theta}(x_i|\theta)\right] \pi(\theta)$$

$$= f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) \cdot f_{\mathbf{X}}(\mathbf{x}).$$

Then,

$$f(x_1, ..., x_n, x_{n+1}) = \int_{\mathbb{R}} \left[\prod_i^{n+1} f_{X_i|\Theta}(x_i|\theta)\right] \pi(\theta) d\theta = \int_{\mathbb{R}} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) \cdot f_{\mathbf{X}}(\mathbf{x}) d\theta.$$

Finally, we can express the predictive distribution as,

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \int_{\mathbb{R}} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) \cdot f_{\mathbf{X}}(\mathbf{x}) d\theta \ / \ f(x_1, ..., x_n) \ , \ \text{ or,}$$

$$f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) = \int_{\mathbb{R}} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta.$$

Then, the Bayesian Premium (or the expected value of the predictive distribution, *a posteriori*) may be obtained as follows:

$$\begin{aligned}
E[X_{n+1}|\mathbf{X} = \mathbf{x}] &= \int_{\mathbb{R}} x_{n+1} f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x}) dx_{n+1} \\
&= \int_{\mathbb{R}} x_{n+1} [\int f_{X_{n+1}|\Theta}(x_{n+1}|\theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta] dx_{n+1} \\
&= \int_{\mathbb{R}} [\int x_{n+1} f_{X_{n+1}|\Theta}(x_{n+1}|\theta) dx_{n+1}] \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta \\
&= \int_{\mathbb{R}} \mu_{n+1}(\Theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) d\theta \\
&= \sum_{\Theta} \mu_{n+1}(\Theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}), \text{ in the discrete case.}
\end{aligned}$$

Thus, the Bayesian Premium is the integral (or summation in the discrete setting) over Θ of the hypothetical mean, $\mu_{n+1}(\Theta)$, multiplied by the posterior density function, $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})$. Computationally, this last expression is generally much easier to implement than using the distribution of $f_{X_{n+1}|\mathbf{X}}(x_{n+1}|\mathbf{x})$ directly.

Example 4:

In this random experiment, there is a big bowl and two boxes (Box 1 and Box 2). The bowl consists of a large quantity of balls, 80% of which are white and 20% of which are red. In Box 1, 60% of the balls are labeled 0, 30% are labeled 1 and 10% are labeled 2. In Box 2, 15% of the balls are labeled 0, 35% are labeled 1 and 50% are labeled 2. In the experiment, a ball is selected at random from a bowl. The color of the selected ball from the bowl determines which box to use (if the ball is white, then use Box 1, if red, use Box 2). Then balls are drawn at random from the selected box (Box i) repeatedly with replacement and the values of the series of selected balls are recorded. The value of the first selected ball is X_1, the value of the second selected ball is X_2, and so on.

Suppose that a random person performs this random experiment (we do not know whether he uses Box 1 or Box 2) and that his first ball is a 1 ($X_1 = 1$) and his second ball is a 2 ($X_2 = 2$). What is the predicted value X_3 of the third selected ball?

Solution:

For convenience, we will denote "draw of a white ball from bowl" by $\theta = 1$ and "draw of a red ball from bowl" by $\theta = 2$. Box 1 and Box 2 represent conditional distributions. Bowl is a distribution for the parameter θ. The distribution given the bowl is a probability distribution over the space of all parameter values (the prior distribution). The prior distribution of θ and the conditional distributions of X given θ are restated as follows:

$$\pi_\theta(1) = 0.8,\ \pi_\theta(2) = 0.2$$

$$f_{X|\Theta}(0|\theta = 1) = 0.60,\ f_{X|\Theta}(1|\theta = 1) = 0.30,\ f_{X|\Theta}(2|\theta = 1) = 0.10$$

$$f_{X|\Theta}(0|\theta=2)=0.15,\ f_{X|\Theta}(1|\theta=2)=0.35,\ f_{X|\Theta}(2|\theta=2)=0.50.$$

The following shows the conditional means $E[X|\theta]$ and the unconditional mean $E[X]$:

$$E[X|\theta=1]=0.6(0)+0.3(1)+0.1(2)=0.50,$$

$$E[X|\theta=2]=0.15(0)+0.35(1)+0.5(2)=1.35,$$

$$E[X]=0.8(0.50)+0.2(1.35)=0.67.$$

The Unconditional Distributions:

$$f_X(0)=0.6(0.8)+0.15(0.2)=0.51,$$

$$f_X(1)=0.3(0.8)+0.35(0.2)=0.31,$$

$$f_X(2)=0.1(0.8)+0.50(0.2)=0.18.$$

The Marginal Probabilities:

$$f_{X_1,X_2}(1,2)=0.1(0.3)(0.8)+0.5(0.35)(0.2)=0.059.$$

The Posterior Distributions of θ:

$$\pi_{\Theta|X_1,X_2}(1|1,2)=\tfrac{0.1(0.3)(0.8)}{0.059}=\tfrac{24}{59},$$

$$\pi_{\Theta|X_1,X_2}(2|1,2)=\tfrac{0.5(0.35)(0.2)}{0.059}=\tfrac{35}{59}.$$

The Predictive Distributions of X:

$$f_{X_3|X_1,X_2}(0|1,2)=0.6\tfrac{24}{59}+0.15\tfrac{35}{59}=\tfrac{19.65}{59},$$

$$f_{X_3|X_1,X_2}(1|1,2)=0.3\tfrac{24}{59}+0.35\tfrac{35}{59}=\tfrac{19.45}{59},$$

$$f_{X_3|X_1,X_2}(2|1,2) = 0.1\tfrac{24}{59} + 0.50\tfrac{35}{59} = \tfrac{19.90}{59}.$$

The posterior distribution $\pi_\theta(\cdot|1,2)$ is the conditional probability distribution of the parameter θ given the observed data $X_1 = 1$ and $X_2 = 2$. This is a result of applying Bayes theorem. The predictive distribution $f_{X_3|X_1,X_2}(\cdot|1,2)$ is the conditional probability distribution of a new observation given the past observed data of $X_1 = 1$ and $X_2 = 2$. Since both of these distributions incorporate the past observations, the Bayesian estimate of the next observation is the mean of the predictive distribution.

$$E[X_3|X_1 = 1, X_2 = 2] = 0f_{X_3|X_1,X_2}(0|1,2) + 1f_{X_3|X_1,X_2}(1|1,2) + 2f_{X_3|X_1,X_2}(2|1,2)$$

$$= 0\tfrac{19.65}{59} + 1\tfrac{19.45}{59} + 2\tfrac{19.90}{59} = \tfrac{59.25}{59} = 1.0042372,$$

$$E[X_3|X_1 = 1, X_2 = 2] = E[X|\theta = 1]\pi_{\Theta|X_1,X_2}(1|1,2) + E[X|\theta = 2]\pi_{\Theta|X_1,X_2}(2|1,2)$$

$$= 0.5\tfrac{24}{59} + 1.35\tfrac{35}{59} = \tfrac{59.25}{59} = 1.0042372. \quad \blacksquare$$

3.3 Conjugate Priors in Bayesian Credibility

One very nice property for some combinations of distributions is that the past experience, $\mathbf{X}$, produces a posterior distribution that is the same distribution as the original prior, but the parameters have been updated based on the experience, $\mathbf{X}$. Priors that behave this way are called *conjugate priors*, when combined with an appropriate distribution that follows this property. One of the most important conjugate priors is the Normal-Normal case, and we will consider this case in more details in this section.

Assume that $X_i|\Theta \sim N(\theta, v)$, and $\Theta \sim N(\mu, a)$. Then, the posterior distribution is

$$\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = \frac{f(\mathbf{x},\theta)}{f_{\mathbf{X}}(\mathbf{x})} \propto (2\pi v)^{-\frac{n}{2}} exp[-\frac{1}{2v}(x_i-\theta)^2](2\pi a)^{-\frac{n}{2}} exp[-\frac{1}{2a}(\theta-\mu)^2]$$

$$\propto exp[-\frac{1}{2v}\sum_i (x_i-\theta)^2 - \frac{1}{2a}(\theta-\mu)^2]$$

$$= exp[-\frac{1}{2v}(\sum_i x_i^2 - 2\theta\sum_i x_i + n\theta^2) - \frac{1}{2a}(\theta-\mu)^2]$$

$$\propto exp[-\frac{1}{2v}(-2\theta n\bar{x} + n\theta^2) - \frac{1}{2a}(\theta-\mu)^2]$$

$$= exp[\frac{\theta n\bar{x}}{v} - \frac{n\theta^2}{2v} - \frac{\theta^2}{2a} + \frac{\theta\mu}{a}]$$

$$= exp[(-\frac{n}{2v} - \frac{1}{2a})\theta^2 + (\frac{n\bar{x}}{v} + \frac{\mu}{a})\theta].$$

Now, define $\mu_* = [\frac{n\bar{x}}{v} + \frac{\mu}{a}]a_*$, and let $a_* = [\frac{n}{v} + \frac{1}{a}]^{-1}$. Then, substitution into the previous expression yields,

$$= exp[-\frac{1}{2}(\frac{n}{v} + \frac{1}{a})\theta^2 + (\frac{\mu_*}{a_*})\theta]$$

$$\propto exp[-\frac{1}{2a_*}\theta^2 + \frac{\mu_*}{a_*}\theta - \frac{\mu_*^2}{2a_*}]$$

$$= exp[-\frac{1}{2a_*}(\theta^2 - 2\mu_*\theta + \mu_*^2)]$$

$$= exp[-\frac{1}{2a_*}(\theta-\mu_*)^2],$$

which shows that the posterior distribution $\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) \sim N(\mu_*, a_*)$.

The mean of the predictive distribution is,

$$E[X_{n+1}|\mathbf{X}=\mathbf{x}] = \int_{\mathbb{R}} \mu_{n+1}(\Theta)\cdot\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta = \int_{\mathbb{R}} \theta\cdot\pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x})d\theta,$$

since $X_i|\Theta \sim N(\theta, v)$, and thus $E(X_i|\Theta) = \theta$. But this expression is simply the formula for the first raw moment (i.e. the expected value) of the posterior distribution, which was already shown to equal μ_*. Thus, the Bayesian Premium in the normal conjugate

prior case is μ_*. However, it is instructive to rearrange the expression for μ_* in a manner that is more revealing:

$$\mu_* = [\tfrac{n\bar{x}}{v} + \tfrac{\mu}{a}][\tfrac{n}{v} + \tfrac{1}{a}]^{-1} = [\tfrac{n\bar{x}}{v} + \tfrac{\mu}{a}][\tfrac{av}{na+v}] = (\tfrac{n}{v})(\tfrac{av}{na+v})\bar{x} + (\tfrac{1}{a})(\tfrac{av}{na+v})\mu$$

$$= (\tfrac{na}{na+v})\bar{x} + (\tfrac{v}{na+v})\mu = (\tfrac{n}{n+\frac{v}{a}})\bar{x} + (\tfrac{\frac{v}{a}}{n+\frac{v}{a}})\mu.$$

If we let $\frac{v}{a} = k$, then the Bayesian Premium for the Normal-Normal case can be written as,

$$\mu_* = E[X_{n+1}] = (\frac{n}{n+k})\bar{x} + (\frac{k}{n+k})\mu,$$

where n is the sample size, and k is the ratio of the within variance to the between variance.

Chapter 4

The Bühlmann Credibility Model

4.1 Target Shooting Example

In the classic paper by Stephen Philbrick[1] there is an excellent target shooting example that illustrates the ideas of Bühlmann Credibility. Assume there are four marksmen each shooting at his own target. Each marksmans shots are assumed to be distributed around his target, marked by one of the letters A, B, C, and D, with an expected mean equal to the location of his target. Each marksman is shooting at a different target.

If the targets are arranged as in Figure 1, the resulting shots of each marksman would tend to cluster around his own target. The shots of each marksman have been distinguished by a different symbol. So for example the shots of marksman B are shown as triangles. We see that in some cases one would have a hard time deciding which marksman had made a particular shot if we did not have the convenient labels.

The point E represents the average of the four targets A, B, C, and D. Thus E is the grand mean. If we did not know which marksman was shooting we would estimate that the shot would be at E; the a priori estimate is E.

[1] Philbrick, S. (1981).

Once we observe a shot from an unknown marksman, we could be asked to estimate the location of the next shot from the same marksman. Using Bühlmann Credibility our estimate would be between the observation and the a priori mean of E. The larger the credibility assigned to the observation, the closer the estimate is to the observation. The smaller the credibility assigned to the data, the closer the estimate is to E.

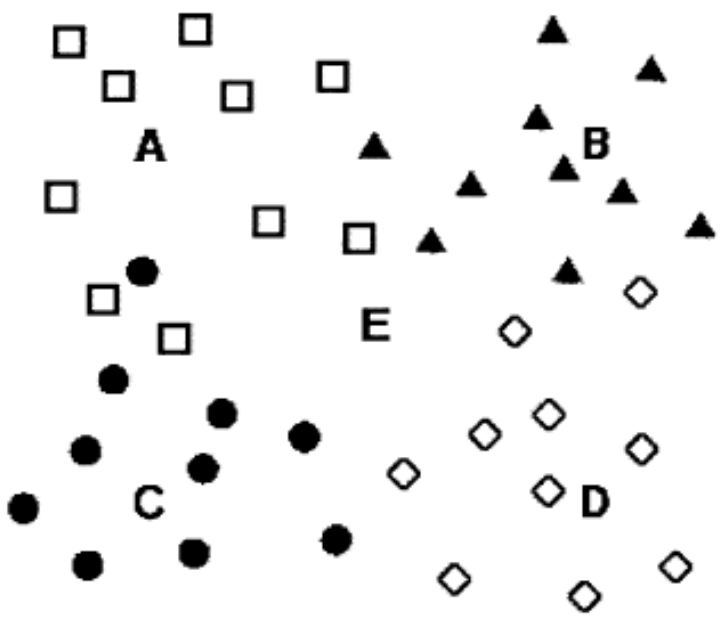

Figure 4.1: Target Shooting Figure 1

There are a number of features of this target shooting example that control how much Bühlmann Credibility is assigned to our observation. We have assumed that the marksmen are not perfect; they do not always hit their target. The amount of spread of their shots around their targets can be measured by the variance. The average spread over the marksmen is the Expected Value of the Process Variance (EPV). The better the marksmen, the smaller the EPV and the more tightly clustered around the targets

the shots will be.

The worse the marksmen, the larger the EPV and the less tightly the shots are spread. The better the marksmen, the more information contained in a shot. The worse the marksmen, the more random noise contained in the observation of the location of a shot. Thus when the marksmen are good, we expect to give more weight to an observation (all other things being equal) than when the marksmen are bad. Thus the better the marksmen, the higher the credibility.

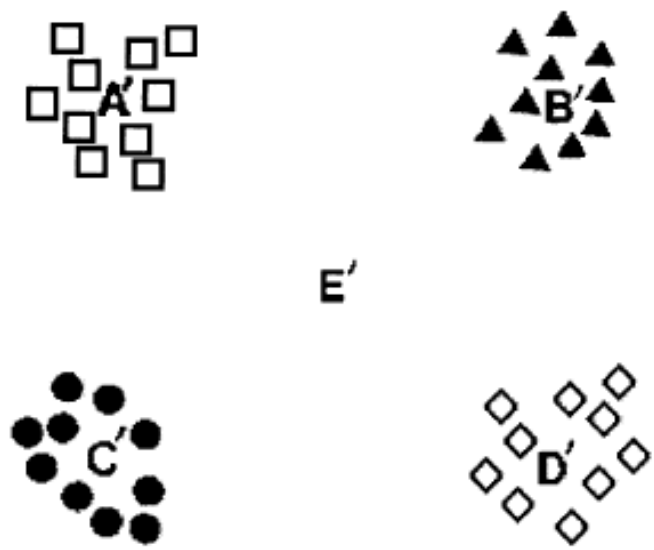

Figure 4.2: Target Shooting Figure 2

The smaller the Expected Value of the Process Variance the larger the credibility. This is illustrated by Figure 2. It is assumed in Figure 2 that each marksman is better than was the case in Figure 1. The EPV is smaller and we assign more credibility to the observation. This makes sense, since in Figure 2 it is a lot easier to tell which marksman is likely to have made a particular shot based solely on its location.

Another feature that determines how much credibility to give an observation is how far apart the four targets are placed. As we move the targets further apart (all other things being equal) it is easier to distinguish the shots of the different marksmen. Each target is a hypothetical mean of one of the marksmens shots. The spread of the targets can be quantified as the Variance of the Hypothetical Means.

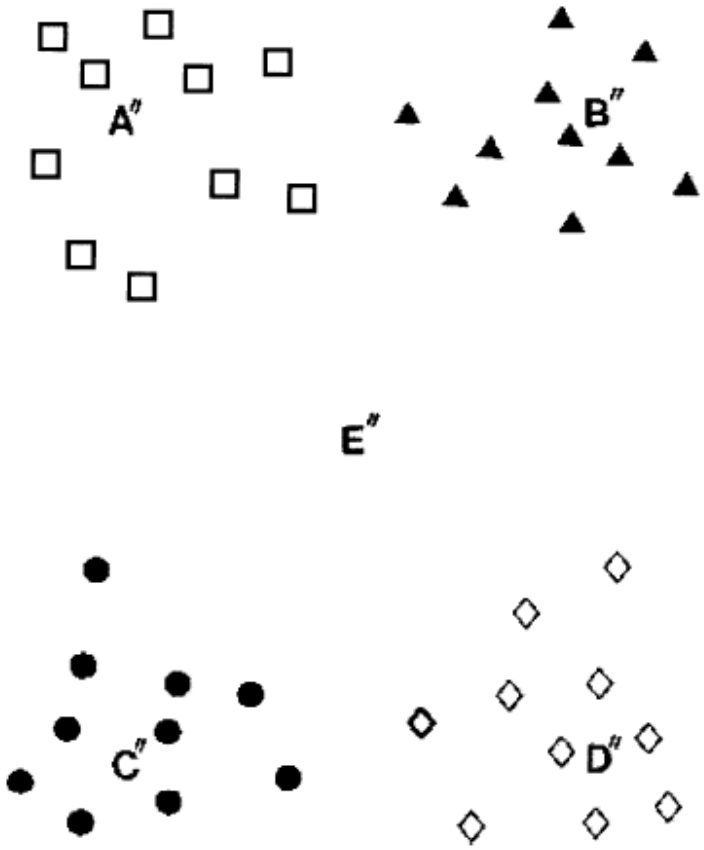

Figure 4.3: Target Shooting Figure 3

As illustrated in Figure 3, the further apart the targets the more credibility we would assign to our observation. The larger the VHM the larger the credibility. It is easier to distinguish which marksman made a shot based solely on its location in Figure 3 than

in Figure 1.

The third feature that one can vary is the number of shots observed from the same unknown marksman. The more shots we observe, the more information we have and thus the more credibility we would assign to the average of the observations.

Each of the three features discussed is reflected in the formula for Bühlmann Credibility

$$Z = \frac{N}{N+K} = \frac{N(VHM)}{N(VHM)+EPV}$$

Thus, as the EPV increases, Z decreases; as VHM increases, Z increases; and as N increases, Z increases.

There are two separate reasons why the observed shots vary. First, the marksmen are not perfect. In other words the Expected Value of the Process Variance is positive. Even if all the targets were in the same place, there would still be a variance in the observed results. This component of the total variance due to the imperfection of the marksmen is quantified by the EPV.

Second, the targets are spread apart. In other words, the Variance of the Hypothetical Means is positive. Even if every marksman were perfect, there would still be a variance in the observed results, when the marksmen shoot at different targets. This component of the total variance due to the spread of the targets is quantified by the VHM.

One needs to understand the distinction between these two sources of variance in the observed results. Also one has to know that the total variance of the observed shots is

a sum of these two components: $Total\ Variance = EPV + VHM$.

4.2 Credibility Parameters

Intuition says that two factors appear important in finding the right balance between class rating and individual risk experience rating:How homogeneous are the classes? If all of the risks in a class are identical and have the same expected value for losses, then why bother with individual experience rating? Just use the class rate. On the other hand, if there is significant variation in the expected outcomes for risks in the class, then relatively more weight should be given to individual risk loss experience.

Each risk in the class has its own individual risk mean called its hypothetical mean. The Variance of the Hypothetical Means (VHM) across risks in the class is a statistical measure for the homogeneity or vice versa, heterogeneity, within the class. A smaller VHM indicates more class homogeneity and, consequently, argues for more weight going to the class rate. A larger VHM indicates more class heterogeneity and, consequently, argues for less weight going to the class rate.

How much variation is there in an individual risks loss experience? If there is a large amount of variation expected in the actual loss experience for an individual risk, then the actual experience observed may be far from its expected value and not very useful for estimating the expected value. In this case, less weight, i.e., less credibility, should be assigned to individual experience. The process variance, which is the variance of the risks random experience about its expected value, is a measure of the variability in an

individual risks loss experience. The Expected Value of the Process Variance (EPV) is the average value of the process variance over the entire class of risks.

Let $\bar{X}_i$ represent the sample mean of n observations for a randomly selected risk$i$. Because there are n observations, the variance in the sample mean $\bar{X}_i$ is the variance in one observation for the risk divided by n. Given risk i, this variance is PV_i where PV is the process variance of one observation. Because risk i was selected at random from the class of risks, an estimator for its variance is $E[PV_i/n] = E[PV_i]/n - EPV/n$. This is the Expected Value of the Process Variance for risks in the class divided by the number of observations made about the selected risk. It measures the variability expected in an individual risks loss experience.

Letting $\hat{\mu}$ represent the overall class mean, a risk selected at random from the class will have an expected value equal to the class mean $\hat{\mu}$. The variance of the individual risk means about μ is the VHM, the Variance of the Hypothetical Means. There are two estimators for the expected value of the ith risk: (1) the risks sample mean $\bar{X}_i$, and (2) the class mean $\hat{\mu}$. How should these two estimators be weighted together? A linear estimate with the weights summing to 1.00 would be:

$$Estimate = w\bar{X}_i + (1 - w)\hat{\mu}.$$

An optimal method for weighting two estimators is to choose weights proportional to the reciprocals of their respective variances. This results in giving more weight to the estimator with smaller variance and less weight to the estimator with larger variance. In many situations this will result in a minimum variance estimator. The resulting weights

are:

$$w = \frac{\frac{1}{EPV/n}}{\frac{1}{EPV/n} + \frac{1}{VHM}}$$

and

$$(1 - w) = w = \frac{\frac{1}{VHM}}{\frac{1}{EPV/n} + \frac{1}{VHM}}$$

Thus,

$$w = \frac{n}{n + \frac{EPV}{VHM}} \text{ and } (1 - w) = 1 - \frac{n}{n + \frac{EPV}{VHM}}.$$

Setting $K = EPV/VHM$, the weight assigned to the risks observed mean is

$$w = \frac{n}{n+K}.$$

The actual observation during time t for that particular risk or group will be denoted by x_t, which will be the observation of corresponding random variable X_t, where t is an integer. For example, X_t may represent the following:

- Number of claims in period t;
- Loss ratio in year t;
- Loss per exposure in year t.

An individual risk is a member of a larger population and the risk has an associated risk parameter Θ that distinguishes the individuals risk characteristics. It is assumed that the risk parameter is distributed randomly through the population and Θ will denote

the random variable. The distribution of the random variable X_t depends upon the value of Θ: $f_{x|\Theta}(x_t|\Theta)$. For example, Θ may be a parameter in the distribution function of X_t.

If X_t is a continuous random variable, the mean for X_t given $\Theta = \theta$, is the conditional expectation, $E_{X|\Theta}[X_t|\Theta = \theta] = \int x_t f_{X|\Theta}(x_t|\theta)dx_t = \mu(\theta)$, where the integration is over the support of $f_{X|\Theta}(x_t|\theta)$. If X_t is a discrete random variable, then a summation should be used:

$$E_{X|\Theta}[X_t|\Theta = \theta] = \sum_{all x_t} x_t f_{X|\Theta}(x_t|\theta).$$

The risk parameter represented by the random variable Θ has its own probability density function (pdf): $f_\Theta(\theta)$. The pdf for Θ describes how the risk characteristics are distributed within the population. If two risks have the same parameter θ, then they are assumed to have the same risk characteristics including the same mean $\mu(\theta)$.

The unconditional expectation of X_t is:

$$E[X_t] = \iint_{\mathbb{R}} x_t f_{X,\Theta}(x_t, \theta)dx_t d\theta = \iint_{\mathbb{R}} x_t f_{X,\Theta}(x_t, \theta) f_\Theta(\theta)dx_t d\theta =$$

$$\int_{\mathbb{R}} \left[\int x_t f_{X,\Theta}(x_t, \theta)dx_t\right] f_\Theta(\theta)d\theta = E_\Theta[E_{X|\Theta}[X_t|\Theta]] = E_\Theta[\mu(\theta)] = \mu.$$

The conditional variance of X_t given $\Theta = \theta$ is

$$Var_{X|\Theta}[X_{t|\Theta} = \theta] = E_{X|\Theta}[(X_t - \mu(\theta))^2|\Theta = \theta] = \iint_{\mathbb{R}} (X_t - \mu(\theta))^2 f_{X|\Theta}(x_t|\theta)dx_t = \sigma^2(\theta).$$

This variance is also called the process variance for the selected risk. The unconditional variance of X_t, also referred to as the total variance, is given by the Total Variance formula:

$$Var[X_t] = Var_\Theta[E_{X|\Theta}[X_t|\Theta]] + E_\Theta[Var_{X|\Theta}[X_t|\Theta]].$$

4.3 Derivation of the Bühlmann Credibility Model

The Bühlmann model assumes that for any selected risk, the random variables $\{X_1, X_2, ..., X_n, X_{n+1}, ...\}$ are independently and identically distributed. For the selected risk, each X_t has the same probability distribution for any time period t, both for the $X_1, X_2, .., X_n$ random variables in the experience period, and future outcomes $X_{n+1}, X_{n+2}, ...$ As Hans Bühlmann described it, homogeneity in time is assumed.

The characteristics that determine the risks exposure to loss are assumed to be unchanging and the risk parameter θ associated with the risk is constant through time for the risk. The means and variances of the random variables for the different time periods are equal and are labeled $\mu(\theta)$ and $\sigma^2(\theta)$, respectively.

Hypothetical Mean: $\mu(\theta) = E_{X|\Theta}[X_1|\theta] = ... = E_{X|\Theta}[X_N|\theta] = E_{X|\Theta}[X_{N+1}|\theta]$.

Process Variance: $\sigma^2(\theta) = Var_{X|\Theta}[X_1|\theta] = ... = Var_{X|\Theta}[X_N|\theta] = Var_{X|\Theta}[X_{N+1}|\theta]$.

The hypothetical means and process variances will vary among risks, but they are assumed to be unchanging for any individual risk in the Bühlmann model. To apply Bühlmann credibility, the average values of these quantities over the whole population of risks are needed, along with the variance of the hypothetical means for the population:

1. Population Mean: $\mu = E_\Theta[\mu(\Theta)] = E_\Theta[E_{X|\Theta}[X_t|\Theta]]$

2. Expected Value of Process Variance: $EPV = E_\Theta[\sigma^2(\Theta)] = E_\Theta[Var_{X|\Theta}[X|\Theta]]$

3. Variance of Hypothetical Means: $VHM = Var_{\Theta}[\mu(\Theta)] = E_{\Theta}[(\mu(\Theta) - \mu)^2]$

The population mean $\mu = E_{\Theta}[E_{X|\Theta}[X_t|\Theta]]$ provides an estimate for the expected value of X_t in the absence of any prior information about the risk. The EPV indicates the variability to be expected from observations made about individual risks. The VHM is a measure of the differences in the means among risks in the population.

Because $\mu(\Theta)$ is unknown for the selected risk, the mean $\bar{X} = (\frac{1}{N})\sum_{t=1}^{N} X_t$, is used in the estimation process. It is an unbiased estimator for $\mu(\theta)$,

$$E_{X|\Theta}[\bar{X}|\theta] = E_{X|\Theta}[(\tfrac{1}{N})\textstyle\sum_{t=1}^{N} X_t|\theta] = (\tfrac{1}{N})\textstyle\sum_{t=1}^{N} E_{X|\Theta}[X_t|\theta] = (\tfrac{1}{N})\textstyle\sum_{t=1}^{N} \mu(\theta) = \mu(\theta).$$

The conditional variance of $\bar{X}$, assuming independence of the X_t given θ, is

$$Var_{X|\Theta}[\bar{X}|\theta] = Var_{X|\Theta}[(\tfrac{1}{N})\textstyle\sum_{t=1}^{N} X_t|\theta] =$$

$$(\tfrac{1}{N})^2\textstyle\sum_{t=1}^{N} Var_{X|\Theta}[X_t|\theta] = (\tfrac{1}{N}^2)\textstyle\sum_{t=1}^{N} \sigma^2(\theta) = \frac{\sigma^2(\theta)}{N}.$$

The unconditional variance of $\bar{X}$ is

$$Var[\bar{X}] = Var_{\Theta}[E_{X|\Theta}[\bar{X}|\Theta]] + E_{\Theta}[Var_{X|\Theta}[\bar{X}|\Theta]] =$$

$$Var_{\Theta}[\mu(\Theta)] + \frac{E_{\Theta}[\sigma^2(\Theta)]}{N} = VHM + \frac{EPV}{N}$$

The Bühlmann credibility assigned to estimator $\bar{X}$ is given by the well-known formula

$$Z = \frac{N}{N+K},$$

where N is the number of observations for the risk and $K = EPV/VHM$. Multiplying the numerator and denominator by (VHM/N) gives an alternative form:

$$Z = \frac{VHM}{VHM + \frac{EPV}{N}}.$$

Therefore $Z = \frac{N}{N+K}$, can be written as

$$Z = \frac{Var_{\Theta}[\mu(\Theta)]}{Var[X]}.$$

The numerator is a measure of how far apart the means of the risks in the population are, while the denominator is a measure of the total variance of the estimator. The credibility weighted estimate for

$$\mu(\theta) = E_{X|\Theta}[X_t|\theta], \text{ for } t = 1, 2, ..., N, N+1, ...,$$

$$\hat{\mu}(\theta) = Z \cdot \bar{X} + (1 - Z) \cdot \mu.$$

The estimator $\hat{\mu}(\theta)$ is a linear least squares estimator for $\mu(\theta)$. This means that

$E[[Z \cdot \bar{X} + (1 - Z) \cdot \mu] - \mu(\Theta)^2]$ is minimized when $Z = N/(N + K)$.

Continuing Example 4:

Suppose that random person performs this random experiment (we do not know whether he uses Box 1 or Box 2) and that his first ball is a 1 ($X_1 = 1$) and his second ball is a 2 ($X_2 = 2$). What is the predicted value X_3 of the third selected ball?

Solution:

The following restates the prior distribution of Θ and the conditional distribution of $X|\Theta$. We denote "white ball from the bowl" by $\Theta = 1$ and "red ball from the bowl" by $\Theta = 2$.

$$\pi_\theta(1) = 0.8, \ \pi_\theta(2) = 0.2$$

$$f_{X|\Theta}(0|\theta = 1) = 0.60, \;\; f_{X|\Theta}(1|\theta = 1) = 0.30,$$

$$f_{X|\Theta}(2|\theta = 1) = 0.10,$$

$$f_{X|\Theta}(0|\theta = 2) = 0.15, \;\; f_{X|\Theta}(1|\theta = 2) = 0.35,$$

$$f_{X|\Theta}(2|\theta = 2) = 0.50.$$

The following computes the conditional means (hypothetical means) and conditional variances (process variances) and the other parameters of the Bühlmann method.

Hypothetical Means:

$$E[X|\Theta = 1] = 0.60(0) + 0.30(1) + 0.10(2) = 0.50,$$

$$E[X|\Theta = 2] = 0.15(0) + 0.35(1) + 0.50(2) = 1.35,$$

$$E[X^2|\Theta = 1] = 0.60(0) + 0.30(1) + 0.10(4) = 0.70,$$

$$E[X^2|\Theta = 2] = 0.15(0) + 0.35(1) + 0.50(4) = 2.35.$$

Process Variances:

$$Var[X|\Theta = 1] = 0.70 - 0.50^2 = 0.45,$$

$$Var[X|\Theta = 2] = 2.35 - 1.35^2 = 0.5275.$$

Expected Value of the Hypothetical Means:

$$\mu = E[X] = E[E[X|\Theta]] = 0.80(0.50) + 0.20(1.35) = 0.67.$$

Expected Value of the Process Variance:

$$EPV = E[Var[X|\Theta]] = 0.8(0.45) + 0.20(0.5275) = 0.4655.$$

Variance of the Hypothetical Means:

$$VHM = Var[E[X|\Theta]] = 0.80(0.50)^2 + 0.20(1.35)^2 - 0.67^2 = 0.1156$$

Bühlmann Credibility Factor:

$$K = \frac{4655}{1156},$$

$$Z = \frac{2}{2+\frac{4655}{1156}} = \frac{2312}{6967} = 0.33185.$$

Bühlmann Credibility Estimate:

$$C = \frac{2312}{6967}\frac{3}{2} + \frac{4655}{6967}(0.67) = \frac{6586.85}{6967} = 0.9454356. \quad \blacksquare$$

Chapter 5

The Bühlmann - Straub Credibility Model

5.1 Credibility Parameters

The requirement that the random variables $X_1, X_2, ..., X_N, X_{N+1}...$ for a risk be identically distributed is easily violated in the real world. For example:

- The work force of a workers compensation policyholder may change in size from one year to the next.
- The number of vehicles owned by a commercial automobile policyholder may change through time.
- The amount of earned premium for a rating class varies from year to year.

In all of these cases, one should not assume that variables $X_1, X_2, ..., X_N, X_{N+1}...$ are identically distributed, although an assumption of independence may be warranted. A risks exposure to loss may vary and it is assumed that this exposure can be measured. Some measures of exposure to loss are:

- Amount of insurance premium

- Number of employees
- Payroll
- Number of claims

The Bühlmann-Straub model assumes that the means of the random variables are equal for the selected risk, but that the process variances are inversely proportional to the size of the risk during each observation period. For example, when the risk is twice as large, the process variance is halved. For the Bühlmann-Straub model occurs following assuptions:

Hypothetical Mean for Risk θ per Unit of Exposure: $\mu(\theta) = E_{X|\Theta}[X_N|\theta]$.

Process Variance for Risk θ: $Var_{X|\theta}[X_N|\theta] = \frac{\sigma^2(\theta)}{m_N}$.

The random variables X_t represent number of claims, monetary losses, or some other quantity of interest per unit of exposure, and m_t is the measure of exposure.

How should random variables $X_1, X_2, ..., X_N, X_{N+1}...$ associated with a selected risk (or group of risks) be combined to estimate the hypothetical mean $\mu(\theta)$? A weighted average using the exposures m_t will give a linear estimator for $\mu(\theta)$ with minimum variance:

$$\hat{\mu} = \sum_{t=1}^{N} m_t.$$

The weighted average is

$$\bar{X} = \sum_{t=1}^{N} (\frac{m_t}{m}) X_t.$$

Recall that the variance of each X_t given θ is $\sigma^2(\theta)/m_t$. For a weighted average $\bar{X} = \sum_{t=1}^{N} w_t X_t$, the variance of $\bar{X}$ will be minimized by choosing the weights w_t to be inversely proportional to the variances of the each individual X_t. That is, random variables with smaller variances should be given more weight. So, the following weights $w_t = m_t/m$ are called for under the current assumptions.

The conditional expected value and variance of X given risk parameter θ are

$$E_{X|\Theta}[\bar{X}|\theta] = E_{X|\Theta}[\sum_{t=1}^{N} \tfrac{m_t}{m} X_t|\theta] = \sum_{t=1}^{N} \tfrac{m_t}{m} E_{X|\Theta}[X_t|\theta] = \sum_{t=1}^{N} \tfrac{m_t}{m} \mu(\theta) = \mu(\theta),$$

$$Var_{X|\Theta}[\bar{X}|\theta] = Var_{X|\Theta}[\sum_{t=1}^{N} \tfrac{m_t}{m} X_t|\theta] = \sum_{t=1}^{N} \tfrac{m_t}{m} Var_{X|\Theta}[X_t|\theta] = \sum_{t=1}^{N} \tfrac{m_t}{m} (\tfrac{\sigma^2(\theta)}{m_t}) = \tfrac{\sigma^2(\theta)}{m}.$$

The EPV and VHM are defined to be $EPV = E_\theta[\sigma^2(\Theta)]$ and $VHM = Var_\theta[\mu(\Theta)]$ where the expected value is over all risk parameters in the population. The loss per unit of exposure is used because the exposure can vary through time and from risk to risk. The unconditional mean and variance of $\bar{X}$ are

$$E[\bar{X}] = E_\Theta[E_{X|\Theta}[\bar{X}|\Theta]] = E_\Theta[\mu(\Theta)] = \mu,$$

$$Var[\bar{X}] = Var_\Theta[E_{X|\Theta}[\bar{X}|\Theta]] + E_\Theta[Var_{X|\Theta}[\bar{X}|\Theta]]] =$$

$$Var_\Theta[\mu(\Theta)] + \tfrac{E_\Theta[\sigma^2(\Theta)]}{m} = VHM + \tfrac{EPV}{m}.$$

The credibility assigned to the estimator $\bar{X}$ of $\mu(\theta)$ is

$$Z = \tfrac{VHM}{VHM + \frac{EPV}{m}} \Rightarrow Z = \tfrac{m}{m+K}.$$

The total exposure m replaces N in the Bühlmann formula and the parameter K is defined as usual:

$$K = \frac{EPV}{VHM} = \frac{E_\Theta[\sigma^2(\Theta)]}{Var_\Theta[\mu(\Theta)]}$$

The Bühlmann model is actually a special case of the more general Bühlmann-Straub model with $m_t = 1\ \forall t$. The credibility weighted estimate is

$$\hat{\mu}(\theta) = Z \cdot \bar{X} + (1 - Z) \cdot \mu.$$

5.2 Nonparametric Estimation

In a nonparametric setup, no assumptions are made about the form or parameters of the distributions of X_{it}, nor are any assumptions made about the distribution of the risk parameters Θ_i. In the Bühlmann-Straub model, the N_i outcomes for risk i have the same means but the process variances are inversely related to the exposure. The number of observations N_i has a subscript indicating that the number of observations can vary by risk in the Bühlmann-Straub model.

The Bühlmann-Straub Model is more complicated because a risks exposure to loss can vary from year to year, and the number of years of observations can change from risk to risk. The reason that Bühlmann-Straub can handle varying numbers of years is because the number of years of data for a risk is reflected in the total exposure for the risk. Estimators for risk mean and variance are the following:

$$\bar{X} = \sum_{i=1}^{R} m_i \bar{X}_i / m \text{ and } EPV = \sum_{i=1}^{R} (N_i - 1)\hat{\sigma}_i{}^2 / (\sum_{i=1}^{R} (N_i - 1))$$

In the Bühlmann-Straub model, the mean is assumed to be constant through time for each risk i:

$$\mu(\theta_i) = E_{X|\theta}[X_{i1}|\theta_i] = E_{X|\theta}.$$

Also, $\bar{X}_i$ is an unbiased estimator for the mean of risk i:

$$E_{X|\Theta}[\bar{X}_i|\Theta_i] = E_{X|\Theta}[(\tfrac{1}{m_i})\sum_{i=1}^{N_i} m_{it}X_{it}|\theta_i] =$$

$$(\tfrac{1}{m_i})\sum_{i=1}^{N_i} m_{it}E_{X|\Theta}[X_{it}|\theta_i] = (\tfrac{1}{m_i})\sum_{i=1}^{N_i} m_{it}\mu(\theta_i) = \mu(\theta_i).$$

The process variance of X_{it} is inversely proportional to the exposure:

$$Var_{X|\theta}[X_{it}|\theta_i] = \sigma^2(\theta_i)/m_{it}.$$

This means that for risk i,

$$\sigma^2(\theta_i) = m_{it}E_{X|\Theta}[(X_{it} - \mu(\theta_i))^2|\theta_i], for\ t = 1\ to\ N_i.$$

$$\sigma^2(\theta_i) = m_{it}E_{X|\Theta}[(X_{it} - \mu(\theta_i))^2|\theta - i], for\ t = 1\ to\ N_i.$$

Summing both sides over t and dividing by the number of terms N_i yields

$$(\tfrac{1}{N_i})\sum_{i=1}^{N_i} \sigma_i^2(\theta_i) = (\tfrac{1}{N_i})\sum_{i=1}^{N_i} m_{it}E_{X|\Theta}[(X_{it} - \mu(\theta_i))^2|\theta - i],\ or$$

$$\sigma_i^2(\theta_i) = E_{X|\Theta}[(\tfrac{1}{N_i})\sum_{i=1}^{N_i} m_{it}(X_{it} - \mu(\theta_i))^2|\theta - i].$$

The quantity $\mu(\theta_i)$ is unknown, so $\bar{X}_i$ is used instead in the estimation process. This reduces the degrees of freedom by one so N_i is replaced by $N_i - 1$ in the denominator:

$$\sigma_i^2(\theta_i) = E_{X|\Theta}[(\tfrac{1}{N_i-1})\sum_{i=1}^{N_i} m_{it}(X_{it} - \bar{X}_i)^2|\theta - i].$$

Thus, an unbiased estimator for σ_i^2 is

$$(\hat{\sigma}_i)^2 = (\tfrac{1}{N_i-1})\sum_{i=1}^{N_i} m_{it}(X_{it} - \bar{X}_i)^2.$$

The EPV can be estimated by combining process variance estimates ${\hat{\sigma}_i}^2$ of the R risks. If they are combined with weights $w_i = (N_i - 1)/(\sum_{i=1}^{R}(N_i - 1))$, then an unbaised estimator for the EPV is

$$E\hat{P}V = \sum_{i=1}^{R} w_i(\hat{\sigma}_i)^2 = \frac{\sum_{i=1}^{R}\sum_{t=1}^{N_i} m_{it}(X_{it}-\bar{X}_i)^2}{\sum_{i=1}^{R}(N_i-1)}$$

The hypothetical mean for risk i is $\mu(\theta_i)$. The variance of the hypothetical means can be written as

$$VHM = E_\theta[(\mu(\Theta_i) - \mu)^2], \text{ where } \mu = E_\theta[\mu(\Theta_i)].$$

Because the observed values of the random variables $\bar{X}_i$ and $\bar{X}$ are estimators for $\mu(\theta_i)$ and μ, respectively, a good starting point for developing an estimator for the variance of the hypothetical means is: $\frac{1}{(R-1)}\sum_{i=1}^{R} m_i(\bar{X}_i - \bar{X})^2$. Each term is weighted by its total exposure over the experience period. However, this is not unbiased. An unbiased estimator is

$$V\hat{H}M = \left(\sum_{i=1}^{R} m_i(\bar{X}_i - \bar{X})^2 - (R-1)E\hat{P}V\right) / \left(m - \left(\frac{1}{m}\right)\sum_{i=1}^{R} m_i^2\right).$$

With the Bühlmann-Straub model, the measure to use in the credibility formula is the total exposure for risk i over the whole experience period. The formulas to compute credibility weighted estimates are

$$\hat{K} = \frac{E\hat{P}V}{V\hat{H}M}, \ \hat{Z}_i = \frac{m_i}{m_i+K}, \text{ and}$$

$$\hat{\mu}(\theta_i) = \hat{Z}_i \cdot \bar{X}_i + (1 - \hat{Z}_i) \cdot \bar{X}.$$

Chapter 6

An Analysis of Colorado Cancer Death Rates

6.1 Introduction to the Cancer Data Set

The case study presented in this chapter will be based on Colorado Health Information Dataset (CoHID), furnished by the Colorado Department of Public Health and Environment. Death data are compiled from information reported on the Certificate of Death. Data items are presented as reported. Information on the certificate concerning time, place, and cause of death is typically supplied by medical personnel or coroners. Demographic information, such as age, race/ethnicity, or occupation, is generally reported on the certificate by funeral directors from information supplied by the available next of kin. CoHID only reports data for Colorado resident deaths.

Colorado Health Information Dataset presents Cause of Death Classification and Total Crude Death Rate, the number of deaths per a specified number of population (i.e., per 1,000 or 100,000). Crude rates are not adjusted for differences in demographic distributions among populations, such as age distributions. This dataset contains information on over 100 causes of death in Colorado from 1990 through the most recent year avail-

Table 6.1: Denver County Deaths by Cancer Data Set

Year	Total Death	Total Population	Total Crude Death Rate per 100,000
2000	909	556,738	163.3
2001	919	563,300	163.1
2002	947	559,090	169.4
2003	906	560,348	161.7
2004	914	560,230	163.1
2005	859	559,459	153.5
2006	828	562,862	147.1
2007	808	570,437	141.6
2008	858	581,903	147.4
2009	846	595,573	142
2010	897	604,875	148.3
2011	870	620,917	140.1
2012	888	634,619	139.9
Average	880.7	579,258	152

Table 6.2: Non-Denver Deaths by Cancer Data Set

Year	Total Death	Total Population	Total Crude Death Rate per 100,000
2000	4,987	3,782,063	131.8
2001	5,215	3,881,213	134.3
2002	5,425	3,945,619	137.5
2003	5,494	3,994,736	137.5
2004	5,271	4,048,581	130.1
2005	5,508	4,103,075	134.2
2006	5,695	4,182,798	136.1
2007	5,782	4,251,347	136
2008	5,851	4,320,035	135.4
2009	6,092	4,381,280	139
2010	6,132	4,445,108	137.9
2011	6,167	4,497,609	137.1
2012	6,426	4,554,064	141.1
Average	5,695.7	4,183,656	136.1

able. The user is able to query death records for information such as: cause of death by race/ethnicity, gender and age. The data can be extracted by zip code, county, or region, allowing user also to analyze it geographically.

The death by cancer report from 2000 to 2012 (inclusive) is captured in Tables 6.1 and 6.2. Tables represent Denver County Dataset and Non-Denver (which is the rest of Colorado State, excluding Denver). The first method to attempt is a Bayesian calculation of the predictive distribution for the upcoming year, irrespective of class (i.e., $\mu_{n+1}(\theta)$). Designating the Denver county as D, and the Non-Denver as N, we have $\bar{x}_D = 152$ and $\bar{x}_N = 136.14$. Now, the prior probabilities are,

$$\pi(\theta_D) = .121618362,\ \pi(\theta_D) = .878381638.$$

However, the joint probabilities are now,

$$f_{\mathbf{X},\Theta_D}(\mathbf{x}, \theta_D) = f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \cdot \pi(\theta_D) = 3.9031 \cdot 10^{-16},$$

$$f_{\mathbf{X},\Theta_N}(\mathbf{x}, \theta_N) = f_{\mathbf{X}|\Theta}(\mathbf{x}|\theta) \cdot \pi(\theta_N) = 2.8912 \cdot 10^{-15},$$

which, when summed, yield a marginal probability function, $f(\mathbf{x}) = 3.2824 \cdot 10^{-15}$. This implies that the posterior probabilities are,

$$\pi_{\Theta_D|\mathbf{X}}(\theta_D|\mathbf{x}) = \frac{f_{\mathbf{X},\Theta_D}(\mathbf{x},\theta_D)}{f_{\mathbf{X}}(\mathbf{x})} = .118909024$$

$$\pi_{\Theta_N|\mathbf{X}}(\theta_N|\mathbf{x}) = \frac{f_{\mathbf{X},\Theta_N}(\mathbf{x},\theta_N)}{f_{\mathbf{X}}(\mathbf{x})} = .881090976$$

Thus, the Bayesian estimate for the average of the Crude Death Rate for an upcoming year is:

$$E(X_{n+1}|x) = \sum_{\Theta} \mu_{n+1}(\Theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = 155.98$$

for Denver county and

$$E(X_{n+1}|x) = \sum_{\Theta} \mu_{n+1}(\Theta) \cdot \pi_{\Theta|\mathbf{X}}(\theta|\mathbf{x}) = 132.7$$

for Non-Denver.

Now we will illustrate the Bühlmann credibility method. An unbiased estimator of μ is $\hat{\mu} = \bar{X} = 144.1922071$. Hence, an unbiased estimator of υ is

$$\hat{\upsilon} = \frac{1}{r}\sum_{i=1}^{r} \hat{\upsilon}_i = \frac{1}{r(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2 = 54.94614462.$$

Since we already have an unbiased estimator of υ given above, an unbiased estimator of a is given by

$$\hat{a} = \frac{1}{r-n}\sum_{i=1}^{r}(\bar{X}_i - \bar{X})^2 - \frac{\hat{\upsilon}}{n} = 128.7470678.$$

Then, $\hat{k} = \hat{\upsilon}/\hat{a} = 0.426775891$. The estimated credibility factor is $Z = 0.968214567$. The estimated credibility premiums (or the crude death rates) are 152.08, for Denver County and 136.29, for Non-Denver.

The next method to employ is the Buhlmann-Straub credibility model. It is a more general Bühlmann setup, with $E(X_{ij}) = E[E(X_{ij}|\Theta_i)] = E[\mu(\Theta_i)] = \mu(\Theta_i)$. The credibility parameters can be calculated as follows: $\bar{X}_D = 152.0380657$, $\bar{X}_N = 136.1433452$. The overall mean is $\hat{\mu} = \bar{X} = 138.0764351$.

Now, we obtain an unbiased estimator of υ, namely,

$$\hat{\upsilon} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}(X_{ij}-\bar{X}_i)^2}{\sum_{i=1}^{r}(n_i-1)} = 91825379.36.$$

Table 6.3: Table of Results for Denver County

Method	Denver County (Crude Death Rate per 100,000)
Bühlmann-Straub Credibility Premium	150.744
μ(weighted)	143.760
Bühlmann-Straub(weighted)	151.271
Bühlmann Credibility Premium	152.087
Posterior probabilities $\pi(\theta)$	0.118909024
Bayesian Credibility Premium	155.980

An unbiased estimator for a may be obtained by replacing υ by an unbiased estimator $\hat{\upsilon}$ and solving for a. That is, an unbiased estimator of a is

$$\hat{a} = (m - m^{-1}\sum_{i=1}^{r} m_i^2)^{-1}\left[\sum_{i=1}^{r} m_i(\bar{X}_i - \bar{X})^2 - \hat{\upsilon}(r-1)\right] = 119.3798741.$$

Then, $\hat{k} = \hat{\upsilon}/\hat{a} = 769186.4316$. The estimated credibility factors are $Z_D = 0.907321771$ and $Z_N = 0.986054528$. The estimated credibility premiums (or the crude death rates) are 150.74, for Denver County and 136.17, for Non-Denver.

For the alternative estimator we would use

$$\hat{\mu} = \frac{\sum_{i=1}^{r} \hat{Z}_i \bar{X}_i}{\sum_{i=1}^{r} \hat{Z}_i} = 143.7602283$$

The credibility premiums are 151.27 for Denver County and 136.24 for Non-Denver. Table 6.3 captures the results of finding the Crude Death Rate for Denver and Non-Denver Countys using credibility methods. Table 6.5 captures the results of finding the Crude Death Rate for Denver County for particular year, using the information in the data set, excluding the predicted year.

Table 6.4: Table of Results for Non-Denver

Method	Non-Denver County (Crude Death Rate per 100,000
Bühlmann-Straub Credibility Premium	136.170
μ(weighted)	143.760
Bühlmann-Straub(weighted)	136.250
Bühlmann Credibility Premium	136.297
Posterior probabilities $\pi(\theta)$	0.881090976
Bayesian Credibility Premium	132.701

Table 6.5: Table of Year - Predicted Crude Death Rate Estimates

	Bühlmann-Straub	Bühlmann-Straub(weighted)	Bühlmann	Bayesian	Actual Rate
2000	149.720	150.292	151.138	154.727	163.300
2001	149.653	150.250	151.154	154.672	163.100
2002	149.303	149.828	150.668	153.585	169.400
2003	149.804	150.377	151.271	153.482	161.700
2004	149.786	150.354	151.155	154.824	163.100
2005	150.398	151.017	151.942	155.868	153.500
2006	150.978	151.571	152.493	156.288	147.100
2007	151.522	152.085	152.979	156.414	141.600
2008	150.954	151.554	152.466	156.287	147.400
2009	151.613	152.131	152.949	156.495	142.000
2010	150.941	151.512	152.393	156.273	148.300
2011	151.809	152.343	153.115	156.509	140.100
2012	152.045	152.490	153.142	156.643	139.900

6.2 Regression Model Diagnostics

Regression Diagnostics refer to statistical techniques that can be used to check whether or not a linear model is appropriate for given bivariate data set. We will apply three diagnostic techniques. First of them is to make a scatterplot and visually assess whether a linear model would be appropriate. This technique is very useful. In most circumstances, a visual inspection of the will tell us right away if a linear model will work or not. If there is a clear linear trend, then a linear regression model will likely be appropriate. If there is an apparent non-linear trend or no trend at all, a linear model would likely be a poor choice as a means for describing the relationship between the variables x and y.

We will use linear regression only for projecting the population size.The Crude Deaths Rates used credibility methods only. As we can see from the plot of Denver County, a linear trend is clearly evident and captures the overall trend well. Thus, we would expect that a linear regression model would be appropriate in this case.

The next technique refers to an R^2 value, the proportion of variability explained be the regression r^2. To obtain the value of R^2, the principle of least squares was used, and as such, the regression line minimized the sum of squared errors between the proposed line and each of the data points. The actual values of this sum of the errors for a given data set is usually abbreviated as SSE and can be shown that

$$SSE = S_{yy} - \frac{S_{xy}^2}{S_{xx}}.$$

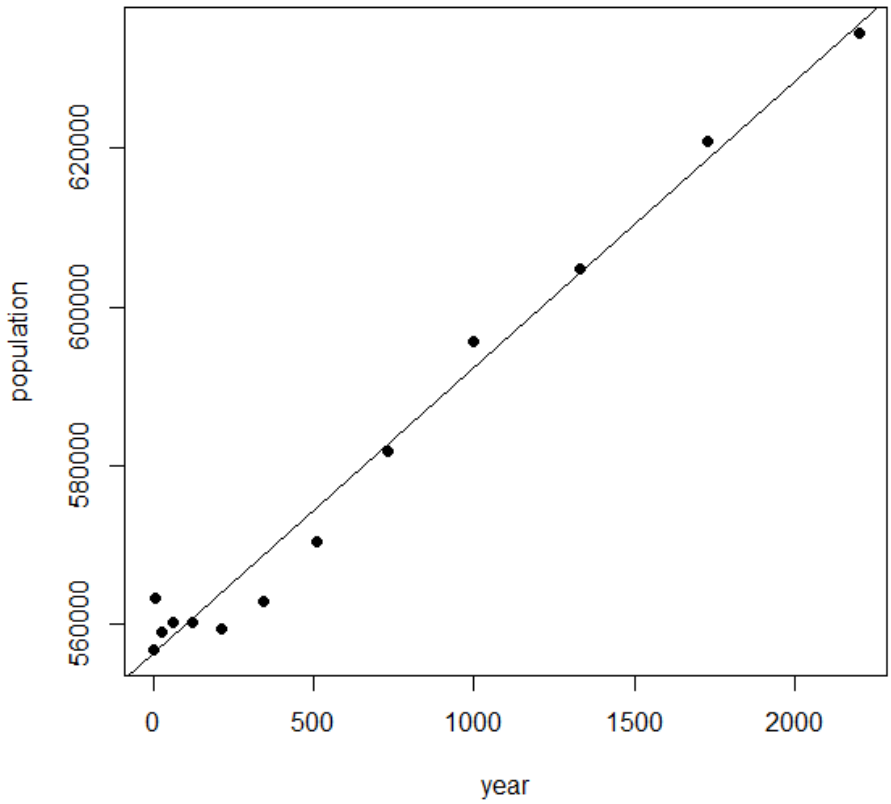

Figure 6.1: Linear Regression Scatterplot for Denver County

In a situation where every point lies exactly on the regression line, the linear model would have accounted for all the variability in y and SSE would equal zero (there would be no errors or vertical distances between the points and the line). Since S_{yy} quantifies the total amount of variability in y, showing that if SSE is the amount of S_{yy} not explained by the regression, then the balance must be the amount of variability in y that was explained by the regression and therefore equal to

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

The R^2 is a measure of how good the regression is performing. Values of R^2 close to 1 imply that the regression is performing well and is likely appropriate for the data

Table 6.6: Summary Statistics for Denver County Dataset

Coefficients	Estimate	Std. Error	$Pr(>\|t\|)$	R^2
Intercept	556,200	1,356	$2 \cdot 10^{-16}$	0.9829
Record.Year	36.14	1.438	$4.54 \cdot 10^{-11}$	0.9829

is was fit to, and values of R^2 close to 0 imply a very poor fit. Table 6.5 captures the linear regression results for Denver data set.

Therefore, the R^2 value is high and so close to one, which suggests that the line is accounting for a very good proportion of the variability in y. Also, using the summary we can obtain the equation of the regression line and using the Denver Dataset, predict the population for the upcoming year 2013.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 556,200 + 36.14(14)^3 = 563,284.$$

$$\widehat{Population} = 556,200 + 36.14(year - 1999)^3 = 563,284.$$

Thus, the predicted population for Denver County in 2013 is 563,284 people.

The third diagnostic technique involves plotting the residuals, on the vertical axis, against the corresponding fitted $\hat{y}$ values on the horizontal axis. In an ideal situation, the plot of the residuals vs fitted values would have no trend or pattern, but just look like random values centered around 0. To the extent that any type of trend is found in this scatterplot, this would suggest that a linear model is less and less appropriate. As can be seen in Figure 6.2, there is no clear trend in the points which is what we want.

All three of the techniques should be used together for the purpose of trying to make a decision regarding linear model. Decisions pertaining to the ultimate appropriateness

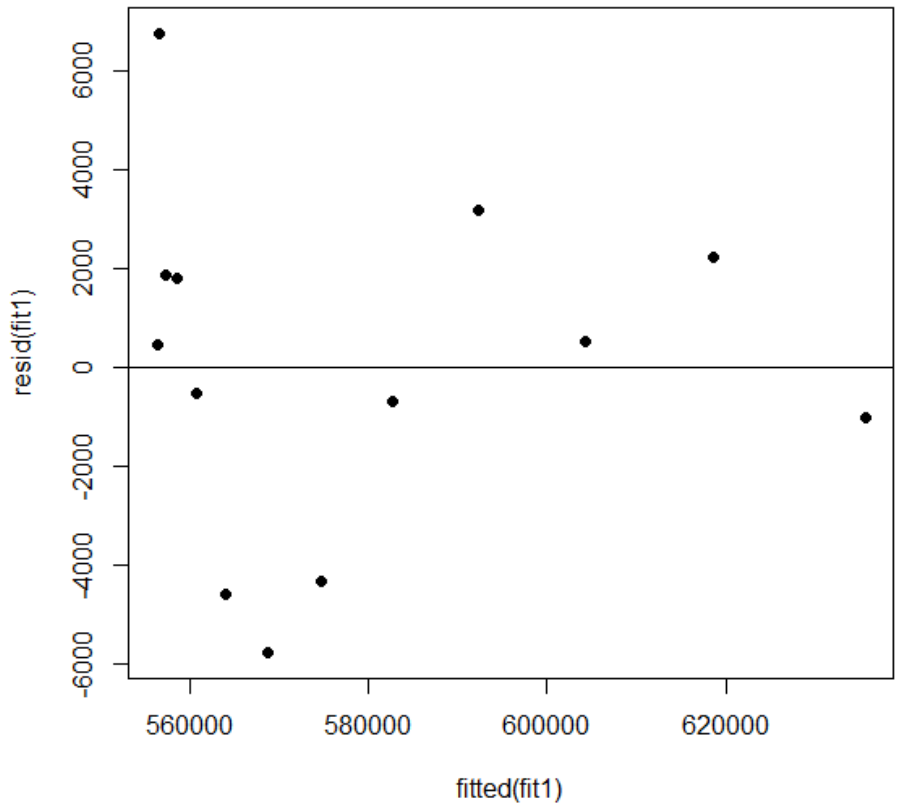

Figure 6.2: Residuals vs. Fitted values plot for Denver County

of the linear model should take into account all of the evidence providing by all of the diagnostics. Thus, all of the diagnostic technoques point to the viability of the linear regression model. We therefore deem the model appropriate. Using the predicted values for Crude Death Rate and Population, the total number of deaths was predicted and is shown in Table 6.6[1].

6.3 Analysis of Results

Three methods for modeling the resulting number of deaths produced different, but close, results. The Bayesian estimate is superior among the credibility models primarily

[1]Regression was used to predict Denver's Population for 2013.

Table 6.7: Results for the Denver County Dataset

Method	Year	Crude Death Rate	Population	Number of Deaths
Bühlmann-Straub	2013	150.744	563,284	849
Bühlmann-Straub(weighted)	2013	151.271	563,284	852
Bühlmann	2013	152.087	563,284	857
Bayesian	2013	155.980	563,284	879

because it accounted for the non-linearity inherent. Two main functions of Bayesian approach were calculated. The posterior distribution tells us how our opinion about the parameter has changed once we have observed the data, which was $\pi(\theta) = 0.118909024$ for Denver County and $\pi(\theta) = 0.881090976$ for Non-Denver. The predictive distribution tells us what the next observation might look like given the information contained in the data (as well as, implicitly, our prior opinion), and the prediction for Denver Crude Death Rate was around 155.98 (for 100,000 people) for the upcoming year of 2013.

Using the Bühlmann model and assume that that for any selected risk, the random variables $\{X_1, X_2, ..., X_n, X_{n+1}, ...\}$ are independently and identically distributed, we managed to find that the Crude Death Rate for year 2013 will be 152.087 (for 100,000 people). The more general Bühlmann-Straub approach, assuming that the random variables $X_1, X_2, ..., X_N, X_{N+1}...$ for a risk be identically distributed is easily violated in the real world , gave the close result of Denver Crude Death Rate equal to 150.744 (for 100,000 people). The Bühlmann-Straub credibility-weighted approach is around 151.271.

Model fitting and diagnostic were done. Three diagnostic techniques were used in order to make a decision regarding linear model. Thus, all of the diagnostic technoques point to the viability of the linear regression model. Therefore we can conclude that the

model appropriate. Using the values for Crude Death Rate and Population, the total number of deaths were predicted and captured in Table 6.6.

Chapter 7

Conclusion

The purpose of this thesis was to explore and apply the fundamental concepts that define the modern practice of credibility modeling. Three different credibility approaches were applied for Denver County Dataset, in order to predict the number of deaths for the upcoming year, and the model was found to be appropriate. Thus, credibility theory can be useful in different aspects of actuarial science, in order to perform experience rating on a risk or group of risks. Whenever the theory can be applied, a credibility interpretation can give a more intelligible meaning to the resulting model predictions.

REFERENCES

Bühlmann, H., Gisler, A. (2005), *"A Course in Credibility Theory and its Applications,"* Berlin: Springer.

Bühlmann, H., Straub, E. (1972), *"Credibility for Loss Ratios,"*(Translated by C.E. Brooks) ARCH, 1972.2.

Colorado Department of Public Health and Environment (2014),"Colorado Health Information Dataset," *www.cohid.dphe.state.co.us/scripts/htmsql.exe/mortalityPub.hsql* (Accessed 10/01/2014).

Dean, C.S. (2005), *"Topics in Credibility Theory,"* Study Note of the Society of Actuaries, 4-17.

Hanneman, R. (2003), "Linear Models: Analyzing data with random effects" (Littell, chapter 4), *http://faculty.ucr.edu/∼hanneman/linear_ models /c4.html* (Accessed 11/12/2013).

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

Hickman, J.C., Heacox, L. (1999), *"Credibility Theory: The Cornerstone of Actuarial Science,"* North American Actuarial Journal, 2.

Klugman, S., Panjer, H., Willmot, G. (1998), *"Loss Models: From Data to Decisions,"* New York: Wiley.

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

Mayerson, A.L. (1965), *"A Bayesian View of Credibility,"* PCAS, 51 , 85-104.

Mahler, H. (1986), *"An Actuarial Note on Credibility Parameters,"* Proceedings of the Casualty Actuarial Society, LXXIII.

Norberg, R. (2004), *"Credibility Theory,"* Encyclopedia of Actuarial Science: Wiley.

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

Philbrick, S. (1981), "An Examination of Credibility Concepts," *Proceedings of the Casualty Actuarial Society*, LXVIII, 195-212.

Venter, G. (1989), *"Credibility,"* Chapter 8 of Foundations of Casualty Actuarial Science. New York: Casualty Actuarial Society.

Appendix 1

Linear Regression Diagnostic Summary for Denver County:

```
Call:
lm(formula = population ~ year, data = denyear)

Residuals:
   Min     1Q Median     3Q    Max
 -5769.8  -1021.2  466.9   1879.2   6775.9

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.562e+05  1.356e+03  410.26 2e-16 ***
Record.Year  3.614e+01  1.438e+00   25.14 4.54e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3605 on 11 degrees of freedom
Multiple R-squared: 0.9829,      Adjusted R-squared: 0.9813
F-statistic: 632 on 1 and 11 DF,  p-value: 4.538e-11
```

Appendix 3

Excel Calculation Sheet:

1	Data from: www.cohid.dphe.state.co.us/scripts/htmsql.exe/mortalityPub.hsql										a=	128,7470678	
2											k=	0,426775891	
3	Denver County (Deaths by Cancer)												
4											Z=	0,968214567	
5	Record Year	Total Deaths	Total Population	Total Crude Death Rate per 100,000			Buhlmann-Straub Metod						
6											Credibility premium 1=		152,087
7	#REF!	909	556 738	163,3		X1=	152,0380657				Credibility premium 2=		136,297
8	#REF!	919	563 300	163,1		X2=	136,1433452						
9	#REF!	947	559 090	169,4		X=	138,0764351						
10	#REF!	906	560 348	161,7								Bayesian Method	
11	#REF!	914	560 230	163,1		v=	91825379,36						
12	#REF!	859	559 459	153,5							x1=	152	
13	#REF!	828	562 862	147,1		a1=	1579281619				x2=	136,1433452	
14	#REF!	808	570 437	141,6		a2=	13229044,1						
15	#REF!	858	581 903	147,4		a=	119,3798741				$\pi(\theta_1)$	0,121618362	
16	#REF!	846	595 573	142							$\pi(\theta 2)$=	0,878381638	
17	#REF!	897	604 875	148,3		k=	769186,4316						
18	#REF!	870	620 917	140,1							$f(x, \theta_1)$	3,90311E-16	
19	#REF!	888	634 619	139,9		Z1=	0,907321771		Z2=	0,98605	$f(x, \theta 2)$	2,89212E-15	
20	All	11 449	7 530 351	152									
21				1980,5		Credibility premium 1=		150,744			F(x)=	3,28243E-15	
22						Credibility premium 2=		136,17					
23	Non-Denver County (Deaths by Cancer)										Posterior probabilities		
24													
25	Record Year	Total Deaths	Total Population	Total Crude Death Rate per 100,000		μ(weighted)=		143,76			$\pi(\theta 1,x)$=	0,118909024	
26											$\pi(\theta 2,x)$=	0,881090976	
27	2000	4 987	3 782 063	131,8592525									
28	2001	5 215	3 881 213	134,3652101		Credibility premium 1=		151,271			C20=	7339954,718	
29	2002	5 425	3 945 619	137,494269		Credibility premium 2=		136,25					
30	2003	5 494	3 994 736	137,5309908							Bayesian premium 1=		155,98
31	2004	5 271	4 048 581	130,1937642									
32	2005	5 508	4 103 075	134,2407828			Buhlmann Metod				C40=	55798324,59	
33	2006	5 695	4 182 798	136,1528814									
34	2007	5 782	4 251 347	136,0039536							Bayesian premium 2=		132,701
35	2008	5 851	4 320 035	135,4387175		X1=	152,3461538						
36	2009	6 092	4 381 280	139,0461235		X2=	136,0382604						
37	2010	6 132	4 445 108	137,9494042		X=	144,1922071						
38	2011	6 167	4 497 609	137,1172994									
39	2012	6 426	4 554 064	141,1047363									
40	All	74 045	54 387 528	136,1433452		V1=	102,0532544						
41			Sum:	1768,497385		V2=	7,839034806						
42						V=	54,94614462						

Figure 7.1: Excel Calculation Sheet