Credibility theory made easy

Autor(en): **Gisler, Alois**

Objekttyp: **Article**

Zeitschrift: **Mitteilungen / Schweizerische Vereinigung der Versicherungsmathematiker = Bulletin / Association Suisse des Actuaires = Bulletin / Swiss Association of Actuaries**

Band (Jahr): **- (1990)**

Heft 1

PDF erstellt am: **24.09.2024**

Persistenter Link: https://doi.org/10.5169/seals-967230

Alois Gisler, Winterthur

Credibility Theory Made Easy

1 Introduction

In classical statistics the following result is well known: Let X_i, $i = 1, 2, \dots, n$, be independent real random variables with $E[X_i] = \mu$ and precision $\varrho_i = \mathrm{Var}^{-1}[X_i]$. Then the minimal variance unbiased linear estimator of μ is given by

$$\widehat{\mu} = \left(\sum \varrho_i\right)^{-1} \sum \varrho_i X_i, \tag{1}$$

i.e. $\widehat{\mu}$ is a weighted mean with the precisions as weights. The precision of $\widehat{\mu}$ is

$$\mathrm{Var}[\widehat{\mu}]^{-1} = \sum \varrho_i \tag{2}$$

In credibility theory one usually wants to estimate the pure risk premium, say $\mu(\Theta)$, based on some statistical information represented by an observable random vector $\boldsymbol{X} = (X_1, X_2, \dots, X_n)'$ and some additional information expressed by a random variable or a constant, say μ_0. In contrast to the above result in classical statistics we do not have to estimate a constant term μ but rather the random variable $\mu(\Theta)$. We will therefore use the terms unbiasedness and precision in a Bayesian sense, i.e.

- an estimator $\widehat{Y}$ of Y is called unbiased if $E[\widehat{Y} - Y] = 0$;
- the precision of an estimator $\widehat{Y}$ is defined by $E[(\widehat{Y} - Y)^2]^{-1}$, that is we use quadratic loss;
- the precision of a random variable X with respect to Y is defined by $E[(X - Y)^2]^{-1}$.

In section 2 we show that there exists a Bayesian equivalent to (1) and (2). Denote by $\widetilde{\mu}(\Theta)$ the best linear unbiased estimator based only on the statistical information $\boldsymbol{X}$. If $X_1, X_2, \dots, X_n$ are conditionally independent with conditional expectation $\mu(\Theta)$, given Θ, it turns out that $\widetilde{\mu}(\Theta)$ is a weighted mean of the X_i with the precisions of the X_i with respect to $\mu(\Theta)$ as weights.

It is also shown that under some basic assumptions the credibility estimator based on both, the statistical information $\boldsymbol{X}$ and the additional information μ_0, is a weighted mean between $\widetilde{\mu}(\Theta)$ and μ_0, the weights again being the precisions with respect to $\mu(\Theta)$. Based on these two fundamental principles and the fact, that credibility estimators are projections on linear subspaces of the Hilbert space of the square integrable random variables, we then find the credibility estimators and their precisions in different models.

In chapter 3 we introduce a basic model covering e.g. the Bühlmann & Straub model. In chapter 4 a class of evolutionary models is considered, whereas chapter 5 is devoted to semilinear credibility. Finally the hierarchical model is dealt with in chapter 6.

The credibility estimators presented in this paper are not new and already known in literature. But the author feels, that this way of looking at credibility gives a deep intuitive insight into the formulae and into credibility theory. The credibility estimators become intuitively plausible and they are derived in a direct and elegant way. Moreover the precisions of the estimators are obtained very easily.

2 Hilbert space technique and basic principles

2.1 Hilbert space technique and some notations

We will make an extensive use of the Hilbert space technique presented e.g. in *De Vylder* (1976 a).

All random variables considered are assumed to belong to $\boldsymbol{L}^2$, which is the Hilbert space of all random variables with finite second order moments. The inner product between two points X and Y in $\boldsymbol{L}^2$ is defined by $\langle X, Y \rangle = E[XY]$.

$\boldsymbol{P}$ is a subspace of $\boldsymbol{L}^2$, if it is non void and contains all linear combinations of its elements. $\boldsymbol{Q}$ is a translated subspace, if it can be written as $\boldsymbol{Q} = Z + \boldsymbol{P}$, where Z is some element of $\boldsymbol{L}^2$ and $\boldsymbol{P}$ is a subspace. Since we may have $Z = 0$, it is seen that a subspace is a particular translated subspace. The point $Q \in \boldsymbol{Q}$ is said to be the orthogonal projection of X on $\boldsymbol{Q}$ ($\text{pro}(X \mid \boldsymbol{Q})$) if $X - Q \perp \boldsymbol{Q}$, i.e. $\langle X - Q, Y_1 - Y_2 \rangle = 0$ for all $Y_1, Y_2 \in \boldsymbol{Q}$.

The following results are basic:

- Linearity
  Let $\boldsymbol{P}$ be any subspace then

$$\text{pro}(aX + bY \mid \boldsymbol{P}) = a \cdot \text{pro}(X \mid \boldsymbol{P}) + b \cdot \text{pro}(Y \mid \boldsymbol{P}) \tag{3}$$

- Iterativity
  Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be any (translated) subspaces with $\boldsymbol{Q} \subset \boldsymbol{P}$. Then

$$\text{pro}(X \mid \boldsymbol{Q}) = \text{pro}(\text{pro}(X \mid \boldsymbol{P}) \mid \boldsymbol{Q}) \tag{4}$$

- "Normed Linearity" in translated subspaces
  Let $\boldsymbol{P}$ be a translated subspace and let $Z = (a+b)^{-1}(aX + bY)$. Then

$$\text{pro}(Z \mid \boldsymbol{P}) = (a+b)^{-1}(a \cdot \text{pro}(X \mid \boldsymbol{P}) + b \cdot \text{pro}(Y \mid \boldsymbol{P})) \tag{5}$$

Notation: Let $\boldsymbol{X} = (X_1, \dots, X_n)'$ be a vector with $X_i \in \boldsymbol{L}^2$ $(i = 1, 2, \dots, n)$.

We denote by
$\boldsymbol{L}(\boldsymbol{X})$ the subspace spanned by the variables $X_1, X_2, \dots, X_n$ and by
$\boldsymbol{L}_Y(\boldsymbol{X})$ the translated subspace $\boldsymbol{Q} = \{Z : Z = \sum a_i X_i,\ E[Z - Y] = 0\}$.

2.2 *Three basic principles*

Theorem 1 (Basic Principle 1) Credibility estimators are projections on subspaces or translated subspaces, i.e. the following results hold true

I) The (inhomogeneous) credibility estimator of Y based on the statistic $\boldsymbol{X}$, i.e. the best estimator of the form $\widehat{Y}^* = a_0 + \sum a_i X_i$, is

$$\widehat{Y} = \text{pro}(Y \mid \boldsymbol{L}(\boldsymbol{X}, 1)) \tag{6}$$

II) The homogeneous credibility estimator of Y based on the statistic $\boldsymbol{X}$, i.e. the best unbiased estimator of Y of the form $\widetilde{Y}^* = \sum a_i X_i$, is

$$\widetilde{Y} = \text{pro}(Y \mid \boldsymbol{L}_Y(\boldsymbol{X})) \tag{7}$$

Interpretation: By definition $\widehat{Y}$ resp. $\widetilde{Y}$ belong to $\boldsymbol{L}(\boldsymbol{X}, 1)$ resp. $\boldsymbol{L}_Y(\boldsymbol{X})$. By definition $\widehat{Y}$ resp. $\widetilde{Y}$ must be the points in $\boldsymbol{L}(\boldsymbol{X}, 1)$ resp. $\boldsymbol{L}_Y(\boldsymbol{X})$ which are closest (with respect to quadratic loss) to Y. It is intuitively clear that the point in a (translated) subspace closest to Y is the orthogonal projection of Y on that (translated) subspace.

Proof: A rigorous proof can be found e.g. in *De Vylder* (1976 a).

Remarks:

- As $\widehat{Y} \in \boldsymbol{L}(\boldsymbol{X}, 1)$ it follows that $\langle Y - \widehat{Y}, 1\rangle = 0$ and hence $E[\widehat{Y}] = E[Y]$. Thus we have

$$\widehat{Y} = \text{pro}(Y \mid L_Y(\boldsymbol{X}, 1)) \tag{8}$$

- Since $\boldsymbol{L}_Y(\boldsymbol{X}) \subset \boldsymbol{L}_Y(\boldsymbol{X}, 1)$, the following relation holds true:

$$\widetilde{Y} = \text{pro}(\widehat{Y} \mid \boldsymbol{L}_Y(\boldsymbol{X})) \tag{9}$$

- From (6) it follows that $\widehat{Y}$ is the (inhomogeneous) credibility estimator if and only if it fulfills the normal equations

$$\langle Y - \widehat{Y}, 1\rangle = 0 \qquad \text{i.e. } E[\widehat{Y}] = E[Y] \tag{10}$$

$$\langle Y - \widehat{Y}, X_i\rangle = 0 \qquad i = 1, 2, \ldots, n \tag{11}$$

Analogously it follows from (7) that $\widetilde{Y}$ is the homogeneous credibility estimator if and only if it fulfills the normal equations

$$E[\widetilde{Y}] = E[Y] \tag{12}$$

$$\langle Y - \widetilde{Y}, X_i - X_j\rangle = 0 \qquad \text{for } \; i, j = 1, 2, \ldots, n \tag{13}$$

Theorem 2 (Basic Principle 2) Let Θ be a risk parameter (random variable) and let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ be an observable vector fulfilling the assumptions

A 1: $X_1, X_2, \ldots, X_n$ are conditionally independent, given Θ

A 2: $E(X_i \mid \Theta] = \mu(\Theta) \qquad i = 1, 2, \ldots, n$

Then

I) $$\widetilde{\mu}(\Theta) := \text{pro}[\mu(\Theta) \mid \boldsymbol{L}_{\mu(\Theta)}(\boldsymbol{X})] = \left(\sum \varrho_i\right)^{-1} \left(\sum \varrho_i X_i\right) \tag{14}$$

where $\varrho_i := E\left[(X_i - \mu(\Theta))^2\right]^{-1}$

II) $$\widetilde{\varrho} := E\left[(\widetilde{\mu}(\Theta) - \mu(\Theta))^2\right]^{-1} = \sum \varrho_i \tag{15}$$

Remarks:

– Theorem 2 states that the homogeneous credibility estimator of $\mu(\Theta)$ based on $\boldsymbol{X}$ is a weighted mean with the precisions of the X_i with respect to $\mu(\Theta)$ as weights. Compared with (1) it becomes obvious that (14) is the Bayesian counterpart to (1).

– The estimator $\widetilde{\mu}(\Theta)$ does not change if all precisions are multiplied by a constant factor. Thus the precisions have to be known up to a constant factor only.

Proof:

Proof of I): We show that $\widetilde{\mu}(\Theta)$ fulfills the normal equations (12) and (13).

$E[\widetilde{\mu}(\Theta)] = E[\mu(\Theta)]$ because of A 2

$$\begin{aligned}\langle \mu(\Theta) - \widetilde{\mu}(\Theta), X_j \rangle &= \text{const}\left\{ \sum \varrho_i \langle \mu(\Theta) - X_i, X_j \rangle \right\} \\ &= \text{const}\varrho_j \langle \mu(\Theta) - X_j, X_j \rangle && \text{(A 1 A 2)} \\ &= \text{const}\varrho_j \langle \mu(\Theta) - X_j, X_j - \mu(\Theta) \rangle \\ &= \text{const} && \text{(A 2)}\end{aligned}$$

and hence

$$\langle \mu(\Theta) - \widetilde{\mu}(\Theta), X_i - X_j \rangle = 0 \qquad \text{for} \quad i \neq j$$

Proof of II)

$$E\left[(\widetilde{\mu}(\Theta) - \mu(\Theta))^2\right] = \left(\sum \varrho_i\right)^{-2} \sum \varrho_i^2 E\left[(X_i - \mu(\Theta))^2\right] = \left(\sum \varrho_i\right)^{-1}$$

Theorem 3 (Basic principle 3) Once more let Θ be a risk parameter and $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ an observable vector with $E[X_i \mid \Theta] = \mu(\Theta)$. Let $\widetilde{\mu}(\Theta)$ be the homogeneous credibility estimator of $\mu(\Theta)$ based on $\boldsymbol{X}$. Let μ_0 be a random variable or a constant fulfilling the conditions

A 1: $E[\mu_0 - \mu(\Theta)] = 0$ (unbiasedness)

A 2: $\langle \mu(\Theta) - \mu_0, \mu_0 \rangle = 0$ (orthogonality)

A 3: μ_0 and $\boldsymbol{X}$ are conditionally independent given Θ

Then

I) $$\widehat{\mu}(\Theta) := \text{pro}(\mu(\Theta) \mid \boldsymbol{L}(\boldsymbol{X}, \mu_0, 1)) = (\varrho_0 + \widetilde{\varrho})^{-1}(\varrho_0 \mu_0 + \widetilde{\varrho}\,\widetilde{\mu}(\Theta)) \tag{16}$$

where $\varrho_0 = E\left[(\mu_0 - \mu(\Theta))^2\right]^{-1}$

$\widetilde{\varrho} = E\left[(\widetilde{\mu}(\Theta) - \mu(\Theta))^2\right]^{-1}$

II) $$\varrho := E\left[(\widehat{\mu}(\Theta) - \mu(\Theta))^2\right]^{-1} = \varrho_0 + \widetilde{\varrho} \tag{17}$$

III) $\widehat{\mu}(\Theta)$ itself fulfills the conditions A 1 and A 2, *i.e.*

$$E[\mu(\Theta) - \widehat{\mu}(\Theta)] = 0, \qquad \langle \mu(\Theta) - \widehat{\mu}(\Theta), \mu(\Theta) \rangle = 0 \tag{18}$$

Interpretation:

- We want to estimate the pure risk premium $\mu(\Theta)$. On the one hand there are available statistical observations X_i which are conditionally unbiased and on the other hand there is the quantity μ_0 saying something about $\mu(\Theta)$. By definition the best unbiased linear estimator based only on the statistical information $\boldsymbol{X}$ is $\widetilde{\mu}(\Theta)$. By theorem 3 the best linear estimator taking into account both, the statistical information $\boldsymbol{X}$ and the information μ_0, is a weighted mean between $\widetilde{\mu}(\Theta)$ and μ_0 with the precisions as weights.
- μ_0 might be a manual premium, a previous credibility estimator, the assessment of a technical expert etc. The orthogonal condition A 2 means that $\mu_0 = \text{pro}(\mu(\Theta) \mid \boldsymbol{L}(\mu_0))$ and hence that μ_0 is in some sense a credibility estimator, if the statistical information $\boldsymbol{X}$ were not available.

Further remark Note that it is not required that $X_1, X_2, \ldots, X_n$ are conditionally independent given Θ.

Proof: To prove I) we have to show that $\widehat{\mu}(\Theta)$ fulfills the normal equations (10) and (11). Obviously $E[\widehat{\mu}(\Theta) - \mu(\Theta)] = 0$.

$$\langle \mu(\Theta) - \widehat{\mu}(\Theta), \mu_0 \rangle = \text{const}\{\varrho_0 \langle \mu(\Theta) - \mu_0, \mu_0 \rangle + \widetilde{\varrho} \langle \mu(\Theta) - \widetilde{\mu}(\Theta), \mu_0 \rangle\}.$$

The first term in { } equals 0 because of A 2. From the conditional unbiasedness of the r.v. X_i follows that $E[\widetilde{\mu}(\Theta) \mid \Theta] = \mu(\Theta)$. Hence the second term in { } equals 0 because of A 3.

$$\langle \mu(\Theta) - \widehat{\mu}(\Theta), X_i \rangle = \text{const}\{\varrho_0 \langle \mu(\Theta) - \mu_0, X_i \rangle + \widetilde{\varrho} \langle \mu(\Theta) - \widetilde{\mu}(\Theta), X_i \rangle\}$$

$$\langle \mu(\Theta) - \mu_0, X_i \rangle \underset{(A\,3)}{=} \langle \mu(\Theta) - \mu_0, \mu(\Theta) \rangle \underset{(A\,2)}{=} \varrho_0^{-1}$$

$$\begin{aligned}\langle \mu(\Theta) - \widetilde{\mu}(\Theta), X_i \rangle &= \langle \mu(\Theta) - \widetilde{\mu}(\Theta), X_i - \widetilde{\mu}(\Theta) + \widetilde{\mu}(\Theta) \rangle \\ &= \langle \mu(\Theta) - \widetilde{\mu}(\Theta), \widetilde{\mu}(\Theta) \rangle = -\widetilde{\varrho}^{-1}\end{aligned}$$

and hence

$$\langle \mu(\Theta) - \widehat{\mu}(\Theta), X_i \rangle = 0$$

which completes the proof of I).

$$\begin{aligned}E\left[(\widehat{\mu}(\Theta) - \mu(\Theta))^2\right] &\underset{(A\,3)}{=} (\varrho_0 + \widetilde{\varrho})^{-2} \Big(\varrho_0^2 E[(\mu_0 - \mu(\Theta))^2] \\ &\qquad + \widetilde{\varrho}^2 E[(\mu(\Theta) - \widetilde{\mu}(\Theta))^2]\Big)^{-1} \\ &= (\varrho_0 + \widetilde{\varrho})^{-1}\end{aligned}$$

which is identical to II)

Finally III) follows directly from the fact that

$$\widehat{\mu}(\Theta) = \text{pro}\left(\mu(\Theta) \mid \boldsymbol{L}(\boldsymbol{X}, \mu_0, 1)\right).$$

For the purposes of this paper it is convenient to state the following results.

Lemma 1 Let μ_0 and $\boldsymbol{X}$ be as in theorem 3. Let $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_m)'$ be such that $E[Z_j] = E[\mu(\Theta)]$, $(\mu(\Theta) - \mu_0) \perp \boldsymbol{L}(\boldsymbol{Z})$, $(\mu(\Theta) - X_i) \perp \boldsymbol{L}(\boldsymbol{Z})$ for $i = 1, 2, \dots, n$.
Then the credibility estimator of $\mu(\Theta)$ based on $\boldsymbol{X}$, μ_0 and Z does not depend on $\boldsymbol{Z}$, i.e.

$$\text{pro}\left(\mu(\Theta) \mid \boldsymbol{L}(\mu_0, \boldsymbol{X}, \boldsymbol{Z}, 1)\right) = \text{pro}\left(\mu(\Theta) \mid \boldsymbol{L}(\mu_0, \boldsymbol{X}, 1)\right) \tag{19}$$

Proof: Let $\widehat{\mu}(\Theta) = \text{pro}\left(\mu(\Theta) \mid \boldsymbol{L}(\mu_0, \boldsymbol{X}, 1)\right)$. We have to show that

$$(\mu(\Theta) - \widehat{\mu}(\Theta)) \perp \boldsymbol{L}(\boldsymbol{Z}).$$

$\widehat{\mu}(\Theta) = a_0\mu_0 + \sum a_i X_i$ with $a_0 + \sum a_i = 1$ because of theorem 3. Hence

$$\langle \mu(\Theta) - \widehat{\mu}(\Theta), Z_j \rangle = a_0 \langle \mu(\Theta) - \mu_0, Z_j \rangle + \sum a_i \langle \mu(\Theta) - X_i, Z_j \rangle = 0$$

for $j = 1, 2, \ldots, m$.

Lemma 2 Let $Y \in \boldsymbol{L}^2$ and $\{\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{in})' \quad i = 1, 2, \ldots, I\}$ be real random vectors. Assume $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are conditionally independent given Y and $E[X_{ij} \mid Y] = Y$ for all i, j. Let $\widetilde{Y}_i = \text{pro}(Y \mid \boldsymbol{L}_Y(\boldsymbol{X}_i))$ for $i = 1, 2, \ldots, I$. Then

$$\widetilde{Y} := \text{pro}(Y \mid \boldsymbol{L}_Y(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_I)) = \left(\sum \widetilde{\varrho}_i\right)^{-1} \sum \widetilde{\varrho}_i \widetilde{Y}_i \tag{20}$$

where $\widetilde{\varrho}_i = E\left[(\widetilde{Y}_i - Y)^2\right]^{-1}$

Proof: We will show that

$$\widetilde{Y} = \text{pro}(Y \mid \boldsymbol{L}_Y(\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_I)).$$

Then (20) is a consequence of theorem 2. Let

$$\widetilde{Y}^* := \text{pro}(Y \mid \boldsymbol{L}_Y(\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_I)) = \sum a_i \widetilde{Y}_i \quad \text{with} \quad \sum a_i = 1.$$

$$\begin{aligned}
\langle Y - \widetilde{Y}^*, X_{ij} - X_{ik} \rangle &= \sum_{\ell} a_\ell \langle Y - \widetilde{Y}_\ell, X_{ij} - X_{ik} \rangle \\
&= a_i \langle Y - \widetilde{Y}_i . X_{ij} - X_{ik} \rangle = 0 \\
\langle Y - \widetilde{Y}^*, X_{ij} - X_{kl} \rangle &= \langle Y - \widetilde{Y}^*, (X_{ij} - \widetilde{Y}_i) + (\widetilde{Y}_i - \widetilde{Y}_k) + (\widetilde{Y}_k - X_{kl}) \rangle \\
&= 0.
\end{aligned}$$

Hence

$$\widetilde{Y}^* = \widetilde{Y}.$$

Lemma 3 Let $Y, Z, \boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ be such that

$$E[Z - Y] = 0, \quad \langle Z - Y, X_i - X_j \rangle = 0 \qquad \text{for} \quad i, j = 1, 2, \ldots, n.$$

Let $\widetilde{Y}$ resp. $\widetilde{Z}$ be the homogeneous credibility estimator of Y resp. of Z based on $\boldsymbol{X}$. Then

$$\widetilde{Y} = \widetilde{Z} \tag{21}$$

Proof:

$$\begin{aligned}\langle Z - \widetilde{Y}, X_i - X_j \rangle &= \langle (Z - Y) + (Y - \widetilde{Y}), X_i - X_j \rangle \\ &= 0 \qquad \text{for} \quad i, j = 1, 2, \ldots, n.\end{aligned}$$

Hence $\widetilde{Y} = \text{pro}(Z \mid \boldsymbol{L}_Z(\boldsymbol{X}))$.

Lemma 4 Let Y and X be points in $\boldsymbol{L}^2$ and $\mu_Y = E[Y]$, $\mu_X = E[X]$. Then

$$\widehat{Y} = \text{pro}(Y \mid \boldsymbol{L}(X, 1)) = \mu_Y + \frac{\text{Cov}\,(X, Y)}{\text{Var}[X]}(X - \mu_X) \tag{22}$$

This result is well known. The proof is an easy exercise in applying the normal equations and is left to the reader.

3 Credibility in a basic model

3.1 Basic Model

Consider a policy characterised by a hidden risk parameter Θ. Suppose the observable vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ and μ_0 satisfy

A 1: $E[X_i \mid \Theta] = \mu(\Theta)$

A 2: The random variables $\mu_0, X_1, X_2, \ldots, X_n$ are conditionally independent, given Θ.

A 3: $E[\mu_0 - \mu(\Theta)] = 0$

A 4: $\langle \mu(\Theta) - \mu_0, \mu_0 \rangle = 0$

Of course we want to estimate the pure risk premium $\mu(\Theta)$.

We denote by

$$\varrho_0 := E\left[(\mu_0 - \mu(\Theta))^2\right]^{-1} \quad \text{the precision of } \mu_0 \text{ with respect to } \mu(\Theta);$$

$$\varrho_i := E\left[(X_i - \mu(\Theta))^2\right]^{-1} \quad \text{the precision of } X_i \text{ with respect to } \mu(\Theta);$$

$$\mu := E[X_i] = E[\mu(\Theta)] \quad \text{the unconditional expectation.}$$

We will see in section 3.3 that this basic model covers the Bühlmann & Straub model as well as some other models encountered in literature.

3.2 Credibility estimators

Theorem 4

I) The credibility estimator based on X and μ_0 is

$$\widehat{\mu}(\Theta) = \left(\varrho_0 + \sum \varrho_i\right)^{-1}\left(\varrho_0\mu_0 + \sum \varrho_i X_i\right) \tag{23}$$

II) The precision of $\widehat{\mu}(\Theta)$ is

$$\varrho := E\left[(\widehat{\mu}(\Theta) - \mu(\Theta))^2\right]^{-1} = \varrho_0 + \sum \varrho_i \tag{24}$$

III) $\widehat{\mu}(\Theta)$ fulfills the conditions A 3 and A 4, i.e.

$$E\left[\widehat{\mu}(\Theta) - \mu(\Theta)\right] = 0 \quad \text{and} \quad \langle \mu(\Theta) - \widehat{\mu}(\Theta), \widehat{\mu}(\Theta)\rangle = 0 \tag{25}$$

Proof: Theorem 4 is a direct consequence of theorem 2 and theorem 3.

Corollary 1 (Recursive credibility formula) Let $\widehat{\mu}_k$ be the credibility estimator of $\mu(\Theta)$ based on $(\mu_0, X_1, X_2, \ldots, X_k)$ and let ϱ_k^* be its precision. The following recursion holds true:

$$\widehat{\mu}_k = (\varrho_{k-1}^* + \varrho_k)^{-1}(\varrho_{k-1}^*\widehat{\mu}_{k-1} + \varrho_k X_k) \qquad k = 1, 2, \ldots \tag{26}$$

$$\varrho_k^* = (\varrho_{k-1}^* + \varrho_k) \tag{27}$$

$$\widehat{\mu}_0 = \mu_0 \qquad \varrho_0^* = \varrho_0$$

3.3 Examples

Example 1: The model of Bühlmann/Straub (1970)

Bühlmann & Straub consider a portfolio of risks. Each risk i ($i = 1, 2, \ldots, I$) is characterised by a hidden risk parameter Θ_i. To each risk i belongs an observation vector $X_i = (X_{i1}, X_{i2}, \ldots, X_{in})'$ where X_{ij} may be interpreted as an observation (e.g. claim amount) of risk i in period j.

Assumptions:

BS 1: Conditionally, given Θ_i, the random variables $X_{i1}, X_{i2}, \ldots, X_{in}$ are independent with

$$E[X_{ij} \mid \Theta_i] = \mu(\Theta_i)$$

$$\mathrm{Var}[X_{ij} \mid \Theta_i] = \frac{\sigma^2(\Theta_i)}{P_{ij}}$$

where P_{ij} are known constants (volume measures)

BS 2: The pairs $(\Theta_1, X_1), (\Theta_2, X_2), \ldots, (\Theta_I, X_I)$ are independent and $\Theta_1, \Theta_2, \ldots, \Theta_I$ are independent and identically distributed (i.i.d.).

From Lemma 1 it follows that the credibility estimator $\mu(\Theta_i)$ depends only on the data X_i and not on X_k for $k \neq i$. (μ, X_i) fulfill the conditions A 1 – A 4 of theorem 4. Hence we immediately obtain the credibility formula, the recursive formula and the formula for the precision.

If we denote by

$$\varrho_{ij} := E\left[(X_{ij} - \mu(\Theta_i))^2\right]^{-1} = v^{-1} P_{ij} \qquad j = 1, 2, \ldots, n_i$$

where $v = E[\sigma^2(\Theta_i)]$

$$\varrho_0 := E\left[(\mu - \mu(\Theta_i))^2\right]^{-1} = w^{-1}$$

$$P_i := \sum_j P_{ij} \qquad \varrho_i := \sum_j \varrho_{ij}$$

we get

$$\widehat{\mu}(\Theta_i) = (\varrho_0 + \varrho_i)^{-1}(\varrho_0 \mu + \varrho_i \widetilde{\mu}^*(\Theta_i)) \tag{28}$$

where

$$\widetilde{\mu}^*(\Theta_i) = \varrho_i^{-1}\Big(\sum_j \varrho_{ij} X_{ij}\Big) = (P_i)^{-1} \sum_j P_{ij} X_{ij}$$

$$\sigma_i = E\left[(\widehat{\mu}(\Theta_i) - \mu(\Theta_i))^2\right]^{-1} = \varrho_0 + \varrho_i \tag{29}$$

Of course (28) can also be written as

$$\widehat{\mu}(\Theta_i) = \alpha_i \widetilde{\mu}^*(\Theta_i) + (1 - \alpha_i)\mu \tag{30}$$

where $\alpha_i = P_i(P_i + v/w)^{-1}$.

If we denote by $\widehat{\mu}_{ki}$ the credibility estimator of $\mu(\Theta_i)$ based on $(X_{i1}, X_{i2}, \ldots, X_{ik})$ and by σ_{ik} its precision, we obtain from corollary 1 the *recursive formula*

$$\widehat{\mu}_{ki} = (\sigma_{i,k-1} + \varrho_{ik})^{-1}(\sigma_{i,k-1}\widehat{\mu}_{k-1} + \varrho_{ik}X_{ik}) \tag{31}$$

$$\sigma_{ik} = \sigma_{i,k-1} + \varrho_{ik} \tag{32}$$

$$\widehat{\mu}_{0i} = \mu \qquad \sigma_{i0} = \varrho_0$$

Example 2: The model used by Campbell (1986)

In addition to the observations X_{ij} fulfilling the assumptions of the Bühlmann & Straub model, a technical expert makes an assessment $\mu^*(\Theta_i)$ for every risk i. It is assumed that the technical assessment fulfills the following assumptions:

$E[\mu^*(\Theta_i) \mid \Theta_i] = \mu(\Theta_i)$;

Conditionally, given Θ_i, $\mu^*(\Theta_i)$ is independent of the random variables X_{ij} $\quad j = 1, \ldots, n_i$.

Hence for any risk i the assumptions A1–A4 of theorem 4 are fulfilled. If we denote by $\widehat{\mu}(\Theta_i)$ the credibility estimator given by (26) and by $\widehat{\widehat{\mu}}(\Theta_i)$ the credibility estimator based on the observation vector $\boldsymbol{X}_i$ and the technical assessment $\mu^*(\Theta_i)$ we immediately get by the recursive formula

$$\widehat{\widehat{\mu}}(\Theta_i) = \alpha\widehat{\mu}(\Theta_i) + \beta\mu^*(\Theta_i)$$

where

$$\alpha = (\tau_i + \sigma_i)^{-1}\sigma_i$$

$$\beta = (\tau_i + \sigma_i)^{-1}\tau_i$$

$\tau_i =$ precision of $\mu^*(\Theta_i)$

$\sigma_i =$ precision of $\widehat{\mu}(\Theta_i)$ $\qquad$ (see (29))

Example 3: Sundt (1987)
In the paper "Credibility and Old Estimates" Sundt discusses under what conditions it is favourable to replace the constant term in the credibility estimator by an old estimator. In the most simple case of the Bühlmann & Straub model the question is under what conditions should the constant term μ in the credibility estimator be replaced by an old estimator $\dot{\mu}(\Theta)$, which is assumed to be conditionally independent of $\boldsymbol{X}$ given Θ. From theorem 4 it becomes obvious that replacement of μ by $\dot{\mu}(\Theta)$ will improve the credibility estimator if and only if the precision of $\dot{\mu}(\Theta)$ is greater than the precision of μ with respect to $\mu(\Theta)$.

3.4 *The homogeneous credibility estimator in the Bühlmann / Straub model*

Let us go back to example 1 of section 3.3 and assume that the overall mean μ is unknown. The following result holds true:

I) The homogeneous credibility estimator of $\mu(\Theta_i)$ based on $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_i$ is

$$\widetilde{\mu}(\Theta_i) = (\varrho_0 + \varrho_i)^{-1}(\varrho_0\widetilde{\mu} + \varrho_i\widetilde{\mu}^*(\Theta_i)) \tag{33}$$

where

$$\widetilde{\mu} = \left(\sum \varrho_i'\right)^{-1}\left(\sum \varrho_i'\widetilde{\mu}^*(\Theta_i)\right) \tag{34}$$

$$\varrho_i' := E\left[(\mu - \widetilde{\mu}^*(\Theta_i))^2\right]^{-1} = \varrho_0\varrho_i(\varrho_0 + \varrho_i)^{-1}$$

II) The precision of $\widetilde{\mu}(\Theta_i)$ is

$$\begin{aligned}\widetilde{\varrho} &:= E\left[(\widetilde{\mu}(\Theta_i - \mu(\Theta_i))^2\right]^{-1} \\ &= (\varrho_0 + \varrho_i)\left(1 + \frac{\varrho_0}{\varrho_0 + \varrho_i}\cdot\frac{\varrho_0}{\sum \varrho_i'}\right)^{-1}\end{aligned} \tag{35}$$

Remarks

- Note that the homogeneous credibility estimator is obtained by replacing the unknown μ in (28) by $\widetilde{\mu}$ which is the homogeneous credibility estimator of μ.

- $\widetilde{\mu}$ is a weighted mean of the $\widetilde{\mu}^*(\Theta_i)$ with the precisions of the $\widetilde{\mu}^*(\Theta_i)$ with respect to μ as weights. Hence the homogeneous credibility estimator of μ is found by a two stage procedure:
  - I) For $i = 1, 2, \ldots, I$ calculate the homogeneous credibility estimator of $\mu(\Theta_i)$ based on X_i, which is a weighted mean of the X_{ij} $(j = 1, 2, \ldots, n_i)$ and the weights being the precisions of the X_{ij} with respect to $\mu(\Theta_i)$.
  - II) Calculate $\widetilde{\mu}$ by taking a weighted mean of the estimators found in step I), the weights being the precisions with respect to μ.
- Denote by $\alpha_i = \varrho_i(\varrho_0 + \varrho_i)^{-1}$ the credibility weight given to $\widetilde{\mu}^*(\Theta_i)$ in (28). Since $\varrho_i' = \alpha_i \varrho_0$, we can replace the ϱ_i' in (34) by α_i, which is the formula usually encountered in literature.

Proof: From (4) and (5) it follows that $\widetilde{\mu}(\Theta_i)$ is obtained by replacing μ in (28) by

$$\widetilde{\mu} = \mathrm{pro}\left(\mu \mid \boldsymbol{L}_\mu(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_k)\right) .$$

Lemma 3 and Lemma 2 yield

$$\widetilde{\mu}_i := \mathrm{pro}\left(\mu \mid \boldsymbol{L}_\mu(\boldsymbol{X}_i)\right) = \widetilde{\mu}^*(\Theta_i)$$

$$\widetilde{\mu} = \sum \varrho_i' \widetilde{\mu}^*(\Theta_i)$$

where

$$\begin{aligned} \varrho_i' &= E\left[(\widetilde{\mu}^*(\Theta_i) - \mu)^2\right]^{-1} \\ &= E\left[(\widetilde{\mu}^*(\Theta_i) - \mu(\Theta_i) + \mu(\Theta_i) - \mu)^2\right]^{-1} \\ &= (\varrho_i^{-1} + \varrho_0^{-1})^{-1} \end{aligned}$$

which is equivalent to (34).

To prove (35) note that

$$\begin{aligned} \mu(\Theta_i) - \widetilde{\mu}(\Theta_i) &= \left(\mu(\Theta_i) - \widehat{\mu}(\Theta_i)\right) + \left(\widehat{\mu}(\Theta_i) - \widetilde{\mu}(\Theta_i)\right) \\ &= \left(\mu(\Theta_i) - \widehat{\mu}(\Theta_i)\right) + \varrho_0(\varrho_0 + \varrho_i)^{-1}(\mu - \widetilde{\mu}) \end{aligned}$$

From (9) we get

$$\langle \mu(\Theta_i) - \widehat{\mu}(\Theta_i),\ \mu - \widetilde{\mu} \rangle = 0 .$$

Hence from (29) and theorem 2 we obtain

$$E\left[(\mu(\Theta_i)-\widetilde{\mu}(\Theta_i))^2\right]=(\varrho_0+\varrho_i)^{-1}+\varrho_0^2(\varrho_0+\varrho_i)^{-2}\left(\sum\varrho_i'\right)^{-1}$$
$$=(\varrho_0+\varrho_i)^{-1}\left(1+\frac{\varrho_0}{\varrho_0+\varrho_i}\cdot\frac{\varrho_0}{\sum\varrho_i'}\right).$$

At this stage it is worthwhile to recall two facts found in the Bühlmann & Straub model:

- Given μ, the credibility estimator of $\mu(\Theta_i)$ depends only on the data X_i and not on X_k for $k \neq i$.
- The homogeneous estimators $\widetilde{\mu}^*(\Theta_i)$ and $\widetilde{\mu}$ are found by a recursive procedure from bottom up.

Exactly the same arguments as above can be applied to the hierarchical model. This will be shown in section 6.

4 The credibility formula in a class of evolutionary models

Let Θ be a random risk parameter and let $(X_1, X_2, \ldots)$ be a sequence of observable random variables where X_i might be interpreted as an observation (e.g. total claim amount) of a particular policy in period i. It is assumed that $X_1, X_2, \ldots$ are conditionally independent given Θ. We further assume that the risk characteristic may change in time. Hence Θ is not a single risk parameter, but rather a sequence $(\Theta_1, \Theta_2, \ldots)$ where Θ_i describes the risk characteristic in year i. We will use the notation $\mu_i(\Theta)$ for $E[X_i \mid \Theta]$ (of course we could use as well $\mu_i(\Theta_i)$ or $\mu(\Theta_i)$).
Gerber/Jones (1975) were among the first to consider such models. Later on, evolutionary models have been discussed in a number of papers (e.g. *Sundt* (1981), *Kremer* (1982), *Sundt* (1982)).
We will derive the credibility formula in the case where the process $\{\mu_i(\Theta);\ i = 1, 2, \ldots\}$ fulfills the assumptions

A 1: $E[\mu_1(\Theta)] = \mu < \infty$ $\qquad$ $\mathrm{Var}[\mu_1(\Theta)] = \lambda < \infty$

A 2: $\mu_{i+1}(\Theta) - \mu = a_i(\mu_i(\Theta) - \mu) + \varepsilon_{i+1}$ $\qquad$ $i = 1, 2, \ldots$
where $a_1, a_2, \ldots$ are constants and where $\mu_1(\Theta), \varepsilon_2, \varepsilon_3, \ldots$ are uncorrelated with $E[\varepsilon_i] = 0$ $\quad$ $\mathrm{Var}[\varepsilon_i] = \sigma_i^2 < \infty$

Remarks:

- If $a_i = 1$ and $\sigma_i^2 = \sigma^2$ for $i = 1, 2, \ldots$ then the process $\{\mu_i(\Theta);\ i = 1, 2, \ldots\}$ is a random walk. This case has been discussed in *Gerber/Jones* (1975).
- If $a_i = a$ with $|a| < 1$, $\sigma_i^2 = \sigma^2$, $\lambda = (1 - a^2)\sigma^2$ then the process $\{\mu_i(\Theta);\ i = 1, 2, \ldots\}$ is a stationary autoregressive process of order 1 (AR(1)-process). The credibility formula for this case can be found e.g. in *Kremer* (1982).
- Cov $(\mu_{i+1}(\Theta), \mu_j(\Theta) = a_i$ Cov $(\mu_i(\Theta), \mu_j(\Theta))$ for $i > j$, which is the general assumption in *Sundt* (1981). Hence the general case has already been covered by Sundt. Below we give an alternative and very simple derivation of the credibility formula.

Given the observations up to time n, we want to estimate the pure risk premium $\mu_{n+1}(\Theta)$ of the next period.
From (3) and A 2 we get

$$\widehat{\mu}_{n+1}(\Theta) = a_n \widehat{\mu}_n'(\Theta) + (1 - a_n)\mu$$

where

$$\widehat{\mu}_n'(\Theta) = \text{pro}\left(\mu_n(\Theta) \mid L(X_1, X_2, \ldots, X_n, 1)\right) \ ;$$
$$E\left[(\widehat{\mu}_{n+1}(\Theta) - \mu_{n+1}(\Theta))^2\right] = E\left[a_n^2(\widehat{\mu}_n'(\Theta) - \mu_n(\Theta))^2\right] + E[\varepsilon_{n+1}^2],$$

which together with theorem 2 yields the following recursion.

Theorem 5

$$\widehat{\mu}_{n+1}(\Theta) = a_n\left\{(\varrho_n + \varphi_n)^{-1}(\varrho_n \widehat{\mu}_n(\Theta) + \varphi_n X_n)\right\} + (1 - a_n)\mu \qquad n = 1, 2, \ldots \tag{36}$$

where

$$\varrho_n = E\left[(\widehat{\mu}_n(\Theta) - \mu_n(\Theta))^2\right]^{-1}$$
$$\varphi_n = E\left[(X_n - \mu_n(\Theta))^2\right]^{-1}$$

$$\varrho_{n+1} = \left\{a_n^2(\varrho_n + \varphi_n)^{-1} + \sigma_{n+1}^2\right\}^{-1} \tag{37}$$
$$\widehat{\mu}_1(\Theta) = \mu \qquad \varrho_1 = \lambda^{-1}$$

5 Semilinear Credibility

Semilinear credibility has been introduced by *De Vylder* (1976 b). The starting point is the homogeneous credibility model defined by a random risk characteristic Θ and an observation vector $\boldsymbol{X} = (X_1, X_2, \dots, X_n)'$, whereby the random variables $X_1, X_2, \dots, X_n$ are conditionally independent and identically distributed given Θ. Let now f be a real function of one real variable and $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)'$ the vector of the transformed variables $Y_i = f(X_i)$.
We will further use the notation

$$\begin{aligned}
\mu_X(\Theta) &= E[X_i \mid \Theta] & \mu_X &= E[\mu_X(\Theta)] \\
\mu_Y(\Theta) &= E[Y_i \mid \Theta] & \mu_Y &= E[\mu_Y(\Theta)] \\
\varrho &= E\left[(\mu_Y(\Theta) - \mu_Y)^2\right]^{-1} & \tau &= E\left[(\mu_Y(\Theta) - Y_i)^2\right]^{-1} \\
\kappa &= \operatorname{Cov}\left[\mu_X(\Theta), \mu_Y(\Theta)\right]
\end{aligned}$$

The best estimator of $\mu_X(\Theta)$ of the form $a_0 + \sum a_i Y_i$ is called the *semilinear credibility estimator*, which we will denote by $\widehat{\mu}_X^f(\Theta)$.

From Lemma 4 we get

$$\mu_X'(\Theta) := \operatorname{pro}(\mu_X(\Theta) \mid \boldsymbol{L}(\mu_Y(\Theta), 1)) = \mu_X + \kappa\varrho(\mu_Y(\Theta) - \mu_Y)$$

Since

$$\langle \mu_X(\Theta) - \mu_X'(\Theta), Y_i \rangle = \langle \mu_X(\Theta) - \mu_X'(\Theta), \mu_Y(\Theta) \rangle = 0$$

it follows that

$$\mu_X'(\Theta) = \operatorname{pro}(\mu_X(\Theta) \mid \boldsymbol{L}(\boldsymbol{Y}, \mu_Y(\Theta), 1)).$$

Hence from (4) we obtain

$$\widehat{\mu}_X^f(\Theta) = \mu_X + \kappa\varrho(\widehat{\mu}_Y(\Theta) - \mu_Y),$$

where $\widehat{\mu}_Y(\Theta)$ is the credibility estimator of $\mu_Y(\Theta)$ based on $\boldsymbol{Y}$.

From theorem 4 it follows that

$$\widehat{\mu}_Y(\Theta) = \mu_Y + (\varrho + n\tau)^{-1} n\tau(\overline{Y} - \mu_Y), \qquad \text{where } \overline{Y} = n^{-1} \sum Y_i$$

and hence

$$\widehat{\mu}_X(\Theta) = \mu_X + (\varrho + n\tau)^{-1} n\tau\kappa\varrho(\overline{Y} - \mu_Y) \tag{38}$$

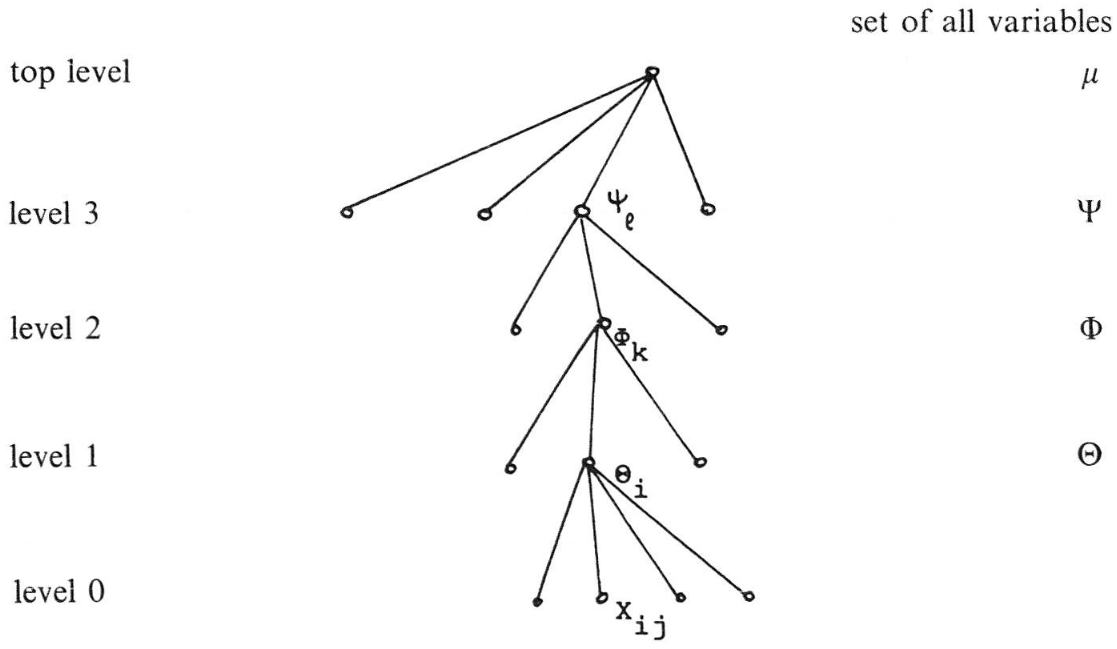
which is equivalent to formula (26) in *De Vylder* (1976 b).

6 Hierarchical credibility

Models with a hierarchical structure have been treated in a number of papers (e.g. *Taylor* (1976), *Norberg* (1986), *Bühlmann/Jewell* (1987)). Hereafter we will strongly rely on the presentation in *Bühlmann/Jewell* (1987). The main result, i.e. the recursive procedure for calculating the credibility estimator, is already well known. But we feel that the following derivation gives a good intuitive insight. Moreover the precisions of the estimators will be given.
For didactical reasons we will consider a model of order 3. The generalization to a hierarchical model of higher order is straightforward.

6.1 Model

The structure of the model is visualized by the following figure:



We will use the notation

e.g. $\Phi(\Psi_\ell)$:= set of all Φ-variables deriving from Ψ_ℓ

$\mathscr{D}(\Phi_k)$:= set of all data (X-variables) deriving from Φ_k

The probability structure is obtained by drawing the variables from "top down" which generates the probability distribution over the whole tree.

Top level: There we have one degenerate random variable, namely the constant μ, which is the overall mean.

Level 3: The random variables Ψ_ℓ $\ell = 1,2,\dots,L$ are i.i.d. with probability distribution $r_3(\psi)$

Level 2: All random variables $\Phi_k \in \Phi(\Psi_\ell)$ are conditionally i.i.d. with probability distribution $r_2(\varphi \mid \Psi_\ell)$

Level 1: All random variables $\Theta_i \in \Theta(\Phi_k)$ are conditionally i.i.d. with probability distribution $r_1(\theta \mid \Phi_k)$

Level 0: Given Θ_i, the random variables X_{ij} $j = 1,2,\dots n_i$ are independent with

$$E[X_{ij} \mid \Theta_i] = \mu(\Theta_i)$$

$$\mathrm{Var}[X_{ij} \mid \Theta_i] = \frac{\sigma^2(\Theta_i)}{P_{ij}}$$

where P_{ij} are known volume measures.

Our aim is to estimate the pure risk premium $\mu(\Theta_i)$ of risk i. We want to find

a) the (inhomogeneous) credibility estimator

$$\widehat{\mu}(\Theta_i) = \text{pro}\left(\mu(\Theta_i) \mid \boldsymbol{L}(\mathscr{D}, 1)\right)$$

b) the homogeneous credibility estimator

$$\widetilde{\mu}(\Theta_i) = \text{pro}\left(\mu(\Theta_i) \mid \boldsymbol{L}_\mu(\mathscr{D})\right)$$

Furthermore we are interested in the precisions of the estimators $\widehat{\mu}(\Theta_i)$ and $\widetilde{\mu}(\Theta_i)$.

6.2 *Relevant quantities and notations*

We will need the following quantities and use the following notations.

Conditional expectations:

$$\mu(\Theta_i) := E[X_{ij} \mid \Theta_i]$$
$$\mu(\Phi_k) := E[\mu(\Theta_i) \mid \Phi_k] \qquad \text{where } \Theta_i \in \Theta(\Phi_k)$$
$$\mu(\Psi_\ell) := E[\mu(\Phi_k) \mid \Psi_\ell] \qquad \text{where } \Phi_k \in \Phi(\Psi_\ell)$$

Structural parameters:

overall mean: $\mu := E[X_{ij}]$
precisions of two neighbouring conditional expectations with respect to each other

level 1 $\quad \varrho := E\left[(\mu(\Theta_i) - \mu(\Phi_k))^2\right]^{-1} \qquad$ where $\Theta_i \in \Theta(\Phi_k)$

level 2 $\quad \sigma := E\left[(\mu(\Phi_k) - \mu(\Psi_\ell))^2\right]^{-1} \qquad$ where $\Phi_k \in \Phi(\Psi_\ell)$

level 3 $\quad \tau := E\left[(\mu(\Psi_\ell) - \mu)^2\right]^{-1}$

Remark: In literature it is more customary to use the inverse of ϱ, σ, τ as structural parameters. Note that e.g. $\sigma^{-1} = E\left[\text{Var}[\mu(\Phi_k) \mid \Psi_\ell]\right]$. In the context of our presentation however, it is more convenient to use the above parametrization.

Statistics:

$$\tilde{\mu}^*(\Theta_i) := \text{pro}\left(\mu(\Theta_i) \mid \boldsymbol{L}_\mu(\mathscr{D}(\Theta_i))\right)$$
$$\tilde{\mu}^*(\Phi_k) := \text{pro}\left(\mu(\Phi_k) \mid \boldsymbol{L}_\mu(\mathscr{D}(\Phi_k))\right)$$
$$\tilde{\mu}^*(\Psi_\ell) := \text{pro}\left(\mu(\Psi_\ell) \mid \boldsymbol{L}_\mu(\mathscr{D}(\Psi_\ell))\right)$$
$$\tilde{\mu} := \text{pro}\left(\mu \mid \boldsymbol{L}_\mu(\mathscr{D})\right)$$

Note that e.g. $\tilde{\mu}^*(\Phi_k)$ is the homogeneous credibility estimator of $\mu(\Phi_k)$ based on the observations X_{ij} deriving from Φ_k.

Precisions of $\tilde{\mu}^*(.)$:

$$\varrho_i^* := E\left[(\tilde{\mu}^*(\Theta_i) - \mu(\Theta_i))^2\right]^{-1}$$
$$\sigma_k^* := E\left[(\tilde{\mu}^*(\Phi_k) - \mu(\Phi_k))^2\right]^{-1}$$
$$\tau_\ell^* := E\left[(\tilde{\mu}^*(\Psi_\ell) - \mu(\Psi_\ell))^2\right]^{-1}$$
$$\xi := E\left[(\tilde{\mu}^* - \mu)^2\right]^{-1}$$

The notations and relevant quantities can be summarized by the following scheme:

level	4	3	2	1	0
variables	μ	Ψ	Φ	Θ	X
index-variables		ℓ	k	i	ij
statistics	$\widetilde{\mu}$	$\widetilde{\mu}^*(\Psi_\ell)$	$\widetilde{\mu}^*(\Phi_k)$	$\widetilde{\mu}^*(\Theta_i)$	
precisions of $\mu^*(.)$	ξ	τ_ℓ^*	σ_k^*	ϱ_i^*	

6.3 *Credibility Estimators*

The next theorem shows that the $\widetilde{\mu}^*(.)$ can be calculated recursively from bottom up by taking at each level a weighted mean of the $\widetilde{\mu}^*(.)$ of the next lower level, the weights being the precisions with respect to the quantity to be estimated. For instance $\widetilde{\mu}^*(\Psi_\ell)$ is a weighted mean of $\{\widetilde{\mu}^*(\Phi_k) : \Phi_k \in \Phi(\Psi_\ell)\}$ and the weights are the precisions of the $\widetilde{\mu}^*(\Phi_k)$ with respect to $\mu(\Psi_\ell)$. In view of the hierarchical structure this result is intuitively very plausible.

Theorem 6

$$\widetilde{\mu}^*(\Theta_i) = (P_i)^{-1} \sum_j P_{ij} X_{ij} \qquad \text{where } P_i = \sum_j P_{ij} \tag{39.1}$$

$$\varrho_i^* = P_i v^{-1} \qquad \text{where } v = E[\sigma^2(\Theta_i)] \tag{39.2}$$

$$\widetilde{\mu}^*(\Phi_k) = \left(\sum \varrho_i'\right)^{-1} \sum \varrho_i' \widetilde{\mu}^*(\Theta_i) \qquad \text{where } \varrho_i' = \varrho \varrho_i^* (\varrho + \varrho_i^*)^{-1} \tag{40.1}$$

$$\sigma_k^* = \sum \varrho_i' \tag{40.2}$$

The sum in (40.1) and (40.2) is taken over $\{i : \Theta_i \in \Theta(\Phi_k)\}$

$$\widetilde{\mu}^*(\Psi_\ell) = \left(\sum \sigma_k'\right)^{-1} \sum \sigma_k' \widetilde{\mu}^*(\Phi_k) \qquad \text{where } \sigma_k' = \sigma \sigma_k (\sigma + \sigma_k)^{-1} \tag{41.1}$$

$$\tau_\ell^* = \sum \sigma_k \tag{41.2}$$

The sum in (41.1) and (41.2) is taken over $\{k : \Phi_k \in \Phi(\Psi_\ell)\}$

$$\widetilde{\mu} = \left(\sum \tau'_\ell\right)^{-1} \sum \tau'_\ell \widetilde{\mu}^*(\Psi_\ell) \qquad \text{where } \tau'_\ell = \tau\tau^*_\ell(\tau + \tau^*_\ell)^{-1} \tag{42.1}$$

$$\xi = \sum \tau'_\ell \tag{42.2}$$

Proof: Consider any two neighbouring points in the tree, e.g. Ψ_ℓ and $\Phi_k \in \Phi(\Psi_\ell)$. From Lemma 3 it follows that the homogeneous credibility estimator of $\mu(\Psi_\ell)$ based only on the data $\mathscr{D}(\Phi_k)$ is equal to $\widetilde{\mu}^*(\Phi_k)$. Theorem 2 then yields, that $\widetilde{\mu}^*(\Psi_\ell)$ is a weighted mean of $\{\widetilde{\mu}^*(\Phi_k);\ \Phi_k \in \Phi(\Psi_\ell)\}$ with the precisions of the $\widetilde{\mu}^*(\Phi_k)$ with respect to $\mu(\Psi_\ell)$ as weights. Since

$$E\left[(\mu(\Psi_\ell) - \widetilde{\mu}^*(\Phi_k))^2\right] = E\left[\{\mu(\Psi_\ell) - \mu(\Phi_k) + \mu(\Phi_k) - \widetilde{\mu}^*(\Phi_k)\}^2\right]$$

and

$$E\left[E\left[\{\mu(\Psi_\ell) - \mu(\Phi_k)\}\{\mu(\Phi_k) - \widetilde{\mu}^*(\Phi_k)\} \mid \Psi_\ell, \Phi_k\right]\right] = 0$$

we obtain

$$E\left[(\mu(\Psi_\ell) - \widetilde{\mu}^*(\Phi_k))^2\right] = (\sigma^{-1} + \sigma_k^{-1})^{-1} = \sigma'_k,$$

which completes the proof of (41.1) and (41.2).
Of course the other formulae in theorem 8 can be proved analogously.

To derive the credibility estimators $\widehat{\mu}(.) = \mathrm{pro}(\mu(.)|\boldsymbol{L}(\mathscr{D}, 1))$ we first introduce the auxiliary random variables

$$\mu'(\Phi_k) := \mathrm{pro}(\mu(\Phi_k) \mid \boldsymbol{L}(\mathscr{D}, \mu(\Psi_\ell), 1)) \qquad \Phi_k \in \Phi(\Psi_\ell)$$
$$\mu'(\Theta_i) := \mathrm{pro}(\mu(\Theta_i) \mid \boldsymbol{L}(\mathscr{D}, \mu(\Phi_k), 1)) \qquad \Theta_i \in \Theta(\Phi_k)$$

Remark: If the conditional expectations at the next higher level were known, then the random variables $\mu'(.)$ would be the credibility estimators.

Consider again two neighbouring points in the tree, e.g. Φ_k and $\Theta_i \in \Theta(\Phi_k)$. Let $X_{ij} \in \mathscr{D}(\Theta_i)$, $X_{mn} \notin \mathscr{D}(\Theta_i)$. By conditioning on the Φ-variables we get

$$(\mu(\Theta_i) - \mu(\Phi_k)) \perp X_{mn}$$
$$(\mu(\Theta_i) - X_{ij}) \perp X_{mn}$$
$$(\mu(\Theta_i) - \mu(\Phi_k)) \perp \mu(\Phi_k).$$

Thus it follows from Lemma 1 that $\mu'(\Theta_i)$ depends only on the data belonging to $\mathscr{D}(\Theta_i)$. Analogously $\mu'(\Phi_k)$ resp. $\widehat{\mu}(\Psi_\ell)$ depend only on the data belonging to $\mathscr{D}(\Phi_k)$ resp. to $\mathscr{D}(\Psi_\ell)$. Hence from theorem 3 we get

$$\widehat{\mu}(\Psi_\ell) = (\tau_\ell^* + \tau)^{-1}(\tau_\ell^* \widetilde{\mu}^*(\Psi_\ell) + \tau\mu) \tag{43.1}$$

$$\mu'(\Phi_k) = (\sigma_k^* + \sigma)^{-1}(\sigma_k^* \widetilde{\mu}^*(\Phi_k) + \sigma\mu(\Psi_\ell)) \qquad \Phi_k \in \Phi(\Psi_\ell) \tag{43.2}$$

$$\mu'(\Theta_i) = (\varrho_i^* + \varrho)^{-1}(\varrho_i^* \widetilde{\mu}^*(\Theta_i) + \varrho\mu(\Phi_k)) \qquad \Theta_i \in \Theta(\Phi_k) \tag{43.3}$$

Remark: Denote by α_i, β_k and γ_ℓ the credibility weights given to $\widetilde{\mu}^*(.)$ in the formula (43), e.g. $\beta_k = \sigma_k^*(\sigma_k^* + \sigma)^{-1}$. Note that the weights ϱ_i', σ_k', τ_ℓ' in (40.1), (41.1) and (42.1) are up to a constant factor the same as these credibility weights, e.g. $\sigma_k' = \sigma\beta_k$. Hence they may be replaced by the credibility weights on that level (e.g. β_k instead of σ_k'), which are the formulae in *Bühlmann & Jewell* (1987).

Since $\widehat{\mu}(.) = \text{pro}(\mu'(.) \mid \boldsymbol{L}(\mathscr{D}, 1))$ the following recursion for computing the credibility estimators from "top down" results:

Theorem 7

$$\widehat{\mu}(\Psi_\ell) = (\tau_\ell^* + \tau)^{-1}(\tau_\ell^* \widetilde{\mu}^*(\Psi_\ell) + \tau\mu) \tag{44.1}$$

$$\widehat{\mu}(\Phi_k) = (\sigma_k^* + \sigma)^{-1}(\sigma_k^* \widetilde{\mu}^*(\Phi_k) + \sigma\widehat{\mu}(\Psi_\ell)) \qquad \Phi_k \in \Phi(\Psi_\ell) \tag{44.2}$$

$$\widehat{\mu}(\Theta_i) = (\varrho_i^* + \varrho)^{-1}(\varrho_i^* \widetilde{\mu}^*(\Theta_i) + \varrho\widehat{\mu}(\Phi_k)) \qquad \Theta_i \in \Theta(\Phi_k) \tag{44.3}$$

At this point it is worthwhile to summarize the procedure:

- In a first step the statistics $\widetilde{\mu}^*(.)$ and the corresponding precisions are calculated from "bottom up" according to theorem 6.
- In a second step the credibility estimators at the different levels are obtained by proceeding from "top down" according to theorem 7.

We will denote the precisions of the credibility estimators $\widehat{\mu}(.)$ at the different levels by τ_ℓ, σ_k, ϱ_i (e.g. $\sigma_k := E\left[(\widehat{\mu}(\Phi_k) - \mu(\Phi_k))^2\right]^{-1}$). The next theorem shows, that they can also be calculated recursively from "top down".

Theorem 8

$$\tau_\ell = (\tau_\ell^* + \tau) \tag{45.1}$$

$$\sigma_k = (\sigma_k^* + \sigma)\left(1 + \frac{\sigma}{\sigma_k^* + \sigma}\,\frac{\sigma}{\tau_\ell}\right)^{-1} \qquad \Phi_k \in \Phi(\Psi_\ell) \tag{45.2}$$

$$\varrho_i = (\varrho_i^* + \varrho)\left(1 + \frac{\varrho}{\varrho_i^* + \varrho}\,\frac{\varrho}{\sigma_k}\right)^{-1} \qquad \Theta_i \in \Theta(\Phi_k) \tag{45.3}$$

Proof: (45.1) is a direct consequence of theorem 2. The proof of (45.2) is analogous to the proof of (35). First note that

$$\mu(\Phi_k) - \widehat{\mu}(\Phi_k) = (\mu(\Phi_k) - \mu'(\Phi_k)) + \sigma(\sigma_k^* + \sigma)^{-1}(\mu(\Psi_\ell) - \widehat{\mu}(\Psi_\ell))$$

Since

$$(\mu(\Phi_k) - \mu'(\Phi_k)) \perp \boldsymbol{L}(\mathscr{D}, \mu(\Psi_\ell), 1)$$

it follows that

$$\langle \mu(\Phi_k) - \mu'(\Phi_k), \mu(\Psi_\ell) - \widehat{\mu}(\Psi_\ell) \rangle = 0$$

Hence we get from theorem 2

$$E\left[(\widehat{\mu}(\Phi_k) - \mu(\Phi_k))^2\right] = (\sigma_k^* + \sigma)^{-1} + \sigma^2(\sigma_k^* + \sigma)^{-2}\tau_\ell^{-1}$$

which is identical to (45.2).
Of course (45.3) is proved in exactly the same way.

As to the homogeneous credibility estimators $\widetilde{\mu}(.) = \text{pro}(\mu(.) \mid \boldsymbol{L}_\mu(\mathscr{D}))$, we get from (3), (9) and theorem 7 the recursion

$$\widetilde{\mu}(\Psi_\ell) = (\tau_\ell^* + \tau)^{-1}(\tau_\ell^* \widetilde{\mu}^*(\Psi_\ell) + \tau\widetilde{\mu}) \tag{46.1}$$

$$\widetilde{\mu}(\Phi_k) = (\sigma_k + \sigma)^{-1}(\sigma_k^* \widetilde{\mu}^*(\Phi_k) + \sigma\widetilde{\mu}(\Psi_\ell)) \qquad \Phi_k \in \Phi(\Psi_\ell) \tag{46.2}$$

$$\widetilde{\mu}(\Theta_i) = (\varrho_i^* + \varrho)^{-1}(\varrho_i^* \widetilde{\mu}^*(\Theta_i) + \varrho\widetilde{\mu}(\Phi_k)) \qquad \Theta_i \in \Theta(\Phi_k) \tag{46.3}$$

Finally using the same arguments as used in the proof of theorem 8 we obtain the formulae for the precisions

$$\widetilde{\tau}_\ell = (\tau_\ell^* + \tau)\left(1 + \frac{\tau}{\tau_\ell^* + \tau}\,\frac{\tau}{\xi}\right)^{-1} \tag{47.1}$$

$$\widetilde{\sigma}_k = (\sigma_k^* + \sigma)\left(1 + \frac{\sigma}{\sigma_k^* + \sigma}\,\frac{\sigma}{\widetilde{\tau}_\ell}\right)^{-1} \qquad \Phi_k \in \Phi(\Psi_\ell) \tag{47.2}$$

$$\widetilde{\varrho}_i = (\varrho_i^* + \varrho)\left(1 + \frac{\varrho}{\varrho_i^* + \varrho}\,\frac{\varrho}{\widetilde{\sigma}_k}\right)^{-1} \qquad \Theta_i \in \Theta(\Phi_k) \tag{47.3}$$

References

Bühlmann, H./Straub, E. (1970): Glaubwürdigkeit für Schadensätze. Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker *70*, 111 – 133.

Bühlmann, H./Jewell, W. S. (1987): Hierarchical Credibility Revisited. Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker *87*, 35 – 54.

Campbell, M. (1986): An integrated system for estimating the risk premium of individual car models in motor insurance. ASTIN Bulletin *16*, 165 – 183.

De Vylder, F. (1976 a): Geometrical credibility. Scandinavian Actuarial Journal, 121 – 149.

De Vylder, F. (1976 b): Optimal Semilinear Credibility. Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker *76*, 27 – 39.

Gerber, H. U./Jones, D. A. (1975): Credibility formulas of the updating type. Transactions of the Society of Actuaries *27*, 31 – 46. Also appeared in Credibility: Theory and applications (ed. P. M. Kahn), 89 – 105. Academic Press, New York.

Kremer, E. (1982): Credibility Theory for Some Evolutionary Models. Scandinavian Actuarial Journal, 129 – 142.

Norberg, R. (1986): Hierarchical Credibility: Analysis of a Random Effect Linear Model with Nested Classification. Scandinavian Actuarial Journal, 204 – 222.

Sundt, B. (1981): Recursive Credibility Estimation. Scandinavian Actuarial Journal, 3 – 21.

Sundt, B. (1983): Finite Credibility Formulae in Evolutionary Models. Scandinavian Actuarial Journal, 106 – 116.

Sundt, B. (1987): Credibility and old estimates. Insurance, Mathematics and Economics *7*, 189 – 194.

Taylor, G. C. (1979): Credibility analysis of a general hierarchical model. Scandinavian Actuarial Journal, 1 – 12.

Alois Gisler
"Winterthur"
Swiss Insurance Company
P.O. Box 357
8401 Winterthur

Summary

It will be shown that under some basic assumptions the credibility estimator is a weighted mean, the weights being the precisions with respect to the quantity to be estimated. Based on this principle and the fact, that credibility estimators are projections on linear subspaces of the Hilbert space of square integrable random variables, we will derive in an elegant way the credibility estimators in a fundamental model as well as in a class of evolutionary models, in the semilinear case and in the hierarchical model. Moreover the precisions of the estimators will be obtained very easily.

Zusammenfassung

Es wird gezeigt, dass unter gewissen Grundvoraussetzungen der Credibility-Schätzer ein gewichtetes Mittel ist, wobei die Gewichte nichts anderes sind als die Präzisionen der Komponenten in Bezug auf die zu schätzende Grösse. Basierend auf diesem intuitiv leicht zugänglichen Grundresultat und unter Zuhilfenahme der Tatsache, dass Credibility-Schätzer Projektionen auf lineare Unterräume des Hilbertraums der quadratintegrierbaren Zufallsgrössen sind, werden auf elegante und konsistente Weise die Credibility-Schätzer hergeleitet in einem Basis-Modell wie auch in einer Klasse von evolutionären Modellen, im semilinearen Fall und im hierarchischen Modell. Zudem erhält man praktisch als Nebenprodukt der Herleitung die Präzisionen der zugehörigen Schätzer.

Résumé

L'article montre que sous certaines hypothèses de base l'estimateur de crédibilité est une moyenne pondérée et les poids les précisions des quantités à estimer. Sur cette base et vu le fait que les estimateurs de crédibilité sont des projections sur des sous-espaces linéaires de l'espace de Hilbert des variables aléatoires intégrables au carré, l'article dérive les estimateurs dans le cas de modèles de base aussi bien que de certains modèles évolutifs et dans le cas des modèles semilinéaire et hiérarchique. De plus il est possible d'obtenir facilement les précisions des estimateurs.