INTRODUCTORY REPORT

EXPERIENCE RATING AND CREDIBILITY

HANS BÜHLMANN
Zürich

1. *The Classical Ratemaking Problem*

Classical statistics deals with the following standard problem of estimation:

Given: random variables $X_1, X_2 \ldots X_n$ independent, identically distributed, and
observations $x_1, x_2 \ldots x_n$,

Estimate: parameter (or function thereof) of the distribution function common to all X_i.

It is not surprising that the "classical actuary" has mostly been involved in solving the actuarial equivalent of this problem in insurance, namely

Given: risks $R_1, R_2 \ldots R_n$ no contagion, homogeneous group,
Find: the proper (common) rate for all risks in the given class.

There have, of course, always been actuaries who have questioned the assumptions of independence (no contagion) and/or identical distribution (homogeneity). As long as ratemaking is considered equivalent to the determination of the mean, there seem to be no additional difficulties if the hypothesis of independence is dropped. But is there a way to drop the condition of homogeneity (identical distribution)?

2. *Pragmatic Solution*

Insurance people are practically minded, and so they have in many circumstances come out with pragmatic solutions, even if the theoreticians were not able to provide them with full theoretical justification for doing so. Within the classical setup such circumstances were happening if

a) you had to rate a risk with no or rather dubious observational data on the class to which it belonged,

b) you had to rate a risk which could not be grouped into a homogeneous class except for one so small that statistical inference drawn from it had little significance.

The pragmatic devices to handle such cases are known to all of us, namely

— participation in profits
— no claim bonus
— premium scales (bonus-malus)
— sliding scale premiums and/or commissions in reinsurance.

All these devices are characterized by the fact that the rate originally fixed is gradually altered during the contractual period of risk coverage. This is called experience rating. It is done, yes, but what about its justification? What does the actuary have to say about it?

3. *Report on Papers Submitted*

Seven papers have been sent in on the subject "Credibility and Experience Rating", and it is interesting to note how actuarial techniques vary in treating this subject.

Let me first speak on the contribution by Harald Bohman, "Experience rating when the company aims at increasing the volume of its business". This paper shows that the scope of experience rating may even be extended beyond "proper-rate making". Aiming at maximum volume of business is the alternative treated by Harald Bohman. His main result, namely that the maximum volume is generally not achieved at the lowest possible premium rates, is certainly of great practical value.

The other six papers treat the problem of experience rating with the aim of finding the correct rate. As all the authors use some form of sequential statistical techniques to tackle the problem, I shall restrict myself to the discussion of such sequential techniques as a means of coping with experience rating.

It is noteworthy that the classical sequential methods, such as those developed by Abraham Wald, do not permit to go beyond the testing of a whole tariff class. Having the two cases above in mind, where practical circumstances force some form of experience rating upon us, I would say that Wald's Sequential Probability Ratio Test can be of help in case a), but unfortunately fails to solve the problem in case b). Lars Benckert brings this out even in the title of his paper "Testing of a Tariff", and I judge it very important to stress the word *tariff* as against *individual risk*. Indeed, Lars Benckert shows us how to control the adequacy of a tariff rate for a whole class using the logarithmic normal distribution for individual claim amounts.

To get a classification of risks inside a tariff class, it is necessary to treat the risk parameters no longer as constants, but as random variables. Only if you take this step of generalization, then you can "hunt accident prones", as Paul Johansen does. I am very glad that he has contributed this paper to the subject under discussion, since it shows very clearly the advantage which can be gained by assuming the risk parameter to be a random variable. This advantage, by the way, is not bought at the expense of an increasingly complicated formula, a fact that for practical purposes will certainly be appreciated as well.

Two papers, namely, "Experience Rating in Subsets of Risks", by Fritz Bichsel, and "Une note sur des systèmes de tarification basés sur des modèles du type Poisson"[1]), by Ove Lundberg, center around the application of the Bayes rule for experience rating purposes. Since Neo-Bayesian techniques seem to really be quite well adapted to experience rating problems, let me expound in greater detail on this approach.

We have said earlier that we would like to have a method which no longer requires the grouping of individual risks into homogeneous classes. What do we actually mean by *homogeneity?* This is best illustrated by the following array of random variables X_{ij} = claim produced by risk i in the accounting period (year) j.

$$
\begin{array}{l}
X_{11}, X_{12}, \ldots X_{1n} \\
\quad\cdot \\
\quad\cdot \\
\quad\cdot \\
X_{m1}, X_{m2}, \ldots X_{mn}.
\end{array}
$$

[1]) Published in Astin Bulletin vol. IV Part I.

The above array represents all "claims random variables" under observation from a class of m risks during n years.

i) This class would be called *homogeneous in the mass* of risks, if $X_{1j}, X_{2j}, X_{3j}, \ldots X_{mj}$ are identically distributed for all fixed j.

ii) Each individual risk would be called *homogeneous in time* if $X_{i1}, X_{i2}, \ldots X_{in}$ are identically distributed for fixed i.

Dropping the requirement of homogeneity means either dropping i) or ii) or both at the same time. All actuarial work so far has, at least according to my knowledge, been done by dropping homogeneity in mass but by still requiring homogeneity in time of the individual risk *). Fritz Bichsel now also tackles the experience rating problem under changes of individual risks in time. He has thus added an additional dimension to the experience rating problem, a dimension which I consider most important from both theoretical and practical viewpoints. If I may make a suggestion it is this: the changes in time considered by Bichsel are random changes; it would be most valuable to treat the problem also under the aspect of trend type changes.

Ove Lundberg seems to have been the first actuary to realize the importance of Bayes procedures for experience rating. The basic principle is already mentioned in his 1940 book.

Now that we other actuaries are understanding more and more the importance of his early work, we particularly welcome his new contribution "Une note sur des systèmes de tarification basés sur des modèles du type Poisson composé". His main result consists in the proof of the consistency of the Bayes estimate derived from a Poisson process. In contrast to the Bayes estimate, Ove Lundberg shows that any estimate based on the time of absence is *not consistent*—a very powerful theoretical undermining of the superiority of the bonus—malus system against the simple bonus system.

---

*) In the discussion Carl Philipson has pointed at some earlier researches of Ammeter ("A Generalization of the Collective Theory of Risk in Regard to Fluctuating Basic Probabilities", Skand. Akt. Tidskrift 1948) and Philipson ("Einige Bemerkungen zur Bonusfrage in der Kraftversicherung" BDGVM 1963; Eine Bemerkung zu Bichsels Herleitung der bedingten zukünftigen Schadenhäufigkeit einer Polya-Verteilung" MVSVM 1964).

But now let's turn to the second term in the heading of the subject discussed: credibility.

It is quite remarkable that our American colleagues have been well advised in solving most of their Experience Rating problems in applying the by now famous *credibility formula*:

$$p_n = (1 - \alpha)\, p_{n-1} + \alpha . r_{n-1}$$

p_k = rate for period k

r_k = loss ratio for period k

α is called *credibility* and assumed to be a function of the volume V, mostly

$$\alpha(V) = \frac{V}{V+k} \text{ or } \alpha(V) = \begin{cases} 1 & \text{if } V \geqslant V_0 \\ \sqrt{\dfrac{V}{V_0}} & \text{if } V < V_0 \end{cases}$$

V_0 is called the *volume of full credibility*.

Hetereogeneity in mass, in time, rate changes for big groups and small groups as well, even some refund formulas are treated by the credibility method on the United States market—and this is the remarkable fact—the credibility formula has been found to do an excellent job. Under such circumstances it does not come as a surprise that many actuaries have tried to prove the credibility formula starting from more general principles. The first to do so was Arthur Bailey, one of the most outstanding American actuaries in the mid-century to whose work Bruno de Finetti has drawn our attention at the Trieste colloquium. Since then I am sure that many of us have made our own personal attempts. Edouard Franckx in his paper "La tarification et son adaptation expérimentale dans le cadre d'une classe de tarif" arrives at the credibility formula by starting from the principle of least squares. It seems to me that in doing so he is the first author to prove the credibility relation independent of the distribution function which governs the individual risks (but still dependent on the prior distribution of the risk parameter or parameters). This would justify our American colleagues' using credibility procedures beyond the assessment of claims frequencies—a point that since Arthur Bailey has worried many of our very best colleagues across the Atlantic.

I regret that there has been only one paper sent in which specifically deals with *credibility techniques*. Marcel Derron with his contribution "Credibility Betterment Through Exclusion of the Largest Claims"[1]) takes up some very interesting thoughts originated by Hans Ammeter in Trieste. Derron computes the credibility improvement for a Pareto-type distribution if the largest claim is excluded. The result may indicate a new way of approach for many other estimation problems which we encounter in our work. I mean that Marcel Derron has very clearly shown to us the importance of truncating or curtailing our basic data to get more reliable information out of them. The importance of his work is underlined by many theoretical statisticians who are presently working on robust estimation methods, and who propose also the exclusion of extreme values to make the information more reliable.

With this I conclude my report on the papers submitted. Permit me though to add two rather special technical remarks as my personal contribution to the discussion.

4. *Equilibrium in an experience rated portfolio*

Consider a set Θ of risk parameters ϑ. Each individual risk (characterized by a value of ϑ) can be observed on a number of random variables $X_i, i = 1 \ldots n$ (To fix the ideas think of claim frequencies or claim sums produced by an individual risk in the year i).

As is usually done I assume the X_i's to be independent and identically distributed (homogeneity *in time*, but *not* in mass). Call the common distribution function $F_\vartheta(x)$ with mean $\mu(\vartheta)$ and variance $\sigma^2(\vartheta)$. Experience rating is then understood as a sequence of estimates for $\mu(\vartheta)$ based on the observations of $X_1, X_2, \ldots X_n$.

Except for one trial on Delaporte's side where the a posteriori *median* is investigated I believe that in the actuarial literature the estimator function for $\mu(\vartheta)$ is always chosen to be equal to

$E[\mu(\vartheta)/X_1, X_2, \ldots X_n]$ = a posteriori *mean*.

Justification for this choice is usually found in the fact that among all functions f depending on the observations only (and integrable of course) the expected square deviation.

$$\int \{\mu(\vartheta) - f(x_1, \ldots x_n)\}^2 \, dF_\vartheta(x_1) dF_\vartheta(x_2) \ldots dF_\vartheta(x_n) \, dS(\vartheta)$$

[1]) Published in Astin Bulletin Vol. IV Part 1.

where $S(\vartheta)$ denotes the "a priori distribution" of the risk parameter (structural function of portfolio), is smallest for the choice

$$f = E[\mu(\vartheta)/X_1, \ldots X_n].$$

I feel that the least expected square deviation is no sufficient justification for the choice of the a posteriori mean as estimator function. Permit me therefore to set forth another justification which Paul Thyrion has indicated to me in a recent letter.

The basic idea can be described by the postulate of equilibrium in all those subclasses of a portfolio which are characterized by experience only. In other words it is postulated that each class of risks with *equal observed risk performance* should pay its own way. Let us try to put this into mathematical language. Let

Θ = set of all possible parameters ϑ

X = set of all possible observed risk performances $(X_1, X_2, \ldots X_n)$

X' = subset of X, $\subset(X')$ = cylinder in $\Theta \times X$ with base $X' \subset X$

P = probability on $\Theta \times X$

Equilibrium for any subset X' means

$$\int \mu(\vartheta) dP = \int f(x_1, x_2, \ldots x_n)\, dP \text{ for all cylinders } \subset(X')$$

The above relation is exactly Kolmogoroff's definition of the conditional expectation

$$E[\mu(\vartheta)/x_1, \ldots x_n]$$

5. *A Distribution Free Credibility Formula*

It is worthwhile to note here that the credibility formula used by our American colleagues is nothing but a linearization of the above estimator function $E[\mu(\vartheta)/X_1, \ldots X_n]$.

I am giving here the least expected square deviation approximation to this a posteriori mean. (Observe that $E[.]$ means expectation with respect to the probability P on the product space $\Theta \times X$, $E_\vartheta[.]$ means expectation on X given the individual risk ϑ).

The best linear approximation to

$E[\mu(\vartheta)/x_1, \ldots x_n]$ is found by solving the following problem

Find: a, b such that

$$a + b\overline{X} \text{ where } \overline{X} = \frac{X_1 + X_2, \ldots X_n}{n}$$

approximates $E[\mu(\vartheta)/x_1, \ldots x_n]$ best

i.e. $E\,\{E[\mu(\vartheta)/x_1,\, x_2, \ldots x_n] - (a + b\overline{X})\}^2 = \text{minimum}$

Lemma: If $E(a_0 + b_0\overline{X} - \mu(\vartheta)^2) \leqslant E(a + b\overline{X} - \mu(\vartheta)^2)$ for arbitrary a and b

then $a_0 + b_0\overline{X}$ is also the best linear approximation to $E[\mu(\vartheta)/x_1, \ldots x_n]$

Proof: $E\,(a_0 + b_0\overline{X} - \mu(\vartheta))^2 = E\,\{a_0 + b_0\overline{X} - E[\mu(\vartheta)/x_1, \ldots x_n]\}^2 + E\,\{E[\mu(\vartheta)/x_1, \ldots x_n] - \mu(\vartheta)\}^2$

Since the second term on the right hand side does not depend on a_0 and b_0 it is clear that the left hand side and the right hand side are mimimized by the same a_0 and b_0 q.e.d.

Reformulating the original problem we hence want to find a and b such that

$$E[a + b\overline{X} - \mu(\vartheta)]^2 = \text{minimum}$$

We find that the above left hand side can be written

$$E[b(\overline{X} - \mu(\vartheta))]^2 + E[a - (1 - b)\mu(\vartheta)]^2$$

which is minimum if i) $a = (1 - b)E[\mu(\vartheta)]$

and ii) $b^2E(\overline{X} - \mu(\vartheta))^2 + (1 - b)^2 \operatorname{Var}[\mu(\vartheta)]$ minimum

$$\text{which leads to } b = \frac{\operatorname{Var}[\mu(\vartheta)]}{\operatorname{Var}[\mu(\vartheta)] + E(\overline{X} - \mu(\vartheta))^2}$$

Assuming now *independence* and identical distribution of the X_i we find

$$E(\overline{X}\text{-}\mu(\vartheta))^2 = \frac{1}{n}\,E(X_1 - \mu(\vartheta))^2 = \frac{1}{n}\,E[\sigma^2(\vartheta)]$$

Hence the credibility relation

$$(1 - b)\cdot E[\mu(\vartheta)] + b\cdot\overline{X}$$

where

$$b = \frac{n}{n + k}$$

$$k = \frac{E[\sigma^2(\vartheta)]}{\mathrm{Var}[\mu(\vartheta)]}$$

Remarks:

1) This relation makes no assumption as to the type of distribution function governing the individual risk or the a priori (structural) distribution function of the parameters.
2) The hypothesis of independence and identical distribution of the observational random variables of the same individual risk could easily be dropped. It amounts to replacing the relation

   $E(X - \mu(\vartheta))^2 = \frac{1}{n} E[\sigma^2(\vartheta)]$ by some other function of n.
3) Since the above relation is generally true it is of great interest to estimate

   $E[\sigma^2(\vartheta)]$ and Var $[\mu(\vartheta)]$

   directly from certain "a priori observations". This problem has not yet been attacked.