

UNBAYESED CREDIBILITY REVISITED

BY RAGNAR NORBERG

University of Oslo, Norway

ABSTRACT

The unbayesed credibility procedure proposed by Gerber is revisited. Its performance is discussed, connections are drawn to earlier literature, and some possible ideas of generalizations are investigated (and found fruitless).

KEYWORDS

Unbayesed credibility, principles of statistical decisions.

1. INTRODUCTION

More than two years have elapsed since GERBER (1982) proposed a procedure for construction of estimators highlighted as Unbayesed Credibility. During this time there has been published no further work on the topic. It is, therefore, worth while having another look at the unbayesed estimations to throw some light on their properties and to inquire if further ideas ought to be pursued along the same lines.

In Section 2 of the present paper two estimation problems considered by GERBER (1982) and Gerber's unbayesed approach to their solution are briefly recapitulated. The properties of the two unbayesed estimators are discussed in Section 3; it is shown that one of them will usually have an infinite expected squared loss. Section 4 stresses the need to build adequate mathematical models and to work strictly within these in search for methods. In particular the properties of any proposed method has to be examined in terms of the performance criterion adopted. Section 5 presents a couple of variations of the unbayesed approach which show that it can lead to many different estimators; the particular form of any unbayesed estimator is due to arbitrary restrictions imposed on the estimating functions rather than being due to the structure of the model itself.

2. REVIEW OF THE UNBAYESED ESTIMATION PROCEDURE

In order to make our presentation fairly selfcontained and to state points clearly, let us recall the unbayesed set-up in neutral mathematical terms. The model framework in Sections 4 and 6 of Gerber's paper is the following.

MODEL. Let X_{ij} , $i = 1, \dots, m (> 1)$, $j = 1, \dots, n$, be a collection of real random variables. For each i the X_{ij} , $j = 1, \dots, n$, have the same distribution, which we denote by F_i . All X_{ij} are mutually independent, and $F = (F_1, \dots, F_m) \in$

$\mathcal{F} \times \cdots \times \mathcal{F}^m$, where \mathcal{F} is the nonparametric family of all those distributions on the real line which possess a finite second order moment.

Let μ and σ^2 denote the mean and the variance, which are well defined functionals on \mathcal{F} , and put $\mu_i = \mu(F_i)$ and $\sigma_i^2 = \sigma^2(F_i)$, $i = 1, \dots, m$.

The vector of means, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$, is to be estimated. More precisely, let \mathcal{P} denote the class of all measurable m -vectorvalued functions of (only) the X_{ij} 's; we seek a $\mathbf{P} = (P_1, \dots, P_m) \in \mathcal{P}$ that is in some sense close to $\boldsymbol{\mu}$. Gerber considers two different measures of closeness, hence two problems, the first of which is the following (numbers in square brackets refer to formulas in Gerber's paper):

PROBLEM 1 Pick $\mathbf{P} \in \mathcal{P}$ such that

$$(1) [20] \quad \sum_{i=1}^m E_F(P_i - \mu_i)^2$$

is "small" (not "minimum" as stated by Gerber, see second remark below).

(Here E_F denotes the integral with respect to $F_1 \times \cdots \times F_1 \times \cdots \times F_m \times \cdots \times F_m$, the joint distribution of the X_{ij} 's.)

A couple of remarks are in order at this stage. First a formal one: Gerber phrases his problem as that of predicting, for each i , a future independent selection $X_{i,n+1}$ from F_i , the performance of a set of predictors P_i being measured by (1) with μ_i replaced by $X_{i,n+1}$. That problem is, however, equivalent to the one stated here because

$$E_F(P_i - X_{i,n+1})^2 = \sigma_i^2 + E_F(P_i - \mu_i)^2.$$

The second remark concerns realities: As it stands, problem 1 is not properly stated since for each choice of \mathbf{P} the expression in (1) is a functional depending on F . One cannot find a \mathbf{P} minimizing (1) for all F (the choice $P_i = \mu(G_i)$, with $G_i \in \mathcal{F}$, $i = 1, \dots, m$, is optimal in $F = (G_1, \dots, G_m)$, but poor in other points F where $\sum_{i=1}^m \{\mu(G_i) - \mu(F_i)\}^2$ is large). Thus, still loosely speaking, we can only require of \mathbf{P} that (1) should not be "too large" in "too many" points F . We leave these considerations for the time being and continue our recapitulation of the unbayesed approach.

Gerber constructs his unbayesed credibility estimator in the following manner:

METHOD 1. (i) As a first step, solve the simple problem of minimizing (1) as $\mathbf{P} = (P_1, \dots, P_m)$ ranges in the class \mathcal{P}' of functions with P_i of the form

$$(2) [7] \quad Z\bar{X}_i + (1-Z)\bar{X},$$

where

$$\bar{X}_i = \frac{1}{n_j} \sum_{j=1}^n X_{ij},$$

$$\bar{X} = \frac{1}{m} \sum_{h=1}^m \bar{X}_h,$$

and Z is a function of F (only). Minimum is readily found to be attained at

$$(3) [21] \quad Z = \frac{n \sum_{h=1}^m (\mu_h - \bar{\mu})^2}{n \sum_{h=1}^m (\mu_h - \bar{\mu})^2 + (m-1)\sigma^2},$$

where

$$\bar{\mu} = \frac{1}{m} \sum_{h=1}^m \mu_h \quad \text{and} \quad \sigma^2 = \frac{1}{m} \sum_{h=1}^m \sigma_h^2.$$

(ii) As a second step, replace numerator and denominator in (3) by their "natural" unbiased estimators [13] and [22] (Gerber's Section 4). Then the right-hand side expression in (3) turns into

$$(4) [23] \quad \hat{Z} = 1 - \frac{(m-1)\hat{\sigma}^2}{n \sum_{h=1}^m (\bar{X}_h - \bar{X})^2},$$

with $\hat{\sigma}^2 = \sum_{h=1}^m \hat{\sigma}_h^2 / m$ and $\hat{\sigma}_h^2 = \sum_{j=1}^n (X_{hj} - \bar{X}_h)^2 / (n-1)$. Upon replacing Z in (2) by \hat{Z} , we obtain a function $\hat{P} \in \mathcal{P}$, which is the unbayesed credibility estimator.

The second problem is the following.

PROBLEM 2. *The same as problem 1, but with (1) replaced by the componentwise expected squared error*

$$(5) [27] \quad E_F(P_i - \mu_i)^2, \quad i = 1, \dots, m,$$

(vector-valued)

The above remarks to problem 1 apply also to problem 2. The unbayesed procedure follows the same outline as in method 1:

METHOD 2. First minimize (5) as $P = (P_1, \dots, P_m)$ ranges in the class \mathcal{P}'' of functions with P_i of the form

$$(6) \quad Z_i \bar{X}_i + (1 - Z_i) \bar{X},$$

where each Z_i is a function of F . Proceeding in analogy to step (ii) of method 1, Gerber arrives at the estimator \hat{P} given by

$$(7) \quad \hat{P}_i = \bar{X}_i - \frac{m-1}{mn} \frac{\hat{\sigma}_i^2}{(\bar{X}_i - \bar{X})}, \quad i = 1, \dots, m.$$

Having summarized the present state of unbayesed credibility, we now set forth to study its merits in terms of concepts from estimation theory.

3 PROPERTIES OF THE UNBAYESED ESTIMATORS

We first consider problem 1 and the unbayesed estimator given by (2) and (4).

It ought, perhaps, to be said that it is unfortunate to speak of \hat{P} as the "real solution" to the problem of minimizing (1) (Gerber's Section 4), confer the second

remark to problem 1 above. Clearly, method 1 is only a preparatory piece of motivating heuristics, and the resulting \hat{P} is so far only a candidate estimator, whose performance has to be examined in terms of the criterion (1). This task has not been undertaken—in fact, not even mentioned—in the previous literature on unbayesed credibility, and no references are made to the closely related literature on compound estimation problems. Therefore, a few remarks are added here on these matters:

Very little is known about the possibility of solving (reasonably precise versions of) problem 1 under the present model with nonparametric \mathcal{F} . Some results on restricted inadmissibility have been established: For certain simple parametric subfamilies $\mathcal{F}_0 \subset \mathcal{F}$ one can construct estimators that in all of \mathcal{F}_0^m dominate old established estimators known to be uniformly optimal on \mathcal{F}_0^m with respect to the traditional performance criterion (5) when one restricts to the class of unbiased estimators. The first results of this kind appeared in fundamental papers by STEIN (1956) and JAMES and STEIN (1961). They considered the subfamily \mathcal{F}_0 of all normal distributions with variance 1 (say) and proved that the estimator

$$(8) \quad P^* = (\bar{X}_1, \dots, \bar{X}_m),$$

which is admissible on \mathcal{F}_0^m with respect to (5) and furthermore is uniformly minimum variance unbiased, does not even remain admissible when criterion (1) is adopted. If $m \geq 3$, it is dominated by the so-called James–Stein estimator P^{**} defined by

$$P_i^{**} = \left(1 - \frac{m-2}{n \sum_{h=1}^m \bar{X}_h^2} \right) \bar{X}_i, \quad i = 1, \dots, m.$$

To most statisticians this result came as a surprise, to some even as an unpleasant one, and there were signs of controversies between defenders of the traditional P^* on the one side and advocates of the new P^{**} on the other. Now there is no reason to discuss which is the better of P^* and P^{**} (on \mathcal{F}_0^m), because that question is settled by emotionless mathematics once the performance criterion is chosen. What can be discussed, is only the choice of criterion. That discussion is, however, not of a purely mathematical nature, but depends on the goals and attitudes of the decision maker.

In closing our comments on problem 1, we note that admissibility on \mathcal{F}_0^m of estimators of the James–Stein type has been extensively treated in the literature, see e.g., BERGER (1976). A survey of James–Stein estimations is given by EFRON and MORRIS (1973).

Let us now examine the unbayesed estimator designed for problem 2. Again by the second remark to problem 1, it is clear that \hat{P} given by (7) does not represent the solution to the problem of minimizing (5) (Gerber's Section 6). In fact, by inspection of (7), it is readily seen that \hat{P}_i assigns the value $+\infty$ to (5) on wide subsets of \mathcal{F}^m : If, for instance, the F_i are normal distributions, then $\hat{\sigma}_i^2$ is independent of $\bar{X}_i - \bar{X}$, and $(\bar{X}_i - \bar{X})^{-1}$ has no expected value. More generally, if the marginal distribution of $\bar{X}_i - \bar{X}$ has a point of increase in 0, then \hat{P}_i and hence $(\hat{P}_i - \mu_i)^2$ are usually not integrable.

Problem 2 in the present nonparametric model is one of the classics of statistics, and to the knowledge of the present author no alternatives to the natural unbiased estimator (8) have been proposed in the pre-unbayesed literature. Thus, in this case it would really be surprising if \hat{P} could be shown to have any good properties. And unpleasant as the interpretation of the model and problem 2 is that the estimation problems are unrelated in every respect; the samples are drawn in an independent manner from populations that have nothing in common, and the losses incurred by error of estimation are measured separately for each problem. A reasonable task for the theory would be to put a firm basis to the intuitive feeling that the estimator of μ_i should depend only on X_{i1}, \dots, X_{in} . If we are not able to justify the deletion of the X_{hj} , $h \neq i$, from the estimation of μ_i , then we would be in serious trouble: How could we then in a rational way choose the statistical basis for a given estimation problem? Which irrelevant data were not to be included? Which advice should we give to the practitioners?

As we have seen, the unbayesed approach gives rise to no such concerns. The traditional and intuitively sound estimator (8) remains an uncontested answer to problem 2.

4. MODEL AND METHOD

After the discussions in section 3 the question arises: What brought Gerber to enter the X_{hj} , $h \neq i$, into the estimation of μ_i ? Why didn't he use the "natural unbiased estimator" (8)? The reason seems to be that he had a particular interpretation in mind; X_{i1}, \dots, X_{in} are spoken of as being the claim amounts in n different years for risk no. i in a portfolio of m insured risks. A few remarks on basic principles of statistical decisions are called for:

The first step in a statistical analysis is to separate out of the situation those features that are believed to have some bearing on the problem and work them into a mathematical model. The model should give a surveyable and, as far as possible, true picture of the phenomena. If, for instance, the data stem from similar automobile insurance risks, the model ought to give precise content to the notion of similarity between these risks. The model in Section 2 fails to reflect the essential circumstance that automobile insurance claims have something in common that distinguishes them from data on e.g. soldiers' heights and turnover of cheese. One reasonable mathematical means of expressing this similarity between the risks is to regard them as selections from one and the same structure distribution (population). Thus the structure distribution is not "essentially superfluous" (Section 4 of Gerber's paper) to those who think they can learn something about a given risk by looking at other risks of the same kind.

Having decided on a model, the purpose of the decision has to be expressed in terms of a performance criterion. When this is done, one is left with the purely mathematical problem of finding decision functions with good performance.

In the traditional credibility analysis based on models with structure distributions, credibility estimators are obtained as logical consequences of the mathematical set-up. They are justified by their (restricted) optimality properties.

This is not the case with unbayesed approach in method 2 above. There the particular credibility appearance of the estimator could only be obtained by the analyst's intervening into the mathematics by prescribing a certain procedure and exempting it from the requirements expressed by the performance criterion. 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 stochastical model studied and on nothing else".

Considered as a statistical framework for the analysis of related risks, problem 1 together with the model in Section 2 plays an intermediate role. As explained above, the model may be judged as inadequate, but it still represents a reasonable partial description of the situation. The connection between the different rating problems is now established through the choice of the performance criterion (1). The unbayesed method 1, however, is until further without any support whatsoever in studies of its performance.

5. SOME VARIATIONS OF THE UNBAYESED PROCEDURE

Until the estimator resulting from method 1 has been investigated with respect to performance, it can, of course, not be excluded. But the unbayesed device as such can be put on test in other ways. One angle of attack arises from noting that the requirement that Z in (2) be independent of ι is quite arbitrary.

Looking for good estimators, we could possibly gain something by allowing Z to depend on ι , that is, let P be of the form (6). But then the unbayesed procedure reduces to that of method 2 and delivers (7), which maximizes the expected loss instead of minimizing it as pointed out already in Section 3.

Let us allow for further flexibility and admit nonhomogeneous estimators of μ_ι of the form

$$P_\iota = Z_{\iota 0} + \sum_{h=1}^m Z_{\iota h} \bar{X}_h.$$

Then we find by the first step of the unbayesed prescription that the optimal approximation to μ_ι is μ_ι . And following the recipe further, we now only have to estimate μ_ι by a natural unbiased estimator. Then we obtain finally (8), which would have resulted immediately if we already at the outset looked for a natural unbiased estimator.

Which ends the present discussion. Some further considerations around the topic of unbayesed estimations can be found in an unpublished report, NORBERG (1983), which can be received upon request.

REFERENCES

- BERGER, J O (1976) Admissible Minimax Estimation of a Multivariate Normal Mean with Arbitrary Quadratic Loss *Annals of Statistics* 4, 223-226
- EFRON, B and MORRIS, C. (1973) Stein's Estimation Rule and its Competitors—An Empirical Bayes Approach *Journal of the American Statistical Association* 68, 117-130
- GERBER, H U (1982) An Unbayesed Approach to Credibility *Insurance. Mathematics & Economics* 1, 271-276
- JAMES, W and STEIN, C (1961) Estimation with Quadratic Loss *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1, 361-379 Univ of California Press
- NEYMAN, J (1954) Current Problems of Mathematical Statistics *Proceedings of the International Congress of Mathematicians, Amsterdam 1954*, Vol 1
- NORBERG, R. (1983) An Unbased Approach to Credibility, *Statistical research report No 1, 1983* Institute of Mathematics, University of Oslo
- STEIN, C (1956) Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1, 361-379. Univ of California Press

CORRIGENDUM

B. Sundt (1982). Asymptotic behaviour of compound distributions and stop-loss premiums. *ASTIN Bulletin* **13**, 89-98.

In the proof of Theorem 3, one uses that $g_s = H(s) - H(s+1)$ instead of the correct expression $g_s = H(s-1) - H(s)$. This means that the displayed asymptotic expression for g_s really concerns g_{s+1} , and that the correct result should be

$$g_s \sim \frac{1-p}{\nu} e^{-\kappa s}.$$

The author became aware of this error when reading Milidiu and Jewell (1984), whose Theorem 2 gives a generalization of this result.

REFERENCE

MILIDIU, R L and W S JEWELL (1984) Asymptotic approximation to compound negative-binomial distributions Submitted for publication