CREDIBILITY-THEORY

Seminar in Financial and Actuarial Mathematics

TU WIEN
Ana-Iulia DOMNISAN
Matrikelnr.: 0627124
Mai 2009

Inhaltsverzeichnis

1 Introduction

Dating back to the I World War, Credibility Theory came to life as a response to the problem of calculating "insurance-premium" based on the information obtained from two main sources: the individual and the collective. Nowadays, Credibility Theory counts to the main topics in actuarial science being defined as a mathematically dense body of formulas, providing tools to deal with randomness of data and having as a main goal the prediction of future events or costs.

1.1 The First Formula

In 1918, A.Whitney adressed the problem of assesing the risk premium m, defined as the expected claims per unit of risk exposed, for an individual risk from a class of similar risks. Using both the individual risk experience and the class risk experience, he proposed that the premium rate be a weighted average described with the help of the following formula:

$$\bar{m} \;=\; Z\hat{m} \;+\; (1 - Z)\mu \;, \quad 0 \leq \; Z \; \leq \; 1$$

where:

- $\bar{m}$ is the observed mean claim amount per unit of risk exposed for the individual contract,
- μ is the corresponding overall mean of the insurance portfolio,
- Z is called the credibility assigned to the information on the individual contract,
- $(1 - Z)$ is refered to as the complement of credibility.

The risk premium here is viewed as a random variable. In modern Credibility Theory, it is a function $m(\Theta)$ of a random element Θ containing the unobservable characteristics of the individual risk. The random nature of Θ expresses the notion of heterogeneity; the individual risk is a random selection from a portfolio of similar risks and the distribution of Θ describes the variation of individual risk characteristics across the portfolio.

1.2 Practical Example

A company has insured twenty drivers over the past ten years. In the following table, the number 1 indicates the fact that the driver i had at least one accident during the year j, the last row shows the number of years with at least one accident:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1			1							1							1	1		
2							1		1	1							1			
3			1				1		1				1							
4																	1			
5									1	1										
6						1			1		1									
7									1					1			1			
8											1									
9						1				1		1								
10									1		1						1		1	
Σ	0	0	2	0	0	2	2	0	6	4	3	1	1	1	0	0	5	1	1	0

From the table one can notice that some of the drivers had no accidents, these drivers are considered to be "good risks". In simple words, the idea that stays at the bottom of credibility theory is that each insured should pay a premium in the next year according to his "behavior" in the previous period of time. According to the individual loss history, one might deduct that some drivers ought to get a zero-premium, whereas the collective loss history insinuates that there should be no awards for good claims experience. The main questions here are:

- Should the "good risks" pay as much as the "bad risks"?
- Is there any chance that some of the drivers might have had bad luck?
- What is the best estimate of the premium rate for a certain driver i?

Supposing the average loss produced by a driver i over all years is $\hat{m}$ and the average over all years and all drivers is μ, then with Whitney's formula one can calculate the future premium as a weighted average of both the individual loss and the collective loss:

$$\bar{m} \;=\; Z\hat{m} \;+\; (1-Z)\mu \;, \quad 0 \leq \; 1$$

where $\bar{m}$ describes the risk premium.

In practical insurance, the competition forces the companies to offer the fairest rates possible. Therefore if company A offers equal rates to all risks in a heterogenous collective, the "good-risks" will pay to much and the "bad-risks" too low. If on the other hand the competing company B offers differentiated premiums according to the loss history of every insured, then this company will be more attractive to "good risks" and less attractive to "bad risks". As a consequence the company A will lose the "good risks" and register an increase in "bad risks". Insurers refer to such an effect as "anti-selection".

1.3 Approaches

In order to finish the example given above, one must calculate Z: how much credibility should be assigned to the information known about the driver i? To solve the matter, following points of view have been expressed, according to how risk classes are evaluated:

1. The **Classical Credibility Model** is also refered to as the limited fluctuations credibility because it attempts to limit the effect that random observations will have on the estimates. As a conclusion, the classical point of view assumes that risks can be classified into homogeneous groups. However in insurance there are no homogeneous risks and the next two models take this into consideration:

2. **Bühlmann Credibility**, also known as the least squares credibility, sets as a goal the minimization of the square of the error between the estimate and the true expected value of the quantity being estimated.

3. **The Bayesian Analysis** is the approach that combines current observation with prior information in order to produce a better estimate. Bayes Theorem is the foundation for this analysis. Research has shown that Bühlmann credibility estimates are the best linear square fits to the Bayesian estimates. For this reason Bühlmann credibility is also referd to as Bayesian credibility. The Bayesian estimate is a linear weightening of current and prior information with weights Z and $(1 - Z)$ where Z is the **Bühlmann credibility**.

2 Mathematical Background

2.1 The Model

The main idea underlying insurance is that individuals exposed to the same risk join together to form a "community-at-risk" and on payment of a premium they transfer their risk to an insurance company.
Let us consider an insurance company with a portfolio of I insured risks numbered $i = 1, 2, \ldots, I$. In a defined insurance period, the risk i produces:

- a number of claims N_i,
- with claim sizes Y_i^ν, where $\nu = 1, 2, \ldots, N_i$,
- together giving the aggregate claim amount $X_i = \sum_{\nu=1}^{N_i} Y_i^\nu$.

In addition the **gross premium** is defined as the premium payable by the insured to the insurer for the bearing of the risk.
The **premium volume** is the sum of all gross premiums over the whole portfolio in the insurance period.
In order to rate the risk one must determine the so called **pure risk premium** (or simply : the risk premium):

$$P_i = \mathbf{E}[X_i].$$

2.2 The Individual Risk

The **individual risk** can be regarded as a black box that produces aggregate claim amounts X_j, where X_j is the claim amount in the well specified time period j. Mathematically, X_j is interpreted as a random variable that holds past information. Our main task is to determine the risk premium for the aggregate claims in a future period: X_{n+1}, on the basis of the observations made in the previous n periods: $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$. Certain assumptions must be made regarding the distribution function of the random variables X_j:

A1: Stationarity – the X_j-s are all identically distributed with (conditional) distribution function $F(x)$. The stationarity assumption allows one to establish a relationship between the past and the future. A (weaker) assumption of stationarity is always needed for the calculation of insurance premiums on the basis of historical data. In practice, the stationarity is oftenly achieved by making adjustments for inflation, indexing, trend elimination etc.

A2: (Conditional) Independence – the random variables X_j, $j = 1, 2, \ldots,$ are (conditionally) independent (given the distribution $F(x)$).

Usually, the distribution function F is either unknown to the insurer or varies from one risk profile to another. Therefore a parametrisation of F can be made: F_ϑ, where ϑ can be seen as an element of some abstract space Θ, also unknown or varying from risk to risk. In practice, ϑ represents the "risk profile".

2.3 The Two Urn Model

Even if the values of the different risk profiles ϑ are mostly unknown to the insurer, based on the a priori knowledge and the statistical information, he does know something about the structure of the collective. Making a parallel to the above presented example, the insurer knows, for instance, that most car drivers are "good risks", seldomly collecting a claim, while there is a small percentage of drivers that make frequent claims. In statistical parlance, all this information can be summarized by a probability distribution $U(\vartheta)$ over the space Θ.

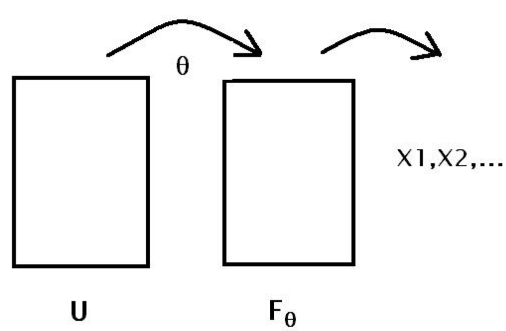
Definition: The probability distribution $U(\vartheta)$ is called the STRUCTURAL FUNCTION of the collective.

$U(\vartheta)$ can be interpreted in two ways:

- in the **empirical Bayesian** interpretation the ϑ-s in the collective are considered to be a random sample from some fixed set Θ; consequently the function $U(\vartheta)$ will describe the idealized frequencies of the ϑ-s over Θ (this is the predominant point of view used here).

- in the **pure Bayesian** interpretation the distribution function $U(\vartheta)$ stands for the personal beliefs, a priori knowledge and experience of the actuary.

Mathematically, the collective rating problem has been described with help of the **Two-Urn Model** :



The first urn contains the collective with distribution function U. From this urn, the individual risk, or equivalently its risk pofile ϑ (Note: In the drawing $\theta = \vartheta$), is selected. This ϑ will now determine the content of the second urn, or equivalently the distribution function F_ϑ. From the second urn, the random values $X_1,\ X_2, \ldots$ are selected, they are independent and indentically distributed with distribution function F_ϑ.
As a result, every risk is now characterized by an individual risk profile ϑ, which is itself the realisation of a random variable Θ and the following remarks can be made:

- $X_1, X_2, \ldots$ are conditionally independent, given $\Theta = \vartheta$, and identically distributed with ditribution function F_ϑ.
  Unconditionally, they are positively correlated as the following shows:

$$\begin{aligned} Cov(X_1, X_2) &= E[Cov(X_1, X_2)|\Theta] + Cov(E[X_1|\Theta], E[X_2|\Theta]) \\ &= Cov(\mu(\Theta), \mu(\Theta)) \\ &= Var[\mu(\Theta)] \end{aligned}$$

  where $\mu(\Theta)$ is a random variable that describes the individual premium.

- Θ is itself a random variable with distribution function U.

2.4 Premium Types

Definition: The INDIVIDUAL PREMIUM is defined by:

$$P^{ind} \ = \ \mu(\Theta) \ = \ E[\ X_{n+1}|\Theta\ ]\ .$$

As one can notice from the formula, the individual premium is a conditional expectation and therefore a random variable. P^{ind} is considered to be a "fictional" premium since it is not a real number, that assigns a fixed cost for assuming a given risk. The size of the "true" premium could only then be deducted if the value ϑ of the random variable Θ would be known.

Definition: The COLLECTIVE PREMIUM is defined by an unconditional expectation (a fixed number) and given with help of the following formula:

$$P^{coll} \ = \ \mu_0 \ = \ \int_\Theta \mu(\vartheta)\ dU(\vartheta) \ = \ E[\ X_{n+1}\ ]\ .$$

However, both P^{ind} and P^{coll} do not satisfy our main goal: to estimate as precisely as possible, for each risk, its premium $\mu(\Theta)$. On the one hand P^{ind} takes into consideration the individual risk, but on the other it neglects the pre-existing claim experience. While P^{coll} does exactly the opposite: considers the risk as belonging to the collective, but does not take into account the individual claim experience. Therefore we define:

Definition: The BAYES PREMIUM as:

$$P^{Bayes} \ = \ \widetilde{\mu(\Theta)} \ = \ E[\ \mu(\Theta)|X\ ]\ ,$$

where X denotes the vector of aggregate claim amounts made over the specified period n.
The value of the Bayes premium (or the BEST EXPERIENCE PREMIUM) depends on the individual claim experience, known to the insurer at the time at which the future risk is to be rated. That is why this premium is regarded as a "true" premium.

3 The Bayes Premium

3.1 Basic Elements

In order to study the Bayes premium in depth, some basic concepts and ideas from statistical decision theory and bayes statistics must be presented:

As before, in this chapter $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ represents the observation vector with distribution function: $F_\vartheta = P_\vartheta[\mathbf{X} \le \mathbf{x}]$, completely or partly unknown (equivalently: the parameter ϑ is completely or partly unknown). Our task is to find a functional $g(\vartheta)$ of the parameter ϑ. This can be achieved with the help of another function $T(\mathbf{X})$, which depends solely on the observation vector $\mathbf{X}$ and is called an ESTIMATOR for $g(\vartheta)$.

Formulation of the Problem:

$\vartheta \in \Theta$: The set of parameters, which contains the true values of ϑ,
$T \in D$: The set of functions to which the estimator function must belong. T can also be seen as a map from the observation space $\mathbf{R}^n$ into the set of all possible values of the functional g: $\{\ g(\vartheta) :\ \vartheta \in \Theta\}$.

Definition: The LOSS FUNCTION is defined by:

$L(\vartheta, T(\mathbf{x}))$: loss, if ϑ is the "true" parameter and $T(\mathbf{x})$ is the value taken by the estimator when the value $\mathbf{x}$ is observed.

Definition: The RISK FUNCTION of the estimator T is:

$$R_T(\vartheta) \ := \ E_\vartheta[L(\vartheta, T)] \ = \ \int_{\mathbf{R}^n} L(\vartheta, T(\mathbf{x}))dF(\mathbf{x}) \ ,$$

where both functions T and L are chosen so that the integral exists.

Remark: In general, it is not possible to find an estimator T for which the risk $R_T(\vartheta)$ will be minimal for all values of ϑ.

Bayes Statistics and the Bayes Premium

Making a parallel from the above presented concepts into Bayes statistics, following corespondences and notations are to be considered:

- $U(\vartheta)$ is called the *a priori* distribution of Θ (before observations have been made),
- $U_{\mathbf{x}}(\vartheta)$ is the *a posteriori* distribution of Θ (after observations have been made).

Definition: The BAYES RISK of the estimator T with respect to the a priori distribution $U(\vartheta)$ is:

$$R(T) \ := \ \int_\Theta R_T(\vartheta)dU(\vartheta).$$

This definition states that there is always a possibility to rank estimators by increasing risks, or differently: there is a complete ordering on the set of estimators.

Definition: The BAYES ESTIMATOR $\widetilde{T}$ is defined by:

$$\widetilde{T} \;:=\; \underbrace{\arg\min}_{T\in\; D_1}\; R(T),$$

where D_1 is the set of all mathematically allowable estimators, which have integrable risk functions. In words: the estimator $\widetilde{T}$ is exactly the estimator that minimizes the Bayes risk $R(\cdot)$.

With help of the following sequence of equations:

$$\begin{aligned} R(T) &= \int_\Theta R_T(\vartheta)dU(\vartheta) \;=\; \int_\Theta E_\vartheta[L(\vartheta,T)]dU(\vartheta) \\ &= \int_\Theta \int_{\mathbf{R}^n} L(\vartheta, T(\mathbf{x}))dF_\vartheta(\mathbf{x})dU(\vartheta) \\ &= \int_{\mathbf{R}^n} \int_\Theta L(\vartheta, T(\mathbf{x}))dU_{\mathbf{x}}(\vartheta)dF_\vartheta(\mathbf{x}) \end{aligned}$$

the next rule for constructing the Bayes estimator can be deduced:

Theorem: For every possible observation $\mathbf{x}$, $\widetilde{T(\mathbf{x})}$ takes the value which minimizes $\int_\Theta L(\vartheta, T(\mathbf{x})dU_{\mathbf{x}}(\vartheta)$. In other words, for every possible observation $\mathbf{x}$, $\widetilde{T(\mathbf{x})}$ is the BAYES ESTIMATOR with respect to the distribution function $U_{\mathbf{x}}(\vartheta)$.

Going back to the Bayesian modelling:

- $\mu(\vartheta)$ represents the correct individual premium, and is the correspondent of the functional $g(\vartheta)$,
- Θ will now be the set of possible individual risk profiles,
- a particular form of the loss function will be used:

$$L(\vartheta, T(\mathbf{x})) \;=\; (\mu(\vartheta) \;-\; T(\mathbf{x}))^2$$

Definition: An estimator $\widehat{\mu(\Theta)}$ is at least as good as another estimator $\widehat{\mu(\Theta)}^*$ if

$$E\left[\left(\widehat{\mu(\Theta)} \;-\; \mu(\Theta)\right)^2\right] \;\leq\; E\left[\left(\widehat{\mu(\Theta)}^* \;-\; \mu(\Theta)\right)^2\right]$$

$E\left[\left(\widehat{\mu(\Theta)} \;-\; \mu(\Theta)\right)^2\right]$ is called the QUADRATIC LOSS of the estimator $\widehat{\mu(\Theta)}$.

The next theorem will prove that $P^{Bayes} = \widetilde{\mu(\Theta)}$ is the **best** possible experience premium:

Theorem: The BAYES ESTIMATOR with respect to the QUADRATIC LOSS FUNCTION is given by the a posteriori expectation of the individual premium $\mu(\Theta)$:

$$\widetilde{\mu(\Theta)} \;=\; E\Big[\mu(\Theta)|\mathbf{X}\Big].$$

Proof: Let $\widehat{\mu(\Theta)}$ be an estimator of $\mu(\Theta)$ and $\widetilde{\mu(\Theta)}$ be its a posteriori expectation $E\Big[\mu(\Theta)|\mathbf{X}\Big]$. Then:

$$\begin{aligned}
E\left[\left(\widehat{\mu(\Theta)} \;-\; \mu(\Theta)\right)^2\right] \;&=\; E\left[E\left[\left(\widehat{\mu(\Theta)} - \widetilde{\mu(\Theta)} + \widetilde{\mu(\Theta)} - \mu(\Theta)\right)^2 \middle|\; \mathbf{X}\right]\right] \\
&=\; E\left[\left(\widehat{\mu(\Theta)} \;-\; \widetilde{\mu(\Theta)}\right)^2\right] \;+\; E\left[\left(\widetilde{\mu(\Theta)} \;-\; \mu(\Theta)\right)^2\right].
\end{aligned}$$

And the theorem is proven since out of all estimators for $\mu(\Theta)$ the smallest quadratic loss is achieved for the estimator $\widetilde{\mu(\Theta)} \;=\; E\Big[\mu(\Theta)|\mathbf{X}\Big]$.
Furthermore, the quadratic loss function leads to the collective premium (respectively the individual premium) in case there is no loss experience available (respectively if Θ is known).

Theorem:

- The QUADRATIC LOSS OF THE BAYES PREMIUM is:

$$E\left[\left(\widetilde{\mu(\Theta)} \;-\; \mu(\Theta)\right)^2\right] \;=\; E\left[Var\Big[\mu(\Theta)|\mathbf{X}\Big]\right].$$

- The QUADRATIC LOSS OF THE COLLECTIVE PREMIUM is:

$$\begin{aligned}
E\left[\left(\mu_0 \;-\; \mu(\Theta)\right)^2\right] \;&=\; Var\Big[\mu(\Theta)\Big] \\
&=\; \underbrace{E\left[Var\Big[\mu(\Theta)|\mathbf{X}\Big]\right]}_{first\ variance\ component} \;+\; \underbrace{Var\left[E\Big[\mu(\Theta)|\mathbf{X}\Big]\right]}_{second\ variance\ component}\;.
\end{aligned}$$

3.2 The Poisson-Gamma Case

In the 1960's the Bonus-Malus system came to life due to the following circumstances: insurers requested the increase of the premium rates, claiming that the current level was insufficient to cover their present risks. The supervision authority agreed to the rise with one condition: the rates ought to be calculated according to the individual claim experience. The main goal of this approach was to distinguish between "good risks", people who had never made a claim, and "bad risks", clients who had made numerous claims.

F.Bichsel, the first non-life actuary in Switzerland, was given the task of contructing a risk rating system which would adjust the premium to the individual risk profile. He came to the conclusion that the individual risk profiles could be classified according to the number of claims made by an individual, where the claim size was of less importance, mainly because of the very high variability in the amounts and no way of actually being able to predict that value.

Mathematical Model

Necessary variables:

- N_j is the number of claims made by a particular driver in year j,
- X_j is its corresponding aggregate claim amount.

For his model, F.Bichsel made the implicit assumption that given an individual risk profile ϑ, the following holds for the aggregate claim amount X_j:

$$E[X_j|\Theta = \vartheta] \; = \; C \; E[N_j|\Theta = \vartheta] \quad (j = 1, 2, \ldots),$$

where

- C is a constant depending on the horsepower of the car,
- $E[N_j|\Theta = \vartheta]$ depends only on the driver of the car.

Furthermore, Bichsel based his model on two other assumptions (which also give the name of the model):

Model Assumptions (Poisson−Gamma)

- *PG1:* Conditionally, given $\Theta = \vartheta$, the N_j-s ($j = 1,\ 2, \ldots, n$) are independent and Poisson distributed with Poisson parameter ϑ:

$$P(N_j = k|\Theta = \vartheta) \; = \; e^{-\vartheta} \, \tfrac{\vartheta^k}{k!}.$$

- *PG2:* Θ has a Gamma distribution with shape parameter γ and scale parameter β, the structural function has density:

$$u(\vartheta) \;=\; \frac{\beta^\gamma}{\Gamma(\gamma)}\, \vartheta^{\gamma-1} e^{-\beta\vartheta}, \quad \vartheta \geq 0.$$

With help of the above made assumptions, the claim frequencies can be calculated:

THE INDIVIDUAL CLAIM FREQUENCY: $\quad \mathrm{F}^{ind} \;=\; \mathrm{E}[\mathrm{N}_{n+1}|\Theta] \;=\; \Theta.$

THE COLLECTIVE CLAIM FREQUENCY: $\quad \mathrm{F}^{coll} \;=\; \mathrm{E}\,[\Theta] \;=\; \frac{\gamma}{\beta}.$

THE BAYES CLAIM FREQUENCY: $\quad \mathrm{F}^{Bayes} \;=\; \frac{\gamma\,+\,N_\bullet}{\beta\,+\,n} \;=\; \alpha\overline{N} \;+\; (1-\alpha)\,\frac{\gamma}{\beta}$

where $\alpha \;=\; \frac{n}{n+\beta}, \quad \overline{N} \;=\; \frac{1}{n}\, \sum_{j=1}^{n} N_j$

and $\mathrm{N}_\bullet \;=\; \sum_{j=1}^{n} N_j.$

In addition the QUADRATIC LOSS of F^{Bayes} is:

$$\begin{aligned}\mathrm{E}\left[(\mathrm{F}^{Bayes}-\Theta)^2\right] \;&=\; (1 - \alpha)\; E\left[(F^{coll}-\Theta)^2\right] \\ &=\; \alpha\; E\left[(\overline{N}-\Theta)^2\right].\end{aligned}$$

Remarks to the model:

- The quantities P^{ind}, P^{coll} and P^{Bayes} can be obtained by multiplying F^{ind}, F^{coll} and F^{Bayes} with C.
- The Bayes premium CF^{Bayes} is a linear function of the claim numbers. This is an example of "credibility theory".
- F^{Bayes} is an average of:
  - · $\overline{N}$ = observed individual claim frequency and
  - · $E\,[\Theta] \;=\; \frac{\gamma}{\beta}$ = a priori expected claim frequency.
- $\alpha \;=\; \frac{n}{n+\beta}$ is called the CREDIBILITY WEIGHT
- $\beta \;=\; \frac{E[\Theta]}{Var[\Theta]}$

  The credibility weight will grow with the increase of the number of observation years n. While the larger β is, the smaller will the weight get. In practice this would be explained as follows: the more information the insurer has on an individual, the greater the weight attached to the individual claim experience will be, on the other hand the more homogenous the whole collective is, the more adequate it will be to use the collective claim experience for the rating of the individual risk.

- The formula of the quadratic loss of F^{Bayes} could be equivalently expressed as:

$$\begin{aligned} \text{quadratic loss of } F^{Bayes} &= (1-\alpha) \cdot \text{quadratic loss of } F^{coll}, \\ \text{quadratic loss of } F^{Bayes} &= \alpha \cdot \text{quadratic loss of } \overline{N}. \end{aligned}$$

  In words: the credibility weight α is the factor by which the quadratic loss is reduced if the Bayes premium will be used instead of the collective premium or, according to the second equation, the factor by which the quadratic loss has to be multiplied if F^{Bayes} is preferred to the observed individual claim frequency $\overline{N}$.

- Estimators:

  - $\mathbf{F^{coll}}$ is the estimator based only on the a priori knowledge from the collective and neglecting the individual claims experience.
  - $\overline{\mathbf{N}}$ is the estimator based only on the individual claims experience, neglecting the a priori knowledge.
  - $\mathbf{F^{Bayes}}$ is a combination of both sources of information. And from statistical point of view is the best since its quadratic loss is smaller than the quadratic losses of both F^{coll} and $\overline{N}$.

- The quadratic loss of F^{coll} is:

$$E\Big[(F^{coll} - \Theta)^2\Big] = E\left[\left(\frac{\gamma}{\beta} - \Theta\right)^2\right] = Var[\Theta] = \frac{\gamma}{\beta^2} .$$

- The quadratic loss of $\overline{N}$ is:

$$E\Big\{E\Big[(\overline{N} - \Theta)^2|\Theta\Big]\Big\} = E\{Var(\overline{N}|\Theta)\} = \frac{E[\Theta]}{n} = \frac{1}{n}\frac{\gamma}{\beta} .$$

- Since we already know that the quadratic loss of F^{Bayes} is smaller than the quadratic loss of $\overline{N}$, then for $n \to \infty$ the last formula will lead to the following result:

$$E\Big[(F^{Bayes} - \Theta)^2\Big] \to \infty .$$

3.3 The Binomial-Beta Case

In group life insurance or group accident insurance, the insurer will take interest in the number of disability cases or equivalently, the disability frequency for a particular group.

Mathematical Model

To simplify the model, a few assumptions will be made:

- each member of the group has the same probability of disablement,
- individual disabilities occur independently,
- once disabled, the person leaves group.

Necessary random variables:

- N_j is the number of new disabilities occurring in the group in the year $j = 1,\ 2, \ldots,$
- V_j is the number of (not disabled) members in the group at the beginning of year $j = 1,\ 2, \ldots$
- $X_j = \frac{N_j}{V_j}$ defines the observed disablement frequency in year $j = 1,\ 2, \ldots$

The task is to identify:

$$X_{n+1} = \frac{N_{n+1}}{V_{n+1}}$$

One might observe that at time n, all random variable with indices smaller than n are known to the insurer and that resolving the above equation resumes to finding out how much N_{n+1} is. Therefore our goal hereon will be to model N_{n+1}.

Model Assumptions (Binomial–Beta)

- *BB1:* Conditionally, given $\Theta = \vartheta$, N_j $(j = 1,\ 2, \ldots, n)$ are independent and binomial distributed with :

$$P[N_j = k|\Theta] = \binom{V_j}{k} \vartheta^k\ (1 - \vartheta)^{V_j - k}.$$

- *BB2:* Θ has a Beta(a, b) distribution with $a, b > 0$ or equivalently, the structural function has density:

$$u(\vartheta) = \frac{1}{B(a,b)}\ \vartheta^{a-1}\ (1-\vartheta)^{b-1}, \qquad 0 \leq \vartheta \leq 1\ ,$$

$$\text{where } B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

As a consequence of the model assumptions the frequencies can be calculated as follows:

THE INDIVIDUAL CLAIM FREQUENCY: $\mathrm{F}^{ind} = \mathrm{E}[\mathrm{X}_{n+1}|\Theta] = \Theta.$

THE COLLECTIVE CLAIM FREQUENCY: $\mathrm{F}^{coll} = \mathrm{E}[\Theta] = \frac{a}{a+b}.$

THE BAYES CLAIM FREQUENCY: $\mathrm{F}^{Bayes} = \frac{a+N_\bullet}{a+b+N_\bullet} = \alpha\overline{N} + (1-\alpha)\frac{a}{a+b}$

$$where\ \overline{N} = \frac{N_\bullet}{V_\bullet}, \quad \alpha = \frac{V_\bullet}{a+b+V_\bullet}$$

In addition the QUADRATIC LOSS of F^{Bayes} is:

$$\mathrm{E}\left[(\mathrm{F}^{Bayes} - \Theta)^2\right] = (1 - \alpha)\ E\left[(F^{coll} - \Theta)^2\right]$$

$$= \alpha\ E\left[(\overline{N} - \Theta)^2\right].$$

3.4 The Normal-Normal Case

This last model is not as relevant to the practical world as the previous two, since insurance data is rarely normally distributed. However it may turn out useful when dealing with large portfolios of data.

Considering an individual risk, we define the following:

- $X = (X_1, \ldots, X_n)'$ is the observation vector,
- X_j is the aggregate claim amount in the j-th year.

Model Assumptions (Normal−Normal)

- *NN1:* Conditionally, given $\Theta = \vartheta$, the X_j's ($j = 1, 2, \ldots, n$) are independent and normally distributed with :

$$X_j \sim \mathcal{N}(\vartheta, \sigma^2)\ .$$

- *NN2:* $\Theta \sim \mathcal{N}(\mu, \tau^2)$, and the structural function has the density:

$$u(\vartheta) = \frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}(\frac{\vartheta\ -\ \mu}{\tau})^2}\ .$$

Under the Model Assumptions we have:

THE INDIVIDUAL PREMIUM: $\mathrm{P}^{ind} = \mathrm{E}[\mathrm{X}_{n+1}|\Theta] = \Theta.$

THE COLLECTIVE PREMIUM: $\mathrm{P}^{coll} = \mathrm{E}[\Theta] = \mu.$

THE BAYES PREMIUM: $\mathrm{P}^{Bayes} = \frac{\tau^2\mu+\sigma^2 X_\bullet}{\tau^2+n\sigma^2} = \alpha\overline{X} + (1-\alpha)E[\Theta]$

$$where\ \overline{X} = \frac{1}{n}X_\bullet, \quad \alpha = \frac{n}{n+\frac{\sigma^2}{\tau^2}}.$$

The QUADRATIC LOSS of P^{Bayes} is:

$$\mathrm{E}\left[(\mathrm{P}^{Bayes} - \Theta)^2\right] = (1 \text{ - } \alpha)\ E\left[(P^{coll} - \Theta)^2\right]$$
$$= \alpha\ E\left[(\overline{X} - \Theta)^2\right].$$

3.5 Similarities between Cases

In all three presented cases, following common features can be noticed:

- The **Bayes premium** is a linear function of the observations and therefore a CREDIBILITY ESTIMATOR. Its formula is a weighted mean between the individual a priori information and the collective premium:

$$P^{Bayes} = \alpha\overline{X} + (1-\alpha)P^{coll} \quad .$$

- The **credibility weight** α is given by the general formula:

$$\alpha = \tfrac{n}{n+\kappa} \quad \text{, where } \kappa \text{ is an appropiate constant.}$$

- The **quadratic loss** of the Bayes premium is given by:

$$\mathrm{E}\left[(\mathrm{P}^{Bayes} - \Theta)^2\right] = (1 \text{ - } \alpha)\ E\left[(P^{coll} - \Theta)^2\right]$$
$$= \alpha\ E\left[(\overline{X} - \Theta)^2\right].$$

- The a posteriori distribution of Θ belongs to the same family as the a priori distribution.

4 Credibility Estimators

"THE CREDIBILITY ESTIMATORS ARE LINEAR BAYES ESTIMATORS!"

In the previous chapters, we denoted the following formula for the Bayes premium:

$$P^{Bayes} = \widetilde{\mu(\Theta)} = E[\ \mu(\Theta)|X\ ]\ ,$$

and proved the fact that this is indeed the best possible estimator in the class of all estimator functions. However, this estimator turns out to be extremely difficult to calculate: first of all because it can only be expressed with help of numerical methods and second of all a lot of information (which in practice cannot be easily obtained) is required: the conditional distribution as well as the a priori distribution.
Credibility theory thus states as a condition the simplicity of the estimator and tries to obtain that by restricting the class of allowable estimator functions to those that are linear in the observations $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ or equivalently, we are looking for the BEST (in Bayesian sense) estimator in the class of all linear estimator functions.
As before, $\mathbf{X}$ defines the observations of the individual risk and $\mu(\Theta)$ stands for the individual premium. Furthermore the components of $\mathbf{X}$ remain, conditional on $\Theta = \vartheta$, independent and identically distributed.

4.1 The Simple Credibility Model

Model Assumptions (Simple Credibility Model)

- *SC1:* The random variables X_j $(j = 1, 2, \ldots, n)$ are, conditional on $\Theta = \vartheta$, independent with the same distribution function F_ϑ with the conditional moments:

$$\begin{aligned} \mu(\vartheta) &= E[X_j|\Theta = \vartheta] \\ \sigma^2(\vartheta) &= Var[X_j|\Theta = \vartheta] \end{aligned}$$

- *SC2:* Θ is a random variable with distribution $U(\vartheta)$.

In this model the individual premium and the collective premium will be defined by the formulas presented in chapter **(2.4)**:

$$\begin{aligned} P^{ind} &= \mu(\Theta) = E[\ X_{n+1}|\Theta\ ], \\ P^{coll} &= \mu_0 = \int_\Theta \mu(\vartheta) dU(\vartheta) = E[X_{n+1}]. \end{aligned}$$

As mentioned before, this time we have to find an estimator for the individual premium $\mu(\Theta)$, which is linear in the observations. Lets denote the best estimator by P^{cred}

or $\widehat{\mu(\Theta)}$, which will have to take the following form:

$$\widehat{\mu(\Theta)} \;=\; \hat{a}_0 \;+\; \sum_{j=1}^{n} \hat{a}_j X_j \quad ,$$

where the real coefficients $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_n$ must satisfy:

$$E\Big[\Big(\mu(\Theta) \;-\; \hat{a}_0 \;-\; \sum_{j=1}^{n} \hat{a}_j X_j\Big)^2\Big] \;=\; \min_{a_0,a_1,\ldots,a_n \in \mathbf{R}} E\Big[\Big(\mu(\Theta) \;-\; a_0 \;-\; \sum_{j=1}^{n} a_j X_j\Big)^2\Big] \quad .$$

In words, $\widehat{\mu(\Theta)}$ will be that estimator which is not only linear in the observations but also for which the quadratic loss function of $\mu(\Theta)$ will be as small as possible. Since in assumption *SC1* we have stated all X_j-s as independent and identically distributed, then the probability distribution of $X_1, X_2, \ldots, X_n$ will remain invariant under permutation of X_j. Therefore all $\hat{a}_i$-s $(i = 1, 2, \ldots, n)$ must be equal and the new form of $\widehat{\mu(\Theta)}$ will be:

$$\widehat{\mu(\Theta)} \;=\; \hat{a} \;+\; \hat{b}\overline{X} \quad , \qquad \textit{where} \quad \overline{X} \;=\; \frac{1}{n}\sum_{j=1}^{n} X_j .$$

Our aim will be now to find these two new parameters $\hat{a}$ and $\hat{b}$, which are solution to the minimizing problem:

$$E\Big[(\mu(\Theta) - \hat{a} - \hat{b}\overline{X})^2\Big] \;=\; \min_{a,b \in \mathbf{R}} E\Big[(\mu(\Theta) - a - b\overline{X})^2\Big].$$

This is equivalent to solving the following set of equations:

$$\frac{\mathrm{d}}{\mathrm{d}a}\, E\Big[(\mu(\Theta) - a - b\overline{X})^2\Big] \;=\; 0 \qquad (1)$$

$$\frac{\mathrm{d}}{\mathrm{d}b}\, E\Big[(\mu(\Theta) - a - b\overline{X})^2\Big] \;=\; 0 \qquad (2)$$

Further:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}a}\, E\Big[(\mu(\Theta) - a - b\overline{X})^2\Big] \;&=\; E\Big[2 \cdot (\mu(\Theta) - a - b\overline{X}) \cdot (-1)\Big] \\
\Leftrightarrow \;& E\Big[\mu(\Theta) - a - b\overline{X}\Big] \;=\; 0
\end{aligned}$$

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}b}\, E\Big[(\mu(\Theta) - a - b\overline{X})^2\Big] \;&=\; E\Big[(-2) \cdot (\mu(\Theta)\overline{X} - a\overline{X} - b\overline{X}^2)\Big] \\
&=\; (-2)\Big(E[\overline{X} \cdot \mu(\Theta)] - E[\overline{X}] \cdot E[\mu(\Theta)] \\
&\quad -\, bE[\overline{X}^2] + b(E[\overline{X}])^2 \\
&\quad +\, E[\overline{X}] \cdot E[\mu(\Theta)] - aE[\overline{X}] - b(E[\overline{X}])^2\Big)
\end{aligned}$$

$$
\begin{aligned}
&= Cov(\overline{X}, \mu(\Theta)) - bVar(\overline{X}) \\
&\quad + E[\overline{X}]\,\underbrace{(E[\mu(\Theta) - a - b\overline{X}])}_{=0} \\
\Leftrightarrow\ & Cov(\overline{X}, \mu(\Theta)) - bVar(\overline{X}) \ = \ 0 \quad .
\end{aligned}
$$

So the final equations to solve are:

$$
\begin{aligned}
E\Big[\mu(\Theta) - a - b\overline{X}\Big] &= 0 \\
Cov(\overline{X}, \mu(\Theta)) - bVar(\overline{X}) &= 0 \quad .
\end{aligned}
$$

Going back to the two model assumptions made at the begining of the chapter, we denote the following:

$$
\begin{aligned}
Cov(\overline{X}, \mu(\Theta)) \ = \ Var(\mu(\Theta)) \quad &=: \quad \tau \ , \\
Var(\overline{X}) \ = \ \frac{E[\sigma^2(\Theta)]}{n} \ + \ Var(\mu(\Theta)) \quad &=: \quad \frac{\sigma^2}{n} \ + \ \tau^2 \ ,
\end{aligned}
$$

and the requested coefficients a and b will look as follows:

$$
\begin{aligned}
b &= \frac{\tau^2}{\tau^2 + \frac{\sigma^2}{n}} = \frac{n}{n + \frac{\sigma^2}{\tau^2}} \quad , \\
a &= (1 - b)\mu_0 \quad .
\end{aligned}
$$

Theorem: The CREDIBILITY ESTIMATOR under the model assumptions *SC1* and *SC2* is given by:

$$
\widehat{\mu(\Theta)} \ = \ \alpha\overline{X} + (1 - \alpha)\mu_0,
$$

where

$$
\begin{aligned}
\mu_0 &= E[\mu(\Theta)], \\
\alpha &= \frac{n}{n + \frac{\sigma^2}{\tau^2}}
\end{aligned}
$$

In words, the credibility estimator P^{cred} is a weighted mean of P^{coll} and the individual observed average $\overline{X}$.

Remarks:

- $\kappa \ := \ \frac{\sigma^2}{\tau^2}$ is called the CREDIBILITY COEFFICIENT.
- by multiplying and deviding through μ_0^2, we can rewrite κ as:

$$
\kappa \ = \ \Big(\frac{\sigma}{\mu_0}\Big)^2 \Big(\frac{\tau}{\mu_0}\Big)^{-2} \quad ,
$$

where

- $\frac{\tau}{\mu_0}$ is the coefficient of variation of $\mu(\Theta)$, which is a good measure for the **heterogeneity** of the portfolio
- $\frac{\sigma}{\mu_0} = \frac{\sqrt{E[Var[X_j|\Theta]}}{E[X_j]}$ is the expected standard deviation within risk devided by the overall expected value, which is a good measure for the within **risk variability**.

- The credibility weight α increases as:
  - the number of the observed years n increases,
  - the heterogeneity of the portfolio increases,
  - the within risk variability decreases.

In case of a collective of similar risks, the structural parameters $\sigma^2,\ \tau^2,\ \mu_0$ that definde the formula for P^{cred} can be estimated using the data from this collective (empirical Bayes procedure). On the other hand these parameters could also be intuitively calculated using the a priori knowledge of an experienced actuary (pure Bayesian point of view).

5 Bibliography

[1] H.BÜHLMANN, A.GISLER – **A Course in Credibility Theory and its Applications** , Springer Verlag, 2005

[2] R.NORBERG – **Credibility Theory**, Department of Statistics, London School of Economics, United Kingdom

[3] H.C.MAHLER, C.G.DEAN – **Credibility**, Casualty Actuarial Department (www.casact.org)

[4] H.SCHMIDLI – **Lecture Notes on Risk Theory**

[5] G.G.VENTER – **Credibility Theory for Dummies** (www.casact.org)