ACTSC432 — Loss Models II

Classnotes for Spring 2019

by

Johnson Ng

BMath (Hons), Pure Mathematics major, Actuarial Science Minor

University of Waterloo

Table of Contents

List of Definitions

List of Theorems

List of Procedures

Preface

For this set of notes, I shall follow the format of which the course is presented, by breaking contents into modules instead of lectures. Also, I will be relying on the standard textbook for this topic, namely Klugman et al. 2012.

1 Introduction and Review of Probability

We shall first take an overview of what this course is about, and we will review on some of the relevant notions from earlier courses.

1.1 Introduction to Credibility Theory

Credibility Theory is a form of statistical inference that

- uses newly observed past events; to
- more accurately re-forecasts uncertain future events.

From Klugman et al. 2012,

> *It is a* ***set of quantitative tools*** *that allows an insurer to perform prospective* ***experience rating*** *(adjust future premiums based on past experience) on a risk or group of risks. If the experience of a policyholder is consistently better than that assumed in the underlying manual rate (also called a* ***pure premium****), then the policyholder may demand a rate reduction.*

That's all fancy mumbo-jumbo so let's go through an example that will hopefully enlighten us.

Example 1.1.1 (Enlightening Example to Credibility Theory)

Suppose automobile insurance policies are classified according to the following factors:

- number of drivers;

- gender of each driver;
- number of vehicles; and
- brand, model, production year, and approximate mileage driver per year.

Policies with identical characteristics are assumed to belong to the same **rating class**, which represents a group of individuals with similar risks.

Suppose there are 2 policies in the same rating class. Both policies are charged with a so-called **manual premium** of $\$1,500$ per year. This is the premium specified in the insurance manual for a policy with similar characteristics.

Let's say that after 3 years, we obtain the following data: We want

	Policy 1	Policy 2
Year 1	0	500
Year 2	200	4000
Year 3	0	2500

Table 1.1: Newly acquired past history for finding 'credibility'

to find out what's a good premium to charge to each policy for Year 4.

Remark 1.1.1

The shall leave the following as remarks.

- ***How is the policyholder's own experience account for?*** *This is a key question that will be addressed in this course.*
- *Risks in a given rating class are* ***not perfectly identical*** *(i.e., no rating system is perfect)*
- *One may refine the rating system by incorporating more factors but it is time-consuming (and no system is perfect).*

Thus, credibility theory is designed such that it

- accounts for heterogeneity within a given rating lass; and
- provides a theoretical justification to charge a premium that reflects to the policyholder's own experience.

1.2 Review of Probability

You are expected to be familiar with the following concepts:

Some examples or more detailed review will be added for each topic if I come to work through them in detail.

- Joint and Marginal Distribution
- Conditional Distribution
- Mixture Distributions (see also ACTSC431)
  - n-point Mixture
- Conditional Expectation

2 Review of Statistics

In this chapter, we will review the following notions:

- Unbiased estimation
- Maximum likelihood estimation
- Bayesian estimation ★

2.1 Unbiased Estimation

Suppose we are given a **parametric model** [1] of X, i.e. the distribution of $X \mid \Theta = \theta$ is known but θ is unknown. Furthermore, we have a **random sample** of X, i.e. we have $\{X_i\}_{i=1}^n$ is an independent and identically distributed (iid) sequence of random variables (rv) such that $X_i \sim X$.

[1] See ACTSC431.

Definition 1 (Estimate)

An ***estimate*** *is a specific value that is obtained when applying an estimation procedure to a set of numbers, and in our case, rvs. We usually denote an estimate by a hat* $\hat{}$.

Definition 2 (Estimator)

An ***estimator*** *is a rule or formula that produces an* *estimate*. *We usually denote an estimator by* $\tilde{}$.

Note 2.1.1

An estimate is a number or a function, while an estimator is an rv or a random function.

Remark 2.1.1

In this course, we will not make a difference between the estimator and the estimate, and will use only $\hat{}$.

Definition 3 (Biased and Unbiased Estimator)

We say that an estimator, $\hat{\theta}$, is ***unbiased*** *if*

$$E[\hat{\theta} \mid \theta] = \theta$$

for all θ. We say that an estimator is ***biased*** *if it is not unbiased, and we define the* ***bias*** *as*

$$\text{bias}_{\hat{\theta}}(\theta) = E[\hat{\theta} \mid \theta] - \theta.$$

Let's have ourselves a silly example.

Example 2.1.1

Let $(X_1, \ldots, X_n)$ be a random sample of $\text{Exp}(\beta)$. The **sample mean**

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

is an unbiased estimator for the mean β; observe that by the **linearity of the expectation**, we have

$$E[\overline{X}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n}(n\beta) = \beta.$$

Example 2.1.2

Let $\{X_i\}_{i=1}^{n}$ be a random sample of $X \sim \text{Unif}(0, \theta)$. Let us construct two unbiased estimators for θ using

1. the sample mean $\overline{X}$; and

2. order statistics $X_{(n)} := \max_{1 \leq i \leq n}\{X_i\}$.

Solution

1. Observe that

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} \cdot n\left(\frac{\theta}{2}\right) = \frac{\theta}{2}.$$

This tells us that if we picked $\hat{\theta} = 2\overline{X}$, then we would end up with

$$E[2\overline{X}] = \theta.$$

Thus $2\overline{X}$ is an unbiased estimator of θ.

2. Using the Darth Vader rule [2], since the X_i's form a random sample of X, and the bounds for each X_i is 0 and θ, we have that

$$\begin{aligned}
E[X_{(n)}] &= \int_0^\infty \overline{F}_{X_{(n)}}(x)\,dx \\
&= \int_0^\infty (1 - P(\max\{X_1, X_2, \ldots, X_n\}) \leq x)\,dx \\
&= \int_0^\infty (1 - P(X_1 \leq x)P(X_2 \leq x)\ldots P(X_n \leq x))\,dx \\
&= \int_0^\theta \left(1 - (\frac{x}{\theta})^n\right)dx \\
&= \theta - \frac{1}{n+1}\left(\frac{x^{n+1}}{\theta^n}\right)\Big|_{x=0}^{x=\theta} = \frac{n}{n+1}\theta,
\end{aligned}$$

where we note that we can change the bounds as such since $X \sim \text{Unif}(0, \theta)$ implies that

$$P(X \leq \theta) = \begin{cases} \frac{x}{\theta} & 0 \leq x \leq \theta \\ 1 & x > \theta \end{cases}.$$

Thus, to get an unbiased estimator for θ, we simply need to consider

$$\hat{\theta} = \frac{n+1}{n}X_{(n)},$$

which then

$$E\left[\frac{n+1}{n}X_{(n)}\right] = \theta.$$

[2] The Darth Vader rule is given as: if X is a **non-negative** rv, then

$$E[X] = \int_0^\infty \overline{F}_X(x)\,dx,$$

where $\overline{F}_X$ is the survival function of X.

Proposition 1 (Sample Mean as the Unbiased Estimator of the Mean)

Let $\{X_i\}_{i=1}^n$ be a random sample of X which has mean μ. Then $\overline{X}$ is an unbiased estimator of μ.

Proof

We have that

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}(n\mu) = \mu. \qquad \square$$

Definition 4 (Sample Variance)

Let $\{X_i\}_{i=1}^n$ be a random sample of X which has mean μ and variance σ^2. We define the ***sample variance*** *as*

$$\hat{\sigma}^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Proposition 2 (Sample Variance as the Unbiased Estimator of the Variance)

Let $\{X_i\}_{i=1}^n$ be a random sample of X which has mean μ and variance σ^2. Then the sample variance $\hat{\sigma}^2$ is an unbiased estimator of σ^2.

Proof

First, note that

$$\begin{aligned}
\operatorname{Var}(\overline{X}) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} n \operatorname{Var}(X_i) \\
&= \frac{1}{n}\sigma^2.
\end{aligned}$$

Thus

$$\begin{aligned}
E\left[\sum_{i=1}^{n}(X_i-\overline{X})^2\right] &= E\left[\sum_{i=1}^{n}(X_i-\mu+\mu-\overline{X})^2\right] \\
&= \sum_{i=1}^{n} E\left[(X_i-\mu)^2\right] + \sum_{i=1}^{n} E\left[(\mu-\overline{X})^2\right] \\
&\quad + 2E\left[\sum_{i=1}^{n}(X_i-\mu)(\mu-\overline{X})\right] \\
&= n\sigma^2 + n\operatorname{Var}(\overline{X}) \;^{3} + 2nE[(\overline{X}-\mu)(\mu-\overline{X})] \;^{4} \\
&= n\sigma^2 - n\operatorname{Var}(\overline{X}) \\
&= n\sigma^2 - n\left(\frac{1}{n}\sigma^2\right) \\
&= (n-1)\sigma^2.
\end{aligned}$$

It follows that

$$E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right] = \sigma^2. \qquad \square$$

[4] This relies on the fact that $\overline{X}$ is the unbiased estimator for μ (cf. Proposition 1). We then use the definition of the variance to achieve this.

[4] We used the fact that

$$\sum_{i=1}^{n}(X_i-\mu) = \sum_{i=1}^{n} X_i - n\mu = n\overline{X} - n\mu.$$

Also, note that

$$\operatorname{Var}(\overline{X}) = E[(\overline{X}-\mu)^2].$$

Remark 2.1.2

In general, unbiasedness is ***not preserved*** *under parameter transformations. E.g.,* $\frac{1}{\overline{X}}$ *is generally not unbiased for* μ*, where* μ *is the mean of* $\overline{X}$.

Some unbiased estimators can also be unreasonable.

Example 2.1.3

Consider $X \sim \text{Poi}(\lambda)$, where $\lambda > 0$. Note that

$$E[(-1)^X] = e^{\lambda(-1-1)} = e^{-2\lambda}$$

by the **probability generating function** method, and we see that $(-1)^X$ is an unbiased estimator of $e^{-2\lambda}$. However, we see that $(-1)^x$ only takes on values ± 1, which is nowhere close to $e^{-2\lambda}$.

Intuitively, $e^{-2\overline{X}}$ would be a "better" estimator despite the fact that it is biased.

Despite shortcomings like the above, unbiasedness is generally a good property for an estimator to have.

2.2 *Mean Squared Error*

Definition 5 (Mean Squared Error)

Suppose $\hat{\theta}$ is an estimator for the parameter θ. The ***mean squared error (MSE)*** *of $\hat{\theta}$ is defined as*

$$\mathrm{MSE}_{\hat{\theta}}(\theta) := E\left[(\hat{\theta}-\theta)^2\right] = \mathrm{Var}(\hat{\theta}) + \mathrm{bias}_{\hat{\theta}}(\theta)^2.$$

Proof

It is not immediately clear how the two expressions are the same. We shall prove it here. First, note that $\mathrm{bias}_{\hat{\theta}}(\theta) = E[\hat{\theta}] - \theta$ is a real value. Using a similar idea as in Proposition 2, we see that

$$\begin{aligned}
E\left[(\hat{\theta}-\theta)^2\right] &= E\left[\left(\hat{\theta}-E[\hat{\theta}]+E[\hat{\theta}]-\theta\right)^2\right] \\
&= E\left[(\hat{\theta}-E[\hat{\theta}])^2\right] + E\left[\left(E[\hat{\theta}]-\theta\right)^2\right] \\
&\quad + 2E[(\hat{\theta}-E[\hat{\theta}])(E[\hat{\theta}]-\theta)] \\
&= \mathrm{Var}(\hat{\theta}) + \mathrm{bias}_{\hat{\theta}}(\theta)^2 \\
&\quad + 2\,\mathrm{bias}_{\hat{\theta}}(\theta)\, \cancelto{0}{E[\hat{\theta}-E[\hat{\theta}]]} \\
&= \mathrm{Var}(\hat{\theta}) + \mathrm{bias}_{\hat{\theta}}(\theta)^2.
\end{aligned}$$

□

Note 2.2.1

The MSE is a measure to evaluate the quality of estimators. The smaller the MSE, the better the estimator.

2.3 Maximum Likelihood Estimation

Definition 6 (Likelihood Function)

*Let $\{X_i\}_{i=1}^n$ be a random sample of X with density $f(x;\underline{\theta})$, where $\underline{\theta}$ is possibly a vector of parameters. The **likelihood function** for $\underline{\theta}$ is defined as*

$$L(\underline{\theta}) = \prod_{i=1}^{n} f(X_i;\underline{\theta}).$$

Definition 7 (Maximum Likelihood Estimation)

*The **maximum likelihood estimation (MLE)** of $\hat{\underline{\theta}}$ of $\underline{\theta}$ is an approach that maximizes $L(\hat{\underline{\theta}})$.*

Note 2.3.1

*Heuristically, under the MLE, $\hat{\underline{\theta}}$ is the **most likely parameter** for the sample $(X_1, \ldots, X_n)$ to be realized.*

Sometimes, the likelihood function is difficult to work with. Fortunately, since $\ln x$ is a increasing bijective function that preserves monotonicity, we can make us of this property to ensure maximality.

Definition 8 (Log-likelihood Function)

*The **log-likelihood function** is defined as*

$$l(\underline{\theta}) = \sum_{i=1}^{n} \ln(f(X_i;\underline{\theta})).$$

Example 2.3.1

Let $\{X_i\}_{i=1}^n$ be a random sample for $\mathrm{N}(\mu, v)$. Find the MLE for μ, v.

Solution

First, we shall work on getting an MLE for μ. The likelihood function here is

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} f(X_i; \mu) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\ &\propto \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \right\}. \end{aligned}$$

Evaluating the derivative and equating it to 0 would be fruitless, since this is an exponentiation. Thus we appeal to the log-likelihood, which is

$$l(\mu) \propto \sum_{i=1}^{n} (X_i - \mu)^2.$$

The derivative log-likelihood is thus

$$\frac{dl}{d\mu} \propto -2 \sum_{i=1}^{n} (X_i - \mu).$$

Equating the above to 0, we get

$$\hat{\mu} = \overline{X}.$$

Now for an MLE of σ^2. For sanity, let us denote $\tau = \sigma^2$. Then the likelihood function, focusing on τ, is

$$\begin{aligned} L(\tau) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(X_i-\mu)^2}{2\tau}} \\ &\propto \tau^{-\frac{n}{2}} e^{-\frac{1}{2\tau} \sum_{i=1}^{n} (X_i-\mu)^2}. \end{aligned}$$

Again, the likelihood involves an exponentiation, so we appeal to the log-likelihood, which is

$$l(\tau) \propto -\frac{n}{2} \ln \tau - \frac{1}{2\tau} \sum_{i=1}^{n} (X_i - \mu)^2.$$

The derivative of the log-likelihood is

$$\frac{dl}{d\tau} = -\frac{n}{2\tau} + \frac{1}{2\tau^2} \sum_{i=1}^{n} (X_i - \mu)^2.$$

Equating the above to 0, we get

$$n = \frac{1}{\hat{\tau}} \sum_{i=1}^{n} (X_i - \hat{\mu})^2,$$

and so

$$\hat{\sigma}^2 = \hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

2.4 Bayesian Estimation

From Klugman et al. 2012,

> *The Bayesian approach assumes that only the data actually observed are relevant and it is the population distribution that is variable.*

Definition 9 (Prior Distribution)

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

Note 2.4.1

- *The parameter θ may be scalar or vector valued.*
- *Determining the prior distribution has always been one of the barriers to the widespread acceptance of the Bayesian methods, since it is almost certainly the case that your experience has provided you with some insight about possible parameter values before the first data point has been observed.*

We shall use the following concepts from multivariate statistics to obtain the following definitions.

Definition 10 (Joint Distribution)

Let $\{X_i\}_{i=1}^n$ be a random sample of the rv X, and Θ another rv that is independent of the X_i's [5], with pdf π. Let $\vec{X} = (X_1, X_2, \ldots, X_n)$. Then the ***joint distribution*** *of $\vec{X}$ and Θ is defined as*

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

[5] Note that Θ does not necessarily have a similar distribution to X.

Definition 11 (Marginal Distribution)

Let $\{X_i\}_{i=1}^n$ be a random sample of the rv X, and Θ another rv that is independent of the X_i's [6], with pdf π. Let $\vec{X} = (X_1, X_2, \ldots, X_n)$. Then the ***marginal distribution*** *of $\vec{X}$ is defined as*

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

[6] Note that Θ does not necessarily have a similar distribution to X.

Once we have obtained data, we can look back at our prior distribution and "update" it to...

Definition 12 (Posterior Distribution)

Let $\{X_i\}_{i=1}^n$ be a random sample of the rv X, and Θ another rv that is independent of the X_i's [7], with pdf π. The ***posterior distribution****, denoted by $\pi_{\Theta|\vec{X}}(\theta \mid \vec{x})$, is the conditional probability distribution of the parameters given the observed data.*

[7] Note that Θ does not necessarily have a similar distribution to X.

It is easy to find out what the general formula of the posterior distribution is. One simply needs to make use of Definition 10 and Definition 11. The proof of the following proposition is left as an easy brain exercise for the reader.

Exercise 2.4.1
Prove Proposition 3.

Proposition 3 (Formula for the Posterior Distribution)

With the assumptions in Definition 12, we have that the posterior distribution can be computed as

$$\begin{aligned}\pi_{\Theta \mid \vec{X}}(\theta \mid \vec{x}) &= \frac{f_{\vec{X},\Theta}(\vec{x},\theta)}{f_{\vec{X}}(\vec{x})} \\ &= \frac{\left(\prod_{i=1}^{n} f_{X_i \mid \Theta}(x_i \mid \theta)\right)\pi(\theta)}{\int_{\forall\theta}\left(\prod_{i=1}^{n} f_{X_i \mid \Theta}(x_i \mid \theta)\right)\pi(\theta)\,d\theta}.\end{aligned}$$

Definition 13 (Posterior Mean)

*The **posterior mean** is defined as the expected value of the posterior distribution.*

Definition 14 (Bayes Estimator)

*The **Bayes estimator** of Θ is the posterior mean of Θ, defined as*

$$\hat{\theta}_B := E[\Theta \mid \vec{X} = \vec{x}] = \int_{\forall\theta} \theta \cdot \pi_{\Theta \mid \vec{X}}(\theta \mid \vec{x}).$$

Note 2.4.2

It can be shown that $\hat{\theta}_B$ minimizes the mean square error

$$\min_{\hat{\theta}} E\left[\left(\hat{\theta} - \Theta\right)^2 \mid \vec{X} = \vec{x}\right].$$

2.4.1 Conjugate Prior Distributions and the Linear Exponential Family

Definition 15 (Conjugate Prior Distribution)

A prior distribution is said to be a ***conjugate prior distribution*** *for a given model if the resulting posterior distribution is from the same family as the prior, although possibly with different parameters.*

8

[8] More examples should be added here.

Example 2.4.1

The following are some important/prominent examples of conjugate prior distributions:

$\pi(\theta)$	$f(x \mid \theta)$	$\pi(\theta \mid \vec{x})$
Gamma	Poisson	Gamma
Normal	Normal	Normal
Beta	Binomial	Beta
Beta	Geometric	Beta

Table 2.1: Important/Prominent Conjugate Prior Distributions

Definition 16 (Linear Exponential Family)

An rv X is said to belong to the ***linear exponential family*** *if its pdf is of the form*

$$f(x, \theta) = \frac{p(x) e^{x r(\theta)}}{q(\theta)},$$

where $p(x)$ is some function of x, and $r(\theta), q(\theta)$ are some functions of θ, and the support of f does not depend on θ.

Example 2.4.2

Some members of the linear exponential family include

- $\mathrm{Exp}(\theta) : f(x, \theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, where $p(x) = 1$, $r(\theta) = -\frac{1}{\theta}$ and $q(\theta) = \theta$.
- $\mathrm{Gam}(\alpha, \theta) : f(x, \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}$.
- $\mathrm{Poi}(\theta) : f(x, \theta) = \frac{\theta^x e^{-\theta}}{x!} = \frac{\frac{1}{x!} e^{x \ln \theta}}{e^\theta}$
- $\mathrm{N}(\theta, v) : f(x, \theta, v) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x-\theta)^2}{2v}} = \frac{(2\pi v)^{-\frac{1}{2}} e^{-\frac{x^2}{2v}} e^{x\frac{\theta}{v}}}{e^{\theta^2/2v}}$

Note 2.4.3

Basically, functions the belong to a linear exponential family is a linear-like function with an exponent.

Theorem 4 (Conjugate Prior Distributions of Linear Exponential Distributions)

Suppose that given $\Theta = \theta$ the rvs $\vec{X}$ are iid with pf

$$f_{X_j \mid \Theta}(x_j \mid \theta) = \frac{p(x_j)e^{r(\theta)x_j}}{q(\theta)},$$

where Θ has the pdf

$$\pi(\theta) = \frac{[q(\theta)]^{-k}e^{\mu k r(\theta)}r'(\theta)}{c(\mu,k)},$$

where μ and k are parameters of the distribution and $c(\mu,k)$ is the normalizing constant [9]. Then the posterior pf $\pi_{\Theta \mid \vec{X}}(\theta \mid \vec{x})$ is of the same form as $\pi(\theta)$, i.e. $\pi(\theta)$ is a conjugate prior distribution function.

[9] The normalizing constant is used to reduce any probability function to a probability density function with a total probability of 1. (Source: Wikipedia)

Proof

Notice that the posterior distribution is

$$\begin{aligned}
\pi(\theta \mid \vec{x}) &= \frac{\left(\prod_{i=1}^n f_{X_i \mid \Theta}(x_i \mid \theta)\right)\pi(\theta)}{\int_{\forall\theta}\left(\prod_{i=1}^n f_{X_i \mid \Theta}(x_i \mid \theta)\right)\pi(\theta)\,d\theta} \\
&\propto \left(\prod_{i=1}^n f_{X_i \mid \Theta}(x_i \mid \theta)\pi(\theta)\right) \\
&= \left(\prod_{i=1}^n \frac{p(x_j)e^{r(\theta)x_j}}{q(\theta)}\right)\left(\frac{[q(\theta)]^{-k}e^{\mu k r(\theta)}r'(\theta)}{c(\mu,k)}\right) \\
&\propto q(\theta)^{-(n+k)}e^{\mu k + n\bar{x} r(\theta)}r'(\theta) \\
&= q(\theta)^{-k^*}e^{\mu^* k^* r(\theta)}r'(\theta),
\end{aligned}$$

where

$$k^* = k + n, \text{ and } \mu^* = \frac{\mu k + \sum x_j}{k+n} = \frac{k}{k+n}\mu + \frac{n}{k+n}\bar{x},$$

and we see that the posterior distribution has the same form as $\pi(\theta)$. $\square$

Example 2.4.3

One non-example is mentioned in Example 2.4.1: the distribution of X_i is not from the linear exponential family, but we still obtain that the posterior distribution has a similar distribution to the posterior distribution.

3 Limited Fluctuation Credibility Theory

The Limited Fluctuation Credibility Theory provides a mechanism for assigning **full** or **partial credibility** to a policyholder's experience. The difficulty with this approach is its lack of a sound underlying mathematical theory that justifies the use of these methods. Despite that fact, it is still widely used today, especially in the United States.

3.1 Limited Fluctuation Credibility

From Klugman et al. 2012,

> *This branch of credibility theory represents the first attempt to quantify the credibility problem.*

This approach is also known as the "**American credibility**". It was first proposed by Mowbray in 1914 [1].

[1] Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society*, I:24–30

The problem can be formulated as follows. Suppose that $\{X_i\}_{i=1}^n$ represents a policyholder's claim amounts in the past n years. Furthermore, we assume that the X_i's have

- the same expected value, i.e. $E[X_i] = \mu$ for some μ; and
- variance, i.e. $\mathrm{Var}(X_i) = \sigma^2$ for some σ.

From our revision in the last section, we know that $\overline{X}$ is an unbiased estimator for μ, and if the X_i's are independent, then $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$.

The goal here is to figure our how much to charge for the next premium, i.e. determining $E[X_{n+1}]$. We have at least the following 3 possibilities:

- ignore past data (no credibility) and charge M, a value, called the **manual premium** [2], obtained from experience on other similar but non-identical policyholders;
- ignore M and charge $\overline{X}$ (full credibility); and a third possibility is to
- choose some combination of M and $\overline{X}$ (partial credibility).

[2] This name is obtained from the fact that it usually comes from a book (manual) of premiums.

From the POV of an insurer, it seems sensible to favor $\overline{X}$ if the experience is "stable", i.e. there is little fluctuation, represented by a small σ^2. Stable values imply that $\overline{X}$ is more reliable as a predictor. Conversely, if $\overline{X}$ is volatile, then M would be a safer choice.

3.2 *Full Credibility*

In **full credibility** theory, there are only 2 outcomes: either we

- assign full credibility, that is to charge $\overline{X}$; or
- no credibility, where we charge M.

One method to 'quantify the stability' of $\overline{X}$ [3] is to infer that $\overline{X}$ is stable if the difference between $\overline{X}$ and μ is small relative to μ with high probability, i.e.

[3] This has become the standard method for 'quantifying stability' for $\overline{X}$.

$$P(|\overline{X}-\mu| \leq \varepsilon\mu) \geq p \tag{3.1}$$

for some $\varepsilon > 0$ and $0 < p < 1$. We may rewrite Equation (3.1) as

$$P\left(\frac{|\overline{X}-\mu|}{\sigma/\sqrt{n}} \leq \frac{\varepsilon\mu}{\sigma/\sqrt{n}}\right) \geq p.$$

Now let y_p be defined as by

$$y_p = \mathrm{VaR}_p\left(\frac{|\overline{X}-\mu|}{\sigma/\sqrt{n}}\right) = \inf\left\{y \in \mathbb{R} : P\left(\frac{|\overline{X}-\mu|}{\sigma/\sqrt{n}} \leq y\right) \geq p\right\}.$$

If $\overline{X}$ is continuous, then the $\geq$ sign above can be replaced with an "$=$" sign [4], and y_p satisfies

[4] See ACTSC431.

$$P\left(\frac{|\overline{X}-\mu|}{\sigma/\sqrt{n}} \leq y_p\right) = p. \tag{3.2}$$

Then the condition for full credibility is

$$y_p \leq \frac{\varepsilon\mu}{\sigma/\sqrt{n}}.$$

Making n the subject, we have that the number of exposure required for full credibility is thus

$$n \geq \left(\frac{y_p}{\varepsilon}\right)^2 \frac{\sigma^2}{\mu^2} = \lambda_0 \frac{\sigma^2}{\mu^2}, \tag{3.3}$$

where we let $\lambda_0 = \left(\frac{y_p}{\varepsilon}\right)^2$ for notational succinctness since it is a constant that depends only p and ε.

(Condition for Full Credibility)

1. *Use the central limit theorem argument for y_p.*
2. *Calculate RHS of Equation (3.3).*

It is often difficult to identify a distribution for $\overline{X}$, of which y_p depends on. Recall the **normal approximation**, which is applicable if n is large [5]:

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx Z_{0,1} \sim \mathrm{N}(0,1)$$

[5] **Is this not circular!?**

Then Equation (3.2) becomes

$$\begin{aligned} p &= P(|Z| \leq y_p) = \Phi(y_p) - \Phi(-y_p) \\ &= \Phi(y_p) - 1 + \Phi(y_p) = 2\Phi(y_p) - 1. \end{aligned}$$

Thus

$$y_p \approx \Phi^{-1}\left(\frac{1+p}{2}\right).$$

Example 3.2.1

Suppose that one has data $\{X_i\}_{i=1}^{10}$ on the claim amounts in the last 10 periods, where

$$X_i = 0 \text{ for } i = 1, \ldots, 6,$$

and

$$X_7 = 253,\ X_8 = 398,\ X_9 = 439,\ X_{10} = 756.$$

Determine the condition for full credibility with $\varepsilon = 0.05$ and $p = 0.9$.

Solution

We need to first determine the sample mean and sample variance, and we shall use the unbiased estimators of μ and σ^2 respectively: they are

$$\overline{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = \frac{0 + 253 + 398 + 439 + 756}{10} = 184.6,$$

and

$$\hat{\sigma}^2 = \frac{1}{10-1}\sum_{i=1}^{10}(X_i - \overline{X})^2 = 267.89^2.$$

We also need

$$y_p = \Phi^{-1}\left(\frac{1+p}{2}\right) = \Phi^{-1}(.95) = 1.645.$$

Then we require that

$$n \geq \left(\frac{1.645}{0.05}\right)^2 \left(\frac{267.89^2}{184.6^2}\right) = 2279.5.$$

We see that the 10 observations definitely do not deserve full credibility.

Full credibility is sometimes given on a number of claims basis (instead of on the claims amount).

Example 3.2.2

Suppose that one has iid data $\{N_i\}_{i=1}^n$ on the number of claims in the past n periods, with $N_i \sim \text{Poi}(\lambda)$. Determine the condition for full credibility in terms of the **expected total number of claims** given that $p = 0.9$ and $\varepsilon = 0.05$.

Solution

Since $N_i \sim \text{Poi}(\lambda)$, we have $E[N_i] = \lambda = \text{Var}(N_i)$. Furthermore,

$$y_p = \Phi^{-1}(0.95) = 1.645.$$

Now since the condition is

$$n \geq \lambda_0 \frac{\sigma^2}{\mu^2} = \frac{\lambda_0}{\lambda},$$

and we want the expected total number of claims, we focus on looking at

$$n\mu = n\lambda \geq \lambda_0.$$

Observe that

$$\lambda_0 = \left(\frac{1.645}{0.05} \right)^2 = 1082.41,$$

we have that the required expected total number of claims should fulfill

$$n\lambda \geq 1082.41.$$

Example 3.2.3 (Compound Poisson for Full Credibility)

Let $\{X_i\}_{i=1}^n$ be a sequence of iid compound Poisson rvs, given by

$$X_i = \sum_{j=1}^{N_i} Y_{i,j} = \begin{cases} \sum_{j=1}^{N_i} Y_{i,j}, & N_i \geq 0 \\ 0 & N_i = 0 \end{cases},$$

where

- $\{N_i\}_{i=1}^n$ are iid with $N_i \sim \text{Poi}(\lambda)$ for each i; and
- $\{Y_{i,j}\}$ are also iid with mean μ_Y and variance σ_Y^2.

Determine the condition for full credibility.

Solution

We require the unconditional sample mean and sample variance of X_i; they are

$$E[X_i] = E[E[X_i \mid N_i]] = E[N_i]E[Y_{i,j}] = \lambda\mu_Y,$$

and

$$\begin{aligned} \text{Var}(X_i) &= \text{Var}(E[X_i \mid N_i]) + E[\text{Var}(X_i \mid N_i)] \\ &= \text{Var}(N_i\mu_Y) + E[N_i\sigma_Y^2] \\ &= \mu_Y^2\lambda + \sigma_Y^2\lambda \\ &= \lambda(\mu_Y^2 + \sigma_Y^2). \end{aligned}$$

Thus, the condition for full credibility is

$$n \geq \lambda_0 \frac{\lambda(\mu_Y^2 + \sigma_Y^2)}{\lambda^2\mu_Y^2} = \frac{\lambda_0}{\lambda} \left(1 + \frac{\sigma_Y^2}{\mu_Y^2} \right).$$

To further illustrate that we can use the concept of full credibility for different things, the following example is provided.

Example 3.2.4

Suppose that the average claim size for a group of insureds is 1500 with a standard deviation of 7500. Furthermore, assume that claim counts have a Poisson distribution. For $\varepsilon = 0.06$ and $p = 0.9$, determine the standard for full credibility based on the

1. total claim amount; and

2. total number of claims,

in terms of the expected total number of claims.

Solution

1. Using the last example and letting

$$E[X_i] = \mu \text{ and } \operatorname{Var}(X_i) = \operatorname{Var}(X_i) = \sigma^2,$$

the standard for full credibility is

$$n \geq \frac{\lambda_0}{\lambda}\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right).$$

We are given that

$$\mu_Y = 1500 \text{ and } \sigma_Y^2 = 7500^2.$$

Thus

$$n \geq \frac{1.645^2}{0.06^2\lambda}\left(1 + \frac{7500^2}{1500^2}\right) = \frac{19543.51}{\lambda}.$$

In terms of the expected total number of claims, we have

$$n\lambda \geq 19543.51.$$

Thus the observed total number of claims of past claims must be at least 19544 to assign full credibility.

2. Using Example 3.2.2, we have

$$n \geq \frac{\lambda_0}{\lambda} = \frac{751.67}{\lambda}.$$

Thus, in terms of the expected total number of claims, we have

$$n\lambda \geq 751.67.$$

Therefore, the observed total number of past claims must be at least 752 to assign full credibility.

3.3 *Partial Credibility*

If full credibility is inappropriate, then we may want to assign **partial credibility** to the past experience $\overline{X}$ in the net premium. Without much mathematical support, it was suggested that we let the net premium be defined as a weighted average of $\overline{X}$ and the manual premium M, i.e.

$$P = Z\overline{X} + (1 - Z)M,$$

where $Z \in [0, 1]$ is known as the **credibility factor** [6] [7], which is a value that needs to be chosen.

[6] It is important to note there that Z is not an rv. It is simply a pretentious choice of notation for what is to come.

[7] It is interesting to remark that Mowbray 1914 considered full but not partial credibility.

In the actuarial literature [8], there are various suggestions for determining Z. However, they are usually justified on intuition rather than theoretically sound grounds. We shall discuss one of the choices here, which is flawed, but is at least simple.

[8] Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons Inc., Hoboken, New Jersey, 4th edition

Recall that the goal of the full-credibility standard is to ensure that the difference between $\overline{X}$ and μ is small with high probability (cf. beginning of Section 3.2). Since $\overline{X}$ is unbiased, to achieve this standard is basically [9] equivalent to controlling the variance of $\overline{X}$. Note that full credibility fails when

[9] This is exactly the case if $\overline{X}$ is normal.

$$n < \lambda_0 \left(\frac{\sigma^2}{\mu^2} \right), \tag{3.4}$$

and since the sample variance (which is unbiased for the variance) is

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n},$$

rearranging Equation (3.4), we have that

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} > \frac{\mu^2}{\lambda_0}.$$

Thus, we choose Z such that it controls the variance of the credibility premium as such:

$$\begin{aligned}\frac{\mu^2}{\lambda_0} &= \mathrm{Var}(P) = \mathrm{Var}(Z\overline{X} + (1-Z)M) \\ &= Z^2\,\mathrm{Var}(\overline{X}) = Z^2 \cdot \frac{\sigma^2}{n}.\end{aligned}$$

Thus, since we want Z as a weighted average, we let

$$Z = \min\left\{\frac{\mu}{\sigma}\sqrt{\frac{n}{\lambda_0}}, 1\right\}.$$

[10] Note that

$$\frac{\mu}{\sigma}\sqrt{\frac{n}{\lambda_0}} = \sqrt{\frac{n}{\lambda_0(\frac{\sigma^2}{\mu^2})}},$$

which is the square root of the **actual number of exposures** divided by the **number of exposures needed for full credibility**. This is also referred to as the **Square-root rule for partial credibility**.

[10] Note that this choice of Z has some consistency with full credibility, since $Z = 1$ iff $n \geq \lambda_0 \frac{\sigma^2}{\mu^2}$.

Example 3.3.1

Suppose that past observations of the number of claims $\{N_i\}_{i=1}^n$ are iid and $N_i \sim \mathrm{Poi}(\lambda)$. Determine the credibility factor Z based on the number of claims.

Solution

Note that

$$\mu = E[N_i] = \lambda \text{ and } \sigma^2 = \mathrm{Var}(N_i) = \lambda.$$

We have that

$$Z = \min\left\{\frac{\mu}{\sigma}\sqrt{\frac{n}{\lambda_0}}, 1\right\} = \min\left\{\sqrt{\frac{n\lambda}{\lambda_0}}, 1\right\}.$$

Example 3.3.2

Consider the setup in Example 3.2.3. Determine the credibility factor Z based on the amount of claims.

Solution

We have that

$$\mu = E[X_i] = \lambda \mu_Y \text{ and } \sigma^2 = \mathrm{Var}(X_i) = \lambda(\mu_Y^2 + \sigma_Y^2).$$

Then since

$$\frac{\mu}{\sigma}\sqrt{\frac{n}{\lambda_0}} = \sqrt{\frac{n\lambda}{\lambda_0} \cdot \frac{\mu_Y^2}{\mu_Y^2 + \sigma_Y^2}},$$

we have that

$$Z = \min\left\{\sqrt{\frac{n\lambda}{\lambda_0} \cdot \frac{\mu_Y^2}{\mu_Y^2 + \sigma_Y^2}}, 1\right\}$$

Different credibility factors may arise depending on the basis of which the credibility is founded upon.

Example 3.3.3

Consider the setup in Example 3.2.4. Further suppose that

- in thelast year, this group of insureds had 600 claims and a total loss of 15600 ; and
- the prior estimate of the total loss was 16500 (this is M).

Estimate the credibility premium for the group based on the

1. total claim amount; and
2. total number of claims.

Solution

1. We are given that $\mu_Y = 1500$, $\sigma_Y = 7500$ and $n\lambda = 600$. Thus

$$\begin{aligned} Z &= \min\left\{\sqrt{\frac{n\lambda}{\lambda_0} \cdot \frac{\mu_Y^2}{\mu_Y^2 + \sigma_Y^2}}, 1\right\} \\ &= \min\left\{\sqrt{\frac{600}{\left(\frac{1.645}{0.06}\right)^2} \cdot \frac{1500^2}{1500^2 + 7500^2}}, 1\right\} \\ &= 0.17522 \end{aligned}$$

Thus the credibility premium for the group is

$$P = 0.17522\overline{X} + (1 - 0.17522)M$$

$$= 0.17522(15600) + (1 - 0.17522)(16500)$$
$$= 16342.302$$

[11]

[11] It is important to note here that $\overline{X} = 15600$ in this case, since this is the total loss over '1' period of time, in particular it is the total amount up to the latest time.

2. Based on the total number of claims, the credibility factor is

$$Z = \min\left\{\sqrt{\frac{n\lambda}{\lambda_0}}, 1\right\} = \min\left\{\sqrt{\frac{600}{\left(\frac{1.645}{0.06}\right)^2}}, 1\right\} = 0.89343.$$

Thus the credibility premium for the group is

$$P = 0.89343\overline{X} + (1 - 0.89343)M = 15696.$$

3.4 Problems with Limited Fluctuation Credibility

- There is no theoretical model for the distribution of X_i's, and so there is no reason why

$$P = Z\overline{X} + (1 - Z)M$$

is a reasonable and more preferable to M.

- The choice of Z is rather arbitrary.
- There is no guidance to the choices of ε and p.
- The limited fluctuation approach does not examine the difference between μ and M. Furthermore, it is usually the case that M is also an estimate, and hence unreliable in itself.

4 Greatest Accuracy Credibility

The **Greatest Accuary Credibility** approach is a model-based approach to the solution of the credibility problem, which is an outgrowth of Bühlmann's classic paper in 1967 [1]. The greater accuracy credibility is also called the **European credibility**.

[1] Bühmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, 4:199–207

In greatest accuracy credibility, we assume that all risk units in a given rating class have an **unknown risk parameter** θ that is associated with their risk level. Since different insureds have different θ values, risk units within a rating class are **not completely homogeneous**. This assumption allows us to quantify the differences between policyholders wrt to the risk characteristics.

Note 4.0.1 (Assumptions)

We shall also always assume that θ exists, but we shall assume that it is not observable, and that we can never know its true value.

Since θ varies by policyholder, there is a probability distribution Θ across the rating class. We denote

- $\pi_\Theta(\theta)$ *as the probability distribution of Θ; and*
- $\Pi_\Theta(\theta)$ *as the cdf of Θ.*

*If θ is a **scalar parameter** [2], then we may interpret*

$$\Pi(\theta) = P(\Theta \leq \theta)$$

as the percentage of policyholders in the rating class with risk parameter Θ less than or equal to θ.

[2] Refer to STAT330.

Furthermore, if we let $\{X_i\}_{i=1}^n$ be the past exposure units [3]*, we will suppose that*

$$\{X_i \mid \Theta = \theta\}_{i=1}^n$$

are iid, with common density function $f_{X|\Theta}(x \mid \theta)$.

[3] which is not necessarily iid

We want to use these assumptions to derive a rate to cover for X_{n+1}.

4.1 The Bayesian Methodology

Definition 17 (Predictive Distribution)

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

Proposition 5 (Formula for Predictive Distribution)

Given exposure units $\{X_i\}_{i=1}^n$, the predictive distribution of a new observation, Y, can be computed as

$$f_{Y|\vec{X}}(y \mid \vec{x}) = \int_{\forall\theta} f_{Y|\Theta}(y \mid \theta)\pi_{\Theta|\vec{X}}(\theta \mid \vec{x}).$$

Proof

By the formula for the posterior distribution, we have that

$$\begin{aligned}\pi_{\Theta|\vec{X}}(\theta \mid \vec{x}) &= \frac{f_{\Theta,\vec{X}}(\theta, \vec{x})}{f_{\vec{X}}(\vec{x})} \\ &= \frac{f_{\vec{X}|\Theta}(\vec{x} \mid \theta)\pi(\theta)}{\int_{\forall\theta} f_{\vec{X}|\Theta}(\vec{x} \mid \theta)\pi(\theta)\,d\theta}.\end{aligned}$$

Also, observe that

$$f_{Y,\vec{X}}(y, \vec{x}) = \int_{\forall\theta} f_{(Y,\vec{X})|\Theta}(y, \vec{x} \mid \theta)\pi(\theta)\,d\theta$$

$$= \int_{\forall\theta} f_{Y|\Theta}(y \mid \theta) f_{\vec{X}|\Theta}(\vec{x} \mid \theta)\pi(\theta)\,d\theta,$$

where the second equality follows from our assumption that the conditional observations are independent. Then

$$\begin{aligned} f_{Y|\vec{X}}(y \mid \vec{x}) &= \frac{f_{Y,\vec{X}}(y,\vec{x})}{f_{\vec{X}}(\vec{x})} \\ &= \frac{\int_{\forall\theta} f_{Y|\Theta}(y \mid \theta) f_{\vec{X}|\Theta}(\vec{x} \mid \theta)\pi(\theta)\,d\theta}{\int_{\forall\theta} f_{\vec{X}|\Theta}(\vec{x} \mid \theta)\pi(\theta)\,d\theta} \\ &= \int_{\forall\theta} f_{Y|\Theta}(y \mid \theta)\pi_{\Theta|\vec{X}}(\theta \mid \vec{x}). \end{aligned}$$ □

Definition 18 (Individual Premium)

*Given the X_{n+1} exposure unit and risk Θ, we define the **individual premium** (or **hypothetical mean**) of X_{n+1} as*

$$\mu_{n+1}(\theta) = E[X_{n+1} \mid \Theta = \theta].$$

Definition 19 (Pure Premium)

*We define the **pure premium** (or **collective premium**) of X_{n+1} as*

$$\mu_{n+1} = E[X_{n+1}].$$

Definition 20 (Bayesian Premium)

*The **Bayesian premium** of X_{n+1} is defined as*

$$E[X_{n+1} \mid \vec{X}] = \int_{\forall\theta} \mu_{n+1}(\theta)\pi_{\Theta|\vec{X}}(\theta \mid \vec{x})\,d\theta.$$

(Finding the Bayesian Premium)

1. *Identify $X_i \mid \Theta = \theta$.*
2. *Identify the prior distribution Θ.*
3. *Identify the posterior distribution $\Theta \mid \vec{X}$.*
4. *Calculate*

$$P = \int_{\forall\theta} E[X_{n+1} \mid \Theta = \theta]\pi_{\Theta|\vec{X}}(\theta \mid \vec{x})\,d\theta.$$

Example 4.1.1

The number of claims for a policyholder in year i is X_i for $i = 1, 2$.

Suppose that $X_1 \mid \Theta = \theta$ and $X_2 \mid \Theta = \theta$ are iid with pmf

$$P(X = 1 \mid \Theta = \theta) = 1 - \theta,$$

and

$$P(X = 2 \mid \Theta = \theta) = \theta.$$

The prior distribution is given as $\Theta \sim \text{Beta}(2, 3)$. Determine the Bayesian premium $E[X_2 \mid X_1 = 2]$.

Solution

Method 1: Using predictive distribution Observe that

$$\begin{aligned}
P(X_2 = 2 \mid X_1 = 2) &= \int_{\forall\theta} P(X_2 = 2 \mid \Theta = \theta)\pi_{\Theta|X_1}(\theta \mid x_1)\, d\theta \\
&= \int_{\forall\theta} \theta \cdot \frac{f_{X_1|\Theta}(2 \mid \theta)\pi(\theta)}{\int_{\forall\theta} f_{X_1|\Theta}(x_1 \mid \theta)\pi(\theta)\, d\theta}\, d\theta \\
&= \int_{\forall\theta} \frac{\theta^2 \pi(\theta)}{E[\Theta]}\, d\theta \\
&= \frac{E[\Theta^2]}{E[\Theta]} = \frac{\frac{1}{5}}{\frac{2}{5}} = \frac{1}{2}.
\end{aligned}$$

Thus

$$P(X_2 = 1 \mid X_1 = 2) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Hence

$$E[X_2 \mid X_1 = 2] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}.$$

Method 2: Using Bayesian premium formula We have that

$$\begin{aligned}
E[X_2 \mid X_1 = 2] &= \int_{\forall\theta} E[X_2 \mid \Theta = \theta]\pi_{\Theta|X_1}(\theta \mid 2)\, d\theta \\
&= \int_{\forall\theta} [1(1-\theta) + 2\theta] \cdot \frac{P(X_1 = 2 \mid \Theta = \theta)\pi(\theta)}{\int_{\forall\theta} P(X1 = 2 \mid \Theta = \theta)\pi(\theta)\, d\theta}\, d\theta \\
&= \int_{\forall\theta} \frac{(1+\theta)\theta\pi(\theta)}{E[\Theta]}\, d\theta \\
&= \frac{E[\Theta] + E[\Theta^2]}{E[\Theta]} \\
&= \frac{\frac{2}{5} + \frac{1}{5}}{\frac{2}{5}} = \frac{3}{2}.
\end{aligned}$$

4.2 The Credibility Premium

The Bayesian premium strongly depends on the assumed distribution of $X_i \mid \Theta = \theta$ and Θ. Furthermore, the Bayesian premium may be difficult to evaluate.

Another method to estimate X_{n+1} which we shall study is to make use of linear combinations of past observations, in particular

$$\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i.$$

The estimates $\hat{\alpha}_0, \ldots \hat{\alpha}_n$ are chosen to **minimize** the mean square error

$$\mathcal{Q}(\alpha_0, \ldots, \alpha_n) = E\left[\left(X_{n+1} - \left[\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i\right]\right)^2\right].$$

Let us now develop the general model in calculating the credibility premium.

Theorem 6 (General Model for Credibility Premium)

*Let $\{X_i\}_{i=1}^n$ be a sequence of past observations (rvs), and X_{n+1} the predictive rv. Then, the solution $(\hat{\alpha}_0, \ldots, \hat{\alpha}_n)$ to the system of linear equations, called the **normal equations**,*

$$E[X_{n+1}] = \hat{\alpha}_0 + \sum_{i=1}^{n} \hat{\alpha}_i E[X_i]$$
$$\mathrm{Cov}(X_j, X_{n+1}) = \sum_{i=1}^{n} \hat{\alpha}_i \mathrm{Cov}(X_i, X_j), \quad \forall j \in \{1, \ldots, n\},$$

minimizes the mean square error

$$\mathcal{Q}(\alpha_0, \ldots, \alpha_n) = E\left[\left(X_{n+1} - \left[\alpha_0 + \sum_{i=1}^{n} \alpha_i X_i\right]\right)^2\right].$$

Proof

First, we take partial derivative wrt α_0, and set the derivative to 0, i.e.

$$\frac{\partial Q}{\partial \alpha_0} = E\left[-2\left(X_{n+1} - \hat{\alpha}_0 - \sum_{i=1}^{n} \hat{\alpha}_i X_i\right)\right] = 0.$$

This gives us

$$E[X_{n+1}] = \hat{\alpha}_0 + \sum_{i=1}^{n} \hat{\alpha}_i E[X_i]. \tag{4.1}$$

Now, we take partial derivatives wrt each α_j, $j \in \{1, \ldots, n\}$, and equate the derivatives to 0, i.e.

$$\frac{\partial Q}{\partial \alpha_j} = E\left[-2X_j\left(X_{n+1} - \hat{\alpha}_0 - \sum_{i=1}^{n} \hat{\alpha}_i X_i\right)\right] = 0.$$

Then we have

$$E[X_j X_{n+1}] = \hat{\alpha}_0 E[X_j] + \sum_{i=1}^{n} \hat{\alpha}_i E[X_i X_j]. \tag{4.2}$$

Multiplying Equation (4.1) by $E[X_j]$, for each $j \in \{1, \ldots, n\}$, we get that

$$E[X_{n+1}]E[X_j] = \hat{\alpha}_0 E[X_j] + \sum_{i=1}^{n} \hat{\alpha}_i E[X_i]E[X_j].$$

Subtracting the above from Equation (4.2), we get

$$\operatorname{Cov}(X_i, X_{n+1}) = \hat{\alpha}_0 E[X_j] + \sum_{i=1}^{n} \hat{\alpha}_i \operatorname{Cov}(X_i, X_j),$$

for $j \in \{1, \ldots, n\}$.

It is then clear that $\hat{\alpha}_0, \ldots, \hat{\alpha}_n$ satisfies the normal equations

$$E[X_{n+1}] = \hat{\alpha}_0 + \sum_{i=1}^{n} \hat{\alpha}_i E[X_i]$$

$$\operatorname{Cov}(X_j, X_{n+1}) = \sum_{i=1}^{n} \hat{\alpha}_i \operatorname{Cov}(X_i, X_j), \quad \forall j \in \{1, \ldots, n\}. \qquad \square$$

Note 4.2.1

The equation

$$E[X_{n+1}] = \hat{\alpha}_0 + \sum_{i=1}^{n} \hat{\alpha}_i E[X_i]$$

*is also called the **unbiased equation** because it requires that the estimate*

$\hat{\alpha}_0 + \sum_{i=1}^n \hat{\alpha}_j X_j$ be unbiased for $E[X_{n+1}]$.

Definition 21 (Estimator for the Credibility Premium)

*We define the **estimator for the credibility premium** as*

$$\hat{P} := \hat{\alpha}_0 + \sum_{i=1}^n \hat{\alpha}_i X_i.$$

Corollary 7 ($\hat{P}$ as Best Linear Estimator)

The α_j's, for $j \in \{0, \ldots, n\}$, also minimizes

1.
$$Q_1(\alpha_0, \ldots, \alpha_n) = E\left[\left(E[X_{n+1} \mid \vec{X}] - \left[\alpha_0 + \sum_{i=1}^n \alpha_i X_i\right]\right)^2\right];$$

and

2.
$$Q_2(\alpha_0, \ldots, \alpha_n) = E\left[\left(E[X_{n+1} \mid \Theta] - \left[\alpha_0 + \sum_{i=1}^n \alpha_i X_i\right]\right)^2\right].$$

*We say that $\hat{P}$ is the **Best Linear Estimator** for*

- X_{n+1};
- *the Bayesian premium $E[X_{n+1} \mid \vec{X}]$; and*
- *the hypothetical mean $E[X_{n+1} \mid \Theta] = \mu_{n+1}(\Theta)$.*

Exercise 4.2.1

Prove Corollary 7 by showing that the derivative of the above equations wrt $\alpha_0, \alpha_1, \ldots, \alpha_n$ still satisfy the normal equations.

The name for Theorem 8 is unfortunate, but I can't think of a good name for it, and it is what is used in lectures.

Theorem 8 (Theorem 1)

Suppose $\{X_i\}_{i=1}^n$ is a sequence of past observations, X_{n+1} is the predictive RV, with

- $E[X_i] = \mu$;
- $\mathrm{Var}(X_i) = \sigma^2$; *and*
- $\mathrm{Cov}(X_i, X_j) = \rho\sigma^2$,

for $i \neq j$, $i, j \in \{1, \ldots, n+1\}$, and $\rho \in (-1, 1)$. Then the credibility premium for X_{n+1} is

$$P = Z\overline{X} + (1 - Z)\mu,$$

where

$$Z = \frac{n\rho}{1 - \rho + n\rho},$$

and

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Proof

By Theorem 6, we have that

$$P = \hat{\alpha}_0 + \sum_{i=1}^{n} \hat{\alpha}_i X_i.$$

We shall use the normal equations to attain this, and we know that we can do quite a number of things with the given assumptions. First,

$$\begin{aligned} \mu = E[X_{n+1}] &= \hat{\alpha}_0 + \sum_{i=1}^{n} \hat{\alpha}_i E[X_i] = \hat{\alpha}_0 + \sum_{i=1}^{n} \hat{\alpha}_i \mu \\ &= \hat{\alpha}_0 + \mu \sum_{i=1}^{n} \hat{\alpha}_i. \end{aligned}$$

Making $\sum_{i=1}^{n} \hat{\alpha}_i$ the subject, we get

$$\sum_{i=1}^{n} \hat{\alpha}_i = 1 - \frac{\hat{\alpha}_0}{\mu}. \tag{4.3}$$

Next, for each $j \in \{1, \ldots, n\}$, the equations with covariances become

$$\begin{aligned}\rho\sigma^2 = \mathrm{Cov}(X_j, X_{n+1}) &= \sum_{i=1}^{n} \hat{\alpha}_i \,\mathrm{Cov}(X_i, X_j) \\ &= \sum_{\substack{i=1\\ i\neq j}}^{n} \hat{\alpha}_i \rho\sigma^2 + \hat{\alpha}_j \sigma^2,\end{aligned}$$

and so dividing both sides by σ^2 and then trying to patch that summation, we get

$$\rho = \sum_{i=1}^{n} \hat{\alpha}_i \rho + \hat{\alpha}_j (1-\rho).$$

Substituting in Equation (4.3), we get

$$\rho = \left(1 - \frac{\hat{\alpha}_0}{\mu}\right) \rho + \hat{\alpha}_j (1-\rho),$$

and making $\hat{\alpha}_j$ the subject,

$$\hat{\alpha}_j = \frac{\hat{\alpha}_0 \rho}{\mu(1-\rho)}.$$

We want to have a more explicit formula for $\hat{\alpha}_0$ and $\hat{\alpha}_j$. Looking at Equation (4.3), we first take the sum of the $\hat{\alpha}_i$'s (save when $i = 0$):

$$\sum_{i=1}^{n} \hat{\alpha}_i = \frac{n\hat{\alpha}_0 \rho}{\mu(1-\rho)}.$$

So

$$1 - \frac{\hat{\alpha}_0}{\mu} = \frac{n\hat{\alpha}_0 \rho}{\mu(1-\rho)},$$

and after rearrangement, we get

$$\hat{\alpha}_0 = \frac{(1-\rho)\mu}{n\rho + 1 - \rho}.$$

Going for $\hat{\alpha}_j$, we get

$$\hat{\alpha}_j = \frac{\hat{\alpha}_0 \rho}{\mu(1-\rho)} = \frac{\rho}{n\rho + 1 - \rho}.$$

Thus

$$P = \frac{(1-\rho)\mu}{n\rho + 1 - \rho} + \sum_{i=1}^{n} \frac{\rho X_i}{n\rho + 1 - \rho}$$

$$= \frac{n\rho}{n\rho + 1 - \rho} \cdot \frac{1}{n} \sum_{i=1}^{n} X_i + \frac{1-\rho}{n\rho + 1 - \rho} \mu,$$

where we note that

$$1 - \frac{n\rho}{n\rho + 1 - \rho} = \frac{1-\rho}{n\rho + 1 - \rho}.$$

Thus if we let

$$Z = \frac{n\rho}{n\rho + 1 - \rho} \text{ and } \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

we have that

$$P = Z\overline{X} + (1 - Z)\mu,$$

as desired. □

4.3 The Bühlmann Model

An example of Theorem 8 is the Bühlmann model, which is one of the (if not the) simplest credibility model.

Definition 22 (The Bühlmann Model)

Under the ***Bühlmann model****, conditional on* Θ *(the risk distribution), for each policyholder, past losses* $X_1, \ldots, X_n$ *have the same mean and variance, and are iid conditional on* Θ*. In particular, in this model, we define the* ***hypothetical mean*** *as*

$$\mu(\theta) := E[X_i \mid \Theta = \theta],$$

and the ***process variance*** *as*

$$v(\theta) = \mathrm{Var}(X_j \mid \Theta = \theta).$$

Furthermore, we also define the ***structural parameters****: the* ***expected hypothetical mean***

$$\mu = E[\mu(\Theta)],$$

*the **mean of the process variance***

$$v = E[v(\Theta)],$$

*and the **variance of the hypothetical mean***

$$a = \mathrm{Var}(\mu(\Theta)).$$

Note 4.3.1

*μ is the estimate to use if we have no information about θ (thus no info about $\mu(\theta)$). In this case, we call μ the **collective premium**.*

It is not difficult to obtain the mean, variance, and covariance of X_j's for each j. We see that the mean of X_j is

$$E[X_j] = E[E[X_j \mid \Theta]] = E[\mu(\Theta)] = \mu.$$

The variance of X_j is

$$\begin{aligned}
\mathrm{Var}(X_j) &= \mathrm{Var}(E[X_j \mid \Theta]) + E[\mathrm{Var}(X_j \mid \Theta)] \\
&= \mathrm{Var}(\mu(\Theta)) + E[v(\Theta)] \\
&= a + v.
\end{aligned}$$

The covariance of X_j with X_i is

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= E[X_i X_j] - E[X_i]E[X_j] \\
&= E[E[X_i X_j \mid \Theta]] - \mu^2 \\
&= E[E[X_i \mid \Theta]E[X_j \mid \Theta]] - \mu^2 \\
&= E[\mu(\Theta)^2] - [\mu(\Theta)]^2 \\
&= \mathrm{Var}(\mu(\Theta)) = a.
\end{aligned}$$

This is exactly what the Bühlmann model assumes. In fact, if we apply Theorem 8, noting that

$$\mathrm{Var}(X_i) = \sigma^2 \text{ and } \mathrm{Cov}(X_i, X_j) = \rho\sigma^2 \implies \rho = \frac{\mathrm{Cov}(X_i, X_j)}{\mathrm{Var}(X_i)},$$

we observe that

$$\mu = \mu,\ \sigma^2 = v + a,\ \rho = \frac{a}{v+a},$$

and so

$$Z = \frac{n\frac{a}{v+a}}{n\frac{a}{v+a} + 1 - \frac{a}{v+a}} = \frac{na}{na+v} = \frac{n}{n+\frac{v}{a}}.$$

The following result follows exactly from our discussion above.

(Finding Bühlmann Credibility Premium)

1. *Find hypothetical mean $\mu(\theta)$ and process variance $v(\theta)$.*
2. *Find structural parameters μ, v, a.*
3. *Calculate the Bühlmann credibility factor Z (and mean loss $\overline{X}$ if necessary).*
4. *Calculate the Bühlmann credibility premium $P = Z\overline{X} + (1-Z)\mu$.*

Theorem 9 (Bühlmann Credibility Premium)

The Bühlmann credibility premium is

$$P = Z\overline{X} + (1-Z)\mu$$

where

$$Z = \frac{n}{n+\frac{v}{a}}$$

is called the ***Bühlmann credibility factor****.*

Note 4.3.2

- *The Bühlmann credibility premium is a weighted average of the sample mean $\overline{X}$ and the collective premium μ.*
- *As n increases, $Z \to 1$, giving more credit to $\overline{X}$, which is reasonable by intuition since our past data is more robust with more exposure.*
- *If the population is fairly homogeneous wrt the risk parameter Θ, then (relatively speaking) the hypothetical means $\mu(\Theta)$ to not vary greatly with Θ, which then gives small variability. In other words, a is small relative to v, and thus Z is nudged closer to 0. This agrees with our intuition, since for a homogeneous population, the overall mean μ is more of value in helping the prediction of next year's claims for a particular policyholder.*
- *If the population is heterogeneous, $\mu(\Theta)$ is more variable, so a is large, and in turn Z is closer to 1. This agrees with intuition, since experience of other policyholders is of less value in predicting future experience of a particular policyholder as compared to past experience.*

Example 4.3.1 (A Poisson-Gamma Example for Bühlmann Credibility)

Let $\{X_i \mid \Theta = \theta\}_{i=1}^n$ with $X_i \mid \Theta = \theta \sim \text{Poi}(\theta)$ for $i \in \{1, \ldots, n\}$, and the prior distribution $\Theta \sim \text{Gam}(\alpha, \beta)$. Find both the Bühlmann credibility premium and the Bayesian premium.

Solution

Bühlmann Credibility Premium We observe that

$$\mu(\theta) = E[X_i \mid \Theta = \theta] = \theta,$$

and

$$v(\theta) = \text{Var}(X_i \mid \Theta = \theta) = \theta.$$

The structural parameters are

$$\mu = E[\mu(\Theta)] = E[\Theta] = \alpha\beta, \quad v = E(v(\Theta)) = E(\Theta) = \alpha\beta,$$

and

$$a = \text{Var}(\mu(\Theta)) = \text{Var}(\Theta) = \alpha\beta^2.$$

Thus the Bühlmann credibility factor is

$$Z = \frac{n}{n + \frac{v}{a}} = \frac{n}{n + \beta^{-1}}.$$

Hence the Bühlmann credibility premium is

$$\begin{aligned} P &= Z\overline{X} + (1 - Z)\mu \\ &= \frac{n}{n + \beta^{-1}}\overline{X} + \frac{\beta^{-1}}{n + \beta^{-1}}\alpha\beta \\ &= \frac{\alpha + n\overline{X}}{n + \beta^{-1}}. \end{aligned}$$

Bayesian Premium We are given that $X_i \mid \Theta = \theta \sim \text{Poi}(\theta)$ and $\Theta \sim \text{Gam}(\alpha, \beta)$. The posterior distribution $\Theta \mid \vec{X}$ is

$$\pi_{\theta \mid \vec{X}}(\theta \mid \vec{x}) \propto \left(\prod_{i=1}^n f_{X_i \mid \Theta}(x_i \mid \theta) \right) \pi(\theta)$$

$$
\begin{aligned}
&\propto \theta^{\alpha-1} e^{-\frac{\theta}{\beta}} \prod_{i=1}^{n} e^{-\theta} \theta^{x_i} \\
&= e^{-(n+\frac{1}{\theta})\theta} \theta^{n\overline{x}+\alpha-1}.
\end{aligned}
$$

It follows that $\Theta \mid \vec{X} = \vec{x} \sim \text{Gam}\left(n\overline{x}+\alpha, \frac{1}{n+\frac{1}{\beta}}\right)$. Thus the Bayesian premium is

$$
\begin{aligned}
&E[X_{n+1} \mid \vec{X} = \vec{x}] \\
&= \int_{\forall\theta} E[X_{n+1} \mid \Theta = \theta] \pi_{\Theta|\vec{X}}(\theta \mid \vec{x})\, d\theta \\
&= \int_{\forall\theta} \cancel{\theta} \frac{1}{\cancel{\theta}\Gamma(n\overline{x}+\alpha)} \left(\frac{\theta}{\frac{1}{n+\frac{1}{\beta}}}\right)^{n\overline{x}+\alpha} e^{-\frac{\theta}{\frac{1}{n+\frac{1}{\beta}}}}\, d\theta \\
&= (n\overline{x}+\alpha)\frac{1}{n+\frac{1}{\beta}} \int_{\forall\theta} \frac{1}{\theta\Gamma(n\overline{x}+\alpha+1)} \left(\frac{\theta}{\frac{1}{n+\frac{1}{\beta}}}\right)^{n\overline{x}+\alpha+1} e^{-\frac{\theta}{\frac{1}{n+\frac{1}{\beta}}}}\, d\theta \\
&= (n\overline{x}+\alpha)\frac{1}{n+\frac{1}{\beta}} \\
&= \frac{n}{n+\beta^{-1}}\overline{x} + \frac{\beta^{-1}}{n+\beta^{-1}}\alpha\beta \\
&= \frac{\alpha+n\overline{x}}{n+\beta^{-1}}.
\end{aligned}
$$

Note 4.3.3

We notice that the Bühlmann credibility premium and the Bayesian premium coincides. This is no accidental coincidence, and we shall see why this is the case later on in exact credibility.

Example 4.3.2 (Disagreement of Bühlmann Credibility Premium and Bayesian Premium)

Consider 2 urns with different proportions of balls marked with 0 or 1.

- Urn 1 has 60% of its balls marked as 0 and 40% marked as 1.
- Urn 2 has 80% of its balls marked as 0 and 20% marked as 1.

An urn is randomly picked with equal probability and a total of 2 balls out of 3 is marked 1 (with replacement).

Calculate the Bühlmann credibility premium and the Bayesian premium for the number on the next ball drawn from the urn.

Solution

In any of the cases, we need to find out what Θ and $X_i \mid \Theta$ are. Let X_i be the number drawn on the ith ball, and Θ the number of the chosen urn. Then the prior distribution is

$$\Theta = \begin{cases} \theta_1 & \text{urn 1 is selected wp } \frac{1}{2} \\ \theta 2 & \text{urn 2 is selected wp } \frac{1}{2} \end{cases}.$$

The conditional probabilities are

$$P(X_i = x \mid \Theta = \theta_1) = \begin{cases} 0.6 & x = 0 \\ 0.4 & x = 1 \end{cases}$$

and

$$P(X_i = x \mid \Theta = \theta_2) = \begin{cases} 0.8 & x = 0 \\ 0.2 & x = 1 \end{cases}.$$

Bühlmann credibility premium The hypothetical means are

$$\mu(\theta_1) = E[X_i \mid \Theta = \theta_1] = 0(0.6) + 1(0.4) = 0.4$$

and

$$\mu(\theta_2) = E[X_i \mid \Theta = \theta_2] = 0(0.8) + 1(0.2) = 0.2.$$

The process variances are

$$v(\theta_1) = \mathrm{Var}(X_i \mid \Theta = \theta_1) = 0.4 - 0.4^2 = 0.24$$

and

$$v(\theta_2) = \mathrm{Var}(X_i \mid \Theta = \theta_2) = 0.2 - 0.2^2 = 0.16.$$

It follows that the structural parameters are

$$\mu = E[\mu(\Theta)] = \frac{1}{2}(0.4) + \frac{1}{2}(0.2) = 0.3,$$

$$v = E[v(\Theta)] = \frac{1}{2}(0.24) + \frac{1}{2}(0.16) = 0.2,$$

and

$$a = \mathrm{Var}(\mu(\Theta)) = (0.4 - 0.3)^2\frac{1}{2} + (0.2 - 0.3)^2\frac{1}{2} = 0.01$$

Thus the Bühlmann credibility factor is

$$Z = \frac{n}{n + \frac{v}{a}} = \frac{n}{n + \frac{0.2}{0.01}} = \frac{n}{n + 20}.$$

Hence the Bühlmann credibility premium is

$$P = \frac{n}{n+20}\frac{2}{3} + \frac{20}{n+20}0.3 = 0.34783.$$

Bayesian premium Let $\vec{X} = X_1 + X_2 + X_3$. Our observation is that $X_1 + X_2 + X_3 = 2$. Thus

$$\begin{aligned}
&\pi_{\Theta|\vec{X}}(\theta_1 \mid 2) \\
&= \frac{P(X_1 + X_2 + X_3 = 2 \mid \Theta = \theta_1)\pi(\theta_1)}{P(X_1 + X_2 + X_3 = 2 \mid \Theta = \theta_1)\pi(\theta_1) + P(X_1 + X_2 + X_3 \mid \Theta = \theta_2)\pi(\theta_2)} \\
&= \frac{\binom{3}{2}(0.4)^2(0.6)\frac{1}{2}}{(\binom{3}{2}(0.4)^2(0.6)\frac{1}{2} + \binom{3}{2}(0.2)^2(0.8)\frac{1}{2})} \\
&= 0.75,
\end{aligned}$$

and so

$$\pi_{\Theta|\vec{X}}(\theta_2 \mid 2) = 0.25.$$

Hence, to the Bayesian premium is

$$\begin{aligned}
E[X_4 \mid X_1 + X_2 + X_3 = 2] &= E[X_4 \mid \Theta = \theta_1]0.75 + E[X_4 \mid \Theta = \theta_2]0.25 \\
&= 0.4(0.75) + 0.2(0.25) \\
&= 0.3 + 0.05 = 0.35.
\end{aligned}$$

Bibliography

Bühmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, 4:199–207.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: From Data to Decisions*. John Wiley & Sons Inc., Hoboken, New Jersey, 4th edition.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society*, I:24–30.

Index