AN UNBASED APPROACH TO CREDIBILITY

by

Ragnar Norberg

Abstract

This is a rejoinder to Gerber's recent note entitled "An unbayesed approach to credibility". It is argued here that the title above would have been more appropriate. Old established principles of statistical decisions are advocated, viz. the necessity of (i) building adequate stochastic models and (ii) examining the properties of any proposed statistical method as in terms of the model assumptions and the performance criterion.

Key words: Credibility, models, methods, principles of statistical decisions.

1. Should automobile insurance premiums depend on the turnover of Appenzeller cheese?

That is the question one is faced with upon reading Gerber (1982). In Sections 4-8 of his note he deals with the following

MODEL:

> Consider m populations. From each population $i = 1,\ldots,m$, we have observations X_{ij}, $j = 1,\ldots,n$, which are i.i.d., with $EX_{ij} = \mu_i$ and $\mathrm{Var}X_{ij} = \sigma_i^2$. All X_{ij} are stochastically independent and the parameters (μ_i,σ_i^2) are functionally independent.

This model says precisely that the samples are unrelated in every respect. They are chosen in an independent manner from populations that have nothing in common. Such a model is suitable for instance if the X_{1j}'s are sold amounts of Appenzeller cheese in different years, the X_{2j}'s are measurements of soldiers' heights,..., the X_{ij}'s are claim amounts in different years for an automobile insurance treaty, etc.

In Section 6 of his paper Gerber addresses himself to the following

PROBLEM:

Estimate the mean μ_i . The performance of an estimator P_i is measured by

$$E(P_i-\mu_i)^2 \ . \qquad (1)\ [27]$$

(Numbers in square brackets refer to formulas in Gerber's paper.) Actually Gerber phrases his problem as that of predicting a future independent selection $X_{i,n+1}$, the performance of a predictor P_i being measured by $E(P_i-X_{i,n+1})^2$. However, as this expression differs from that in (1) only by σ_i^2 , his problem is really the one stated here. Now this problem forms a basic exercise in statistics. It is, therefore, surprising to see it brought to issue anew in a research journal. What more is there to propose?

Gerber proposes the following

METHOD:

(i) Consider estimators of the form

$$P_i = \zeta\bar{X}_i + (1-\zeta)\bar{X}_{i-} \qquad (2)[11]$$

where

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij} \quad \text{and} \quad \bar{X}_{i-} = \frac{1}{m-1}\sum_{h\neq i}\bar{X}_h \ . \qquad (3)[10]$$

And one must wonder why. Why include sold amounts of Appenzeller cheese in the (estimate of the) automobile insurance premium? Already in the model it is stated that the $\bar{X}_h$, $h \neq i$, are irrelevant for the estimation of μ_i . A reasonable way to proceed would be to seek, on grounds of the model and general principles of statistical inference, arguments for <u>leaving out</u> all irrelevant data. If we are not able to justify the deletion of the $\bar{X}_h$, $h \neq i$, from the estimation of μ_i in the present model, then we are in serious trouble; How can we then in a rational way choose the statistical basis for rating of insurance policies? Which (irrelevant) information is not to be included? And which advice shall we give to the practitioners?

But Gerber wants the irrelevant data to appear in the insurance premium and commands the use of a formula of the homogeneous credibility type (2). He proceeds as follows.

METHOD (continued):

(ii) Determine the value ζ_i of ζ in (2) which minimizes (1). The solution is

$$\zeta_i = 1 - \frac{(m-1)\sigma_i^2}{n(m-1)(\mu_i-\mu_{i-})^2+\sigma_{i-}^2+(m-1)\sigma_i^2} \qquad (4)[28]$$

with

$$\mu_{i-} = \frac{1}{m-1}\sum_{h\neq i}\mu_h \quad \text{and} \quad \sigma_{i-}^2 = \frac{1}{m-1}\sum_{h\neq i}\sigma_h^2 .$$

This does not give an estimator since the solution depends on unknown parameters, including the estimand μ_i . To get around this Gerber proposes:

(iii) Replace the unknown parameter funtions $(\mu_i-\mu_{i-})^2$ and σ_{i-}^2 in (4) by "their natural unbiased estimators" (Gerber's Section 4) defined by [29], which gives

$$\hat{\zeta}_i = 1 - \frac{\hat{\sigma}_i^2}{n(\bar{X}_i-\bar{X}_{i-})^2} \qquad (5)[30]$$

and the estimator

$$\hat{P}_i = \hat{\zeta}_i\bar{X}_i + (1-\hat{\zeta}_i)\bar{X}_{i-} . \qquad (6)$$

Thus one arrives at an estimator with the appearance of a credibility formula. From this and a number of similar exercises Gerber concludes in his Section 9 that, "It appears that the classical Bayesian (!) approach is only one out of several methods that produce credibility formulas."

This quotation gives a key to an understanding of the above arithmetics by evincing the lack of distinction between model and method which is a characteristic of Gerber's paper.

2. Model and method

I would approve to the quoted conclusion if it were made precise as follows: "The (empirical) Bayes model is only one out of an infinity of models within which credibility formulas can be produced. The only thing that is needed is the will to produce them. If in our search for an estimation method we confine ourselves to credibility formulas of the form (2), then we will certainly produce a credibility formula of the form (2), no matter what model assumptions and optimality criterion we have adopted." This is true, but trivial. Nevertheless Gerber undertakes to demonstrate by ample examples the validity of this thesis. He picks some models with a minimum of structure, specifying no relations that could motivate the use of the collateral $\bar{X}_h$, $h \neq i$, then conjures forth formula (2) and produces credibility formulas by use of the mean square machinery. In this respect the "direct approach" in Gerber's Sections 5 and 7 is really direct to the point; There credibility formulas are produced without any model at all. (The vague formulations about exchangeability in time and matrix rows that are assumed to be equivalent under permutations are empty as they are never explicated in mathematical terms; They are of no significance to the results.) That exercise demonstrates that credibility formulas can be obtained from any collection of numbers, no matter what is their nature.

Theoretical work is more than mere calculations. One must separate out those features which have some bearing on the problem, work them into a model to give a surveyable and, as far as possible, true picture of the phenomena, and finally investigate the logical consequences of the model assumptions.

The objective of credibility theory is to found a base for, i.e. motivate and explain the rationale of the use of estimators which are credibility weighted means of current and collateral risk experience. As a first step one must specify a model that comprises all relevant knowledge. In particular it must give precise content to the notion of collateral data. Gerber's model fails to reflect the essential circumstance that automobile insurance risks have something in common that distinguishes them from data on cheeses and soldiers' heights. The mathematical way to establish this similarity between the risks is to regard them as selections from one and the same population. Thus the structure distribution is certainly not superfluous to those who think they can learn something about a risk by looking at other similar risks. Having decided on a model and a performance criterion the statistician must, as a second step, examine the properties of any proposed decision rule as in terms of the criterion. This is not done in Gerber's paper.

With this in mind let us investigate in some more detail Gerber's approach and also the usual approach based on the concept of structure distribution.

3. Properties of Gerber's method in Gerber's model

Which consequences can be drawn from the model in Section 1? Very few, indeed; From a parsimonious set of assumptions one cannot deduce much. However, if one steps outside of the model framework by imposing arbitrary restrictions on the method, as it is done in Gerber's method, one can suddenly deduce a lot, and much of it contradicting common sense as we shall see. Arbitrariness in choice of method will usually unveil itself through annoyances in the answers.

One example of this is the fact that the estimator (6) may have an arbitrarily large bias. In fact, its expectation will typically not exist because in most situations the credibility $\hat{\zeta}_i$ defined in (5) is unbounded (it may assume any negative value) and non-integrable.

Another example, which is fatal, follows immediately: in most situations Gerber's estimator will give the value $+\infty$ to the expected squared error (1), which he wants to minimize.

Yet another example is the lack of invariance. One can apply any set of transformations $X_{hj} \to f_h(X_{hj})$, $j = 1,\ldots,n$, to the irrelevant X_{hj}'s, $h \neq i$, and remain in the same model. But the estimator (6) will not remain invariant; In fact it can assume any value. In particular, by performing a scale transformation $X_{hj} \to cX_{hj}$ and making c large we obtain $\hat{P}_i = \bar{X}_i$ in the limit.

Step (iii) in the method is a heavy piece of arbitrariness. On his way to an estimate of μ_i Gerber is required to estimate μ_i, or rather some functions of μ_i and the other parameters. Why is the unbiasedness principle now suddenly useful? If it had been invoked already at the outset, the irrelevant data could have been deleted at once and a lot of trouble would have been avoided. And one would of course have ended up with "the natural unbiased estimator" $\hat{P}_i = \bar{X}_i$. Ad hoc devices of this kind is what one resorts to when the model is not appropriate so that the conclusions aimed at can not be hit by consequent methodology alone.

Why not look for better estimators than those given by formula (2)? Let us allow for less restrictive estimators of the form

$$P_i = \sum_{h=1}^{m} \zeta_h \bar{X}_h \quad . \qquad (7)$$

Upon inserting (7) into (1), we find after some calculation that the optimal coefficients are

$$\zeta_{ih} = \mu_i \frac{n\mu_h/\sigma_h^2}{n\sum_{k=1}^{m} \mu_k^2/\sigma_k^2 + 1} \quad . \qquad (8)$$

If we replace the parameters μ_h and σ_h^2 occurring in (8) by "their natural unbiased estimators", we arrive at

$$\hat{P}_i = \bar{X}_i \frac{n\sum_{h=1}^{m} \bar{X}_h^2/\hat{\sigma}_h^2}{n\sum_{h=1}^{m} \bar{X}_h^2/\hat{\sigma}_h^2 + 1} \quad . \qquad (9)$$

This should be a better estimator than (6) (?). It has always a bias in the direction of 0, (which, by the way, is an improvement in comparison with (6)), and the larger the coefficient of variation of the turnover of cheese, the more low-priced premiums in automobile insurance. Presumably the estimator (9) should also be better the more data we have; By entering more irrelevant information, i.e. increasing m, we once more find that $\hat{P}_i$ approaches $\bar{X}_i$ in the limit.

A further improvement ought to be achieved if we admit also non-homogeneous linear estimators of the form

$$P_i = \zeta_0 + \sum_{h=1}^{m} \zeta_h \bar{X}_h \ . \qquad (10)$$

(This could, of course, be obtained with formula (7) by including in the tarrification basis some observations with variance 0, e.g. the observed number of Grossmünster monasteries in Zürich on different days.) Then we find that the optimal coefficients are $\zeta_{io} = \mu_i$ and $\zeta_{ih} = 0$ for all $h = 1,\ldots,m$, and hence the optimal stage (ii) - estimator of μ_i is μ_i. Good! To obtain a genuine (stage (iii)) estimator of μ_i we need only an estimator of μ_i ... And following Gerber's recipe we take the natural unbiased estimator $\bar{X}_i$ and obtain once more

$$\hat{P}_i = \bar{X}_i \ ! \qquad (11)$$

By way of summary conclusion the unbayesed approach seems to imply that the estimator (6) is better than (11). And vice versa. Many other interesting optimal estimators can be produced. You must only decide at the outset what appearance you would like the optimal estimator to have.

4. Properties of credibility estimators in models with a structure distribution

For this traditional model one finds that by known structure distribution the optimal estimator of the form (10) is [7] with Z defined by [16]. By unknown structure distribution one can arrange an empirical approximation to the optimal solution, which is asymptotically optimal by all structure distributions as $m \to \infty$. This solution is also asymptotically restricted minimax, and hence the linear form of the estimator has been justified.

5. Conclusion

Section 4 shows why the notion of structure distribution is not superfluous. Assuming that risks are chosen from a population, we arrive at credibility formulas without having to rely on tricks and artifices from outside of the model framework. "Unbayesed credibility" and similar ideas may perhaps serve a good purpose; Through their shortcomings they demonstrate the importance of building adequate models and working strictly within these. The following quotation from Neyman (1954) seems pertinent: "... the efforts of the representatives of modern statistical theory are directed towards solving problems that depend only on the stochastic model of the phenomena studied and on nothing else."

References

Gerber, H.U. (1982). An unbayesed approach to credibility. Insurance: Math. & Econ. 1, 271-276.

Neyman, J. (1954): Current problems of mathematical statistics. Proceedings of the International Congress of Mathematicians, Amsterdam 1954, Vol.1, 349-370.

Ragnar Norberg
Matematisk institutt
P.b. 1053, Blindern
Oslo 3
Norway