COMPARING TWO PROBABILITY APPRAISERS

by

Glen Meeden

FSU Statistics Report M400

February 1977

THE FLORIDA STATE UNIVERSITY
DEPARTMENT OF STATISTICS
TALLAHASSEE, FLORIDA
COMPARING TWO PROBABILITY APPRAISERS

by

Glen Meeden

Summary

Consider the situation where an expert is required to express his or her prior opinion by specifying a probability distribution. Suppose after a series of trials one wishes to determine which of two experts is preferred. This note discusses two methods for making this decision.

Key words: Bayesian decision making, prior distribution, scoring function, subjective probability.
1. Introduction

Consider the situation where an expert is required to express his or her opinion by specifying a probability distribution. The distribution is assumed to be over a mutually exclusive and exhaustive list of possible states of nature, where the true or actual state of nature is unknown to the expert at the time of specification. Two common examples where this occurs are weather forecasting and giving odds for outcomes of competitive contests. Suppose two experts are available and one must decide which of two experts to use. Two methods for comparing the experts are discussed in Winkler (1971). In the first, after the true state of nature becomes known the expert is given a score which measures how close his or her distribution was to the actual state of nature. The expert with the better total score over a series of trials is preferred. In the second, the two place systems of bets against the other's stated odds and settle up after the actual state of nature becomes known. The winner after a series of trials is preferred. In this note the consequences of using these two methods for comparing experts is explored. It is shown how the two methods do not necessarily yield the same preferred expert when it can be assumed that the experts' prior distributions are estimates of a true but unknown distribution.

2. The Two Methods.

Let \( A_1, \ldots, A_k \) be the list of mutually exclusive and exhaustive states of nature under consideration. Let \( P = (P_1, \ldots, P_k) \) be the probability distribution assigned to the states of nature by the first expert. We assume that for each \( i, 0 < P_i < 1 \) and \( \sum_{i=1}^{k} P_i = 1 \). Similarly let
\( \lambda = (\lambda_1, \ldots, \lambda_k) \) be the probability distribution assigned to the states of nature by the second expert where for each \( i \) \( 0 < \lambda_i < 1 \) and \( \sum_{i=1}^{k} \lambda_i = 1 \). We assume that the chosen distributions actually reflect the experts' prior opinions. In practice this may be difficult to verify, e.g. see Hogarth (1975) and Savage (1971). Once the actual state of nature is revealed one must decide which of the two experts' distribution was more accurate.

The first method of comparing the experts uses scoring functions. Suppose \( A_i \) is the actual state of nature then the first expert is given the score

\[
1 - (1 - P_i^0)^2 - \sum_{i \neq 1} P_i^0 2
\]

and the second expert is given the score

\[
1 - (1 - \lambda_i^0)^2 - \sum_{i \neq 1} \lambda_i 2
\]

An expert's score can range from zero to one, the larger the better. The score measures how close the experts' distribution was to the true state of nature. After a series of trials (say repeated weather forecasts) the expert with the larger total score is preferred.

In the second method the experts place systems of bets against the other's stated odds. Let \( E \) be a proper subset of the set of all possible states of nature. Let \( P(E) = \sum_{A_i \in E} P_i \) and \( \lambda(E) = \sum_{A_i \in E} \lambda_i \). Suppose the first expert places a stake of \( s \geq 0 \) on \( E \) and a stake of \( t \geq 0 \) against \( E \). Then the second expert pays the first expert

\[
([1 - \lambda(E)]|\lambda(E)) s - t \quad \text{if } E \text{ occurs}
\]

and

\[
(\lambda(E)|(1 - \lambda(E)) t - s \quad \text{if } E \text{ does not occur}
\]

E occurs if the true state of nature is found to belong to \( E \). Each expert
may place any combination of stakes for or against proper subsets of the set \( \{A_1, \ldots, A_k\} \). The total amount each expert may bet is assumed to be one unit. We now find the optimal system of bets for the first expert.

Without loss of generality we assume that the states of nature are numbered in such a way that

\[
(\lambda_1 | p_1) \leq (\lambda_2 | p_2) \leq \ldots \leq (\lambda_n | p_n)
\]

We may assume that at least one of the above inequalities is strict. If not both experts has specified the same distribution. Let \( E \) be as above and suppose the first expert places a stake of \( s \geq 0 \) on \( E \) and a stake of \( t \geq 0 \) against \( E \). Since \( P(E) \) is the first expert's probability that \( E \) will occur the expected gain of this bet for the first expert is

\[
\{(1 - \lambda(E))|\lambda(E))s - t\} P(E) + \{((\lambda(E)|(1 - \lambda(E))) t - s\} (1 - P(E))
\]

\[
= \{[P(E)|\lambda(E)] (1 - \lambda(E)) - (1 - P(E))\} s
\]

\[
+ \{[(1 - P(E))|(1 - \lambda(E))] \lambda(E) - P(E)\} t
\]

Note that the coefficient of \( s \) is positive if and only if the coefficient of \( t \) is negative which is equivalent to \( P(E) > \lambda(E) \). From this it follows that the first expert should place stakes only on those sets \( E \) for which \( \lambda(E)|P(E) < 1 \). If the first expert places a stake of \( s > 0 \) on such an \( E \) and a stake of 0 against \( E \) then the expected gain of the first expert is

\[
\{[P(E) - \lambda(E)]|\lambda(E)\} s
\]

Hence the system of bets which maximizes the first expert's gain is to place his entire stake of one unit on the set \( E \) where \( E \) is such that
\[P(E) - \lambda(E)]|\lambda(E)\] is maximum or equivalently when \(\lambda(E)|P(E)\) is minimum which occurs when \(E = \{A_1\}\). The first expert places his entire stake on a state of nature for which he believes the second expert has underestimated by the largest percentage. Similarly the second expert places her entire stake on \(E = \{A_k\}\), a state of nature for which she believes the first expert has underestimated by the largest percentage. After a series of trials the expert with the larger total winnings is preferred.

3. Comparing the Two Experts.

Both these methods are easy to use; see for example Winkler (1971) and de Finetti (1972). Note that the resulting preferences need not be the same. We now wish to discuss which method should be used to choose a preferred expert. We are assuming that the experts are idealized probability assessors, that is the stated distributions truly reflect their respective beliefs, opinions and information about the situation under study. We only wish to decide which of the two experts' judgment is preferred in the light of experience. Hence a more apt title for this note would be "Comparing two reality assessors whose judgments are expressed through probability distributions". The experts' probability distributions are to be judged by what they say about reality. Without making some assumptions about what is being assessed it does not seem possible to choose between the two methods in any sensible way.

One simple assumption is that the true state of nature is chosen from the set of possible states of nature according to the distribution \(\theta = (\theta_1, \ldots, \theta_k)\) where \(\theta_i\) is the probability \(A_i\) is chosen. In this
case we shall now see that it is easy to compare the two methods for choosing the preferred expert.

If $\theta_i$ is the true distribution of the $A_i$'s then the first expert's expected score is

$$
\sum_{i=1}^{k} \left[ 1 - (1 - p_i)^2 - \sum_{l=1}^{p_i} \theta_i \right] = \sum_{i=1}^{k} \theta_i^2 - \sum_{i=1}^{k} (\theta_i - p_i)^2.
$$

The first expert's expected score is bigger than the second expert's expected score if and only if

$$
\sum_{i=1}^{k} (\theta_i - p_i)^2 < \sum_{i=1}^{k} (\theta_i - \lambda_i)^2.
$$

That is, the first expert is expected to be preferred if and only if his distribution is closer to the true but unknown distribution.

On the other hand, if the second method is used the expected winnings of the first and second experts are $$(\theta_1 - \lambda_2)/\lambda_1$$ and $$(\theta_k - p_k)/p_k$$ respectively. Hence the first experts expected winnings are greater if and only if

$$\frac{\theta_1}{\lambda_1} > \frac{\theta_k}{p_k}.$$

that is the first expert is expected to be preferred if and only if he does a better job of estimation at the two states of nature where their disagreements are the largest.
These remarks help to clarify the differences between the two methods for choosing a preferred expert. The first will tend to pick the expert who on the average does the better job in the overall estimation of the distribution. The second will tend to pick the expert who on the average does the better job estimating the probability of the two states of nature where their disagreement is the largest. Now the two states of nature which yield the greatest percentage disagreement obviously depend on the two experts involved. However it seems reasonable to assume that these states of nature will often be those for much the $\theta_i$'s are very small. Hence it seems that the second method should be preferred when one desires the expert who can give better estimates for the probabilities of the rare or extreme states of nature. On the other hand, the first method should be used when one desires the expert who can use the better overall estimate of the distribution.

In the modern subjective Bayesian approach to statistics the distribution of the expert represents his or her prior beliefs and knowledge about the possible states of nature. The distribution is only a statement about how the expert perceives reality and does not necessarily represent any random phenomena of nature. In fact, de Finetti takes the position that "probability does not exist", see for example de Finetti (1974). If this is the case then the assumption of the existence of a true but unknown distribution $\theta$ is untenable. Consider however the case of the expert trying to assess the probability of the possible outcomes of a contest, say a football game or a chess match. The expert's uncertainty not only reflects the imprecision and
incompleteness of his or her knowledge about the contest but it stems from
the nature of the contest as well. In fact, it is the uncertainty of the
outcome which constitutes one of the major charms of such contests. In such
cases as these, it is possible to imagine an infinite sequence of such
contests played under identical conditions and that the outcome of any
particular match follows the long run distribution. Hence the distribution
\( \theta \) represents the uncertainty inherent within the contest independent of
the expert's beliefs. Such a distribution is clearly a fiction and the
willingness to assume its existence is a matter of personal choice. In
many instances such an assumption is completely unjustified. We believe
however that sometimes it is a reasonable assumption to make.

Usually a person will prefer the expert whose opinions, when applied
to their affairs, will in some sense maximize their expected earnings. In
this note we have discussed two methods for choosing between experts. We
have noted that the choice between the competing experts can be made in a
sensible way, when the expert's prior distributions are estimates of a
true but unknown distribution. When this is not the case there does not
seem to be any known justification for choosing between the two methods.
For example, this problem is discussed in de Finetti (1972). Although
he does believe that it is possible to realize that one expert is doing
a better job than another he thinks that choosing a preferred expert is
a personal decision which cannot be governed by a mechanical rule. We
have argued here that sometimes it is possible to pick a method of deter-
mining the better expert which reflects the decision makers personal goals.
REFERENCES


