ASSESSING THE ACCURACY OF ESTIMATES: A PERSONALIST APPROACH

Rex V. Brown, University of Michigan

Introduction

The orientation of this talk is strictly practical, in the sense that the ultimate beneficiary is intended to be the man of affairs, like a businessman, who may use the fruits of the statistician's labors, rather than the statistician himself. If the approach to be presented does not contribute to the technology of business administration and other applied arts, it has failed in its primary purpose. If it happens also to have some academic interest, so much the better.

At a theoretical level, however, there is nothing very original in the approach. It does, however, appear to represent a novel attack, at an operational level, on a scandalously mishandled problem in decision-oriented quantitative research of all kinds. This problem is a trite and universal one for anyone who has to make decisions in the face of uncertainty; and it is in two parts:

1. How to assess uncertainty about relevant target variables (such as a market share), which I will call the problem of Target Assessment.
2. How to evaluate ways of reducing this uncertainty, which I will call the problem of Research Design.

Now, I am well aware, of course, that the literature abounds with procedures that appear to address these problems. In the face of sample findings, for example, such devices as Confidence Intervals, Maximum Likelihood Estimates, and Tests of Significance, certainly seem to be saying something about Target Assessment. However, the trouble with classical inference tools such as these is that their output is not in a form that is of direct interest to a decision maker. He wants to answer the very personal question, "Where does my target variable probably lie?", whereas a Confidence Interval, for example, is telling him how surprising the observed research results would be if some variable, (not necessarily his target variable) had some hypothetical values.

When he considers a possible Research Design he wants to answer the question, "What research can I do which will make me least uncertain?", rather than, "What research will produce the smallest sampling variance, from among those research designs for which a sampling variance can be objectively calculated?" The weaknesses of classical inference for decision-oriented purposes are well known and do not need to be covered in any detail here.[1]

When assessing his target variable, the decision maker surely wants to come up with a personal probability assissment, possibly in form shown in Figure 1. However, he may not care to represent his assessment as a complete probability or credence distribution (as I prefer to call a personal probability distribution when it refers to an assessment conditional only to what the assessor actually knows).[2] When choosing between research designs, he will also surely want to look ahead to the kind of probabilistic assessment he will make after the research. Presumably, he will opt for the design which, in some sense, promises to produce a personal probability distribution with as little "spread" as possible.

Personalist Inference

What is commonly known as Bayesian inference is, of course, designed to handle exactly this kind of problem and substantial development in the area has been achieved, notably by Savage, Raiffa, and Schlaifer. However, not very much of this development has yet trickled down into real world usage.

I have recently completed a brief survey of what use business is making of Bayesian decision theory. While I found plentiful and even dramatic usage of decision trees and other devices for analyzing decisions given specified assessments of uncertainty about the relevant target variables, I found almost no use of the conventional Bayesian devices for determining those uncertainties. I am referring, of course, to prior-posterior analysis and its derivative preposterior analysis. At least, I was able to find very few instances where executives acted on the implications of such Bayesian analysis.

Now, part of this lack of implementation is no doubt due to the quite natural time lag between a new technology being developed at a theoretical level, and its becoming operational. But part, I think, is because the technique itself, as currently developed, is not always appropriate for use by non-technical decision makers or, as I will call them, executives.

Both prior-posterior and preposterior analysis depend upon the use of Bayes' Theorem and require two critical inputs which the executive has, if not necessarily to supply, at least to agree with. In the first place, a prior distribution must be assessed on the target variable and it may be exceedingly difficult to make an uncontaminated prior assessment after research evidence has already been obtained. Commonly, this is the point in time at which a prior posterior analysis will be initiated. Secondly, a likelihood function must be assessed, specific to actual or potential research findings, and it may be very difficult for a non-technical executive to grasp what it means, let alone contribute to its formulation. Except in those rare cases where the data generating process, conditional on the true value of the target variable, is uncontroverailly known (e.g., random sampling with perfect measurement), informed judgment needs to go into the construction of a Likelihood Function. The most appropriate judgment will often be in a head unable to express it in a Likelihood Function form.

In addition to these difficulties of eliciting needed inputs, I have found that very few executives feel that they understand even the general idea of prior-posterior or preposterior inference. For this reason they are understandably hesitant to trust decisions that may involve millions of dollars to an analysis based on an arcane logic.

Is there any way of avoiding these drawbacks? I think there is, and I would now like to propose an alternative which, while it is Bayesian, in the sense that it accepts personal inputs and its output is interpreted personally, it does not depend on Bayes' Theorem. (I think it would be a very good convention if we could agree to use the word "Bayesian" only for those types of personalist analysis which depend on Bayes' Theorem.)

Credence Decomposition

The alternative I am about to propose depends, not on Bayes' Theorem, but on the equally well known logic of the distribution of functions of random variables. I call it Credence Decomposition and, in the form I shall now present, it can be used both for problems of Target Assessment and Research Design.

The essential steps are very simple and are as follows:

1. The target variable is decomposed, by which I mean that it is expressed as a function of two or more components. A very simple example would be to express future sales of a product as sales per outlet times number of outlets. A slightly more elaborate decomposition (and decompositions can get very elaborate) would be to express future sales as the sum of multiplicative expressions of the above form for each of a number of market sectors.
2. Each component thus defined is assessed probabilistically on whatever evidence is available to the assessor. This could include field work, judgment or published statistics and the supporting reasoning could be any combination of intuition and statistical theory (including, possibly, prior-posterior analysis).
3. A personal probability distribution, e.g., in the form of Figure 1, is derived routinely by any of a number of standard statistical procedures. Computer programs, formulas and other supporting devices have been developed to make this processing as painless as possible.

At this level of generality, Credence Decomposition is a rather trivial (if unexploited) tool. However, there is a variant of Credence Decomposition which is less obvious and which seems to lend itself rather conveniently to problems of target assessment and Research Design. I call this Error Decomposition.

Error Decomposition

In Error Decomposition, what is decomposed is not the target variable of ultimate interest to the executive but the estimating error resulting from a specific piece of quantitative research, e.g., a sample survey.

There are at least two ways of formally defining estimating error. It can be defined as the difference between the target variable and some more or less arbitrary estimate calculated from the research findings. Alternatively, it can be the ratio of the target variable to such an estimate. Either formulation has advantages in different circumstances, and for illustrative purposes, I will discuss only the error ratio form.

Let me take as a concrete example a parking survey situation that has been written up in some technical detail elsewhere.[3]

The case setting is as follows: A British town, which I will call Camford, had a mail survey done to assess the probable demand for parking space if meters were introduced. A list of ten thousand locally registered motorists was obtained and, of these, one thousand were randomly selected and sent a mail questionnaire. Nine hundred returned the questionnaires and of these ten percent or ninety indicated that, if meters were introduced, they would be parked in the downtown area at a given peak hour.

The city engineer's target assement problem is, what should he conclude personally about the actual demand for parking space if meters were introduced, expressed in a form like Figure 1? He also has a research design problem. If he conducts a new survey in another town, should he use the same budget on another mail survey or on a smaller personal survey.

On the target assessment problem, the first thing the city engineer might do, using Credence Decomposition, would be to decompose the total spaces needed (if meters were introduced) as the product of:

> The fraction of local motorists needing space (t)
> TIMES
> Number of local motorists (n)
> TIMES
> Some adjustment factor to allow for spaces needed by out-of-town parkers (f)

If probabilistic assessments can be made for each of these, a credence distribution on the target variable can be derived routinely. The number of local motorists (n) is known to be ten thousand, so no probabilistic assessment is needed of that component. The out-of-town adjustment component (f) can be assessed by direct intuition informally. This leaves the "local fraction," (t), which is the variable which the mail survey addresses. This is the assessment he might use error decomposition for.

Figure 2 shows the essential steps the assessor might go through, viz.: to express total error ratio as a function of component ratios which reflect distinguishable (and assessable) sources of error. The nested rings at the top of the figure and the vertical lines indicate the various ways in which sources of error can creep in between the true value of the target t, "local fraction," and the estimate a' (known to be ten percent).

t/a' is then the total error ratio and the component error ratios are defined in the line in Figure 2 marked "Decomposition." It can be seen that three sources of error are distinguishable: random error, nonresponse error, and reporting error. It can easily be verified that each of these will be one if there is no error of the type involved. The set of boxes on the right hand side of the bottom line of Figure 2 summarize, in the form of ninety-five percent credible intervals, probabilistic assessments that were made for each of the three component error ratios.

The detail of these assessments and the logic behind them are described in the reference cited in Footnote 3. Suffice it to say here that the random error was based on prior-posterior analysis using a flat prior assumption, and the other two errors were assessed intuitively. The resulting distribution of t/a' was approximated by means of a formula which exploits the fact that the relative variance of a product is approximately the sum of the relative variances of the components. (Also discussed more fully in the above reference.) It could also be computed more exactly by a computer program called DECOMP.[4]

The credence distribution on t follows directly from the distribution on t/a' and, as a summary, its credible interval appears in the left hand box in Figure 2.

The city engineer might thus conclude, if he accepts the input assessments, that he can be ninety-five percent sure that the local fraction, t, lies between six percent and twenty-one percent. Conjoined with the knowledge that there are ten thousand local motorists and an assessment of the "out-of-town adjustment" with a credible interval of 1 to 1.2, a credence curve on the real target variable, total spaces needed, was derived, which actually corresponds to Figure 1. His target assessment would therefore be that between three hundred and twenty-two hundred parking spaces will be needed if meters are introduced, with ninety-five percent personal probability.

As for the research design problem, the city engineer would go through virtually the same procedure for each of the alternative research designs considered. If the cost is the same, he might reasonably choose which ever strategy produces the smallest span for the credible interval on the total error ratio. (Alternative research design criteria can be selected, such as prior expectation of posterior variance, but they seem to produce almost identical rankings.)

Conclusion

There has only been time this afternoon to rather briefly sketch the class of algorithms which I call Credence Decomposition. A fuller technical discussion appears in the article referenced in Footnote 3. A general handbook on the technique and its applications is in process of publication.[5]

The general Credence Decomposition technique and the Error Decomposition variant of it seems to work reasonably well in the Marketing Research area in which I am most familiar. At the present time, it certainly seems to suffer from very little competition as far as practical tools for research appraisal are concerned.

Other researchers, notably Professor Charles Mayer of York University, are working on the critical problem of how to make reasonable, empirically based component assessments which are required by this technique or others with the same objectives.[6] I would certainly appreciate hearing from anyone else who may be working in this area.

FOOTNOTES

1 See J. W. Pratt, H. Raiffa, and R. Schlaifer, "Introduction to Statistical Decision Theory," McGraw Hill, 1965, Chapter 20.

2 Thus a sampling distribution conditional on some hypothetical value of a population parameter would not be a credence distribution though it might well be a personal probability distribution.

3 See Rex V. Brown, "The Evaluation of Total Survey Error," The Statistician, Volume 17, No. 4, 1967, (copies of this paper are available from the author).

4 Further information about this program can be obtained from the author.

5 Rex V. Brown, "Research and the Credibility of Estimates," Harvard University, Graduate School of Business Administration, Division of Research, Boston, 1969.

6 Charles S. Mayer, "Assessing the Accuracy of Marketing Research," to be published.