ON COMBINING EXPERT OPINION

SAMRIT CHATTERJEE

TECHNICAL REPORT NO. 49
JUNE 1981

PREPARED UNDER THE AUSPICES OF
SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

SIAM

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON COMBINING EXPERT OPINION

Samprit Chatterjee
Stanford University

and

New York University

TECHNICAL REPORT NO. 49

JUNE 1981

STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH (SIMS)

PREPARED UNDER SUPPORT TO SIMS FROM

Department of Energy (DOE)
Sloan Foundation
Environmental Protection Agency (EPA)
National Science Foundation (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SUMMARY

In various decision contexts, a decision maker has to combine the estimates (information) supplied by several experts whom he has called in for advice. In this paper we present methods which may be used to combine the different estimates. The role of interaction between the experts and the emergence of consensus is also discussed. Methods are presented for combining expert estimates in situations where along with the expert estimates a small amount (sketchy) of observational data is also available.
ON COMBINING EXPERT OPINION

by

Samprit Chatterjee*
New York University

1. Introduction

In any formal Bayesian decision making situation two elements which are always present include an utility function and a prior probability distribution over the possible states of nature. Since the decision maker is going to be facing the consequences which might result from the decisions made, it is felt that the utility function should be generated by the decision maker. The present writer feels that the same however is not true for the prior probability distribution. The prior probability distribution should reflect the best available knowledge on the states of nature (the unknown parameters) be that objective or subjective. To get the best available knowledge and information a decision maker may summon or consult a panel of experts. (The question of who is an expert, what does expertise consist of are interesting and difficult questions but will not be discussed in the present paper.) When the panel of experts agree on the unknown states of nature there is no problem. A problem arises however when the different members of the panel have different estimates of the unknown parameters or come in with different assessments of the unknown states of nature. The present paper will address itself to the question of devising possible ways of combining expert opinion.

*This work was done while the author was visiting the Statistics Department at Stanford University.
The problem of combining expert opinion may also arise in other contexts. In a regulatory setting, for setting tolerances an administrator might conceivably consult several experts. If the experts differ among themselves, what is the most appropriate method of combining the expert opinions. We shall present and explore several schemes for combining expert opinion.

The second problem we consider in this paper will be devising appropriate methodology of combining expert opinion in situations where along with the expert opinion we have sketchy or spotty data on the question at hand. How best to weight the expert opinions and the available sample information, meagre though it may be, to arrive at the best available estimate. For a summary of previous work the reader is referred to [1], [2], [3].

2. The General Problem

The problem of combining expert opinion arises most acutely when the opinions (estimates) differ sufficiently so that acting on it lead to different courses of actions. There are situations when although the experts do not exactly agree, the differences are such that the best course of action based on the several opinions are the same. Under these situations we do not have to resolve the differences between the experts. We are concerned solely with the cases where the different expert estimates lead to different "best" courses of action. In situations where experts differ significantly we will distinguish two broad cases. These are i) interaction between the experts is not possible ii) interaction between the experts is possible. In situations where interaction between experts is possible, the interaction may lead to consensus, or might narrow the difference gap,
so that the difference is slight and lead to the same course of action.
In section 3 of the paper we will examine the question of interaction
between the experts, the problems of reaching consensus. Even after
the interaction between the experts, there may be substantial differences,
and the problem of combining expert opinion still remains. We make a
suggestion based on personal experience, that where several experts differ
every attempt should be made to allow them to interact with each other.
Often the differences are resolved, the divergence narrowed, and the problem
at hand illuminated. There are other situations however when no interaction
between the experts is possible (physical separation, confidentiality, etc.)
and the differing expert estimates have to be combined by the decision maker.

3. **Combining Estimates With No Interaction Between Experts**

Let us suppose we have \( k \) experts, and their estimates are denoted
by \( \theta_1, \theta_2, \ldots, \theta_k \). The \( \theta \)'s may be scalar, vector or even probability
distributions. We describe three possible schemes, each of which has
merit and may be suitable in a particular situation.

a) **Equal Experts Situation**

In situations where the decision maker cannot rank the experts, or
feels that the level of expertise is the same for each of the \( k \) experts,
the most obvious estimate is \( \bar{\theta} \) the average of the \( k \)-estimates.

\[
\bar{\theta} = \frac{1}{k} \sum_{i=1}^{k} \theta_i .
\]  

(1)

An alternative to (1) which should be explored is the trimmed mean. If
\[ \theta(1) \leq \theta(2) \leq \cdots \leq \theta(k) \]

denote the ordered estimates, then the one-trimmed mean is obtained by leaving out the largest \( \theta(k) \) and the smallest \( \theta(1) \) estimate. The one trimmed mean

\[ \bar{\theta}_T = \frac{\theta(2) + \theta(3) + \cdots + \theta(k-1)}{k-2} \]  \hspace{1cm} (2)

By deleting the 2 largest and the 2 smallest values we can get the trimmed mean. The merits of the trimmed mean over the mean is discussed Huber [4] and carries over in the present context. It is significant that in some sporting events (like diving and figure skating) where the team of judges are regarded all as experts of equal ability, a form of trimmed mean is used.

b. **Ranked Experts Situation**

If the decision maker can rank the experts, it is obvious that the estimate (1) should be modified to a weighted estimate \( \bar{\theta}_w \) where

\[ \bar{\theta}_w = \sum_{i=1}^{k} W_i \theta_i \quad W_i \geq 0 \quad \sum W_i = 1. \] \hspace{1cm} (3)

Where the experts can be ranked, there \( W_i \) may be taken as

\[ W_i = \frac{R_i}{R_1} \]

where \( R_i \) is the rank of the \( i \)th expert. The most well regarded expert gets the rank \( k \), the next best \( k-1 \), and so on. The greatest weight is
given to the estimate of the expert who is most well regarded and so on down the line. Other schemes of weighting may be considered, the rank weights appear to us simple and natural.

c. **Pseudo-Bayesian or Weight of Experience Method**

In the traditional Bayesian method, prior information is combined with sample information by multiplying the prior distribution with the sample likelihood. The parameters of the resulting posterior distribution embody all the available information. In the case of combining expert opinion we have essentially no sample information. If however we treat one experts distribution as a prior, and the second experts distribution as the likelihood then Bayes theorem provides us with the mechanism for combining them. This procedure can then be repeated successively to combine all the experts. Since the procedure is symmetric with respect to the experts, it will given the same estimate regardless of which two experts we start with. We give an example to clarify the method.

The unknown probability $\pi$ for an event is to be assessed and we have $k$ experts. The experts assessments are $p_1, p_2, \ldots, p_k$. Besides obtaining the experts probability assessment we also determine from each of the experts the numbers (subjective) of similar cases which they feel they have encountered and on the basis of which they have arrived at the probabilities they have furnished. Suppose $(n_1, n_2, \ldots, n_k)$ be the cases on which expert $1, 2, \ldots, k$ have based their estimate. Let $r_i = n_i p_i$ $i = 1, 2, \ldots, k$. Then we assume that the $i$th experts assessment of $\pi$ is $B(r_i, n_i)$.** From the conjugate properties of the Beta distribution it

**The Beta distribution with parameters $r_i$ and $n_i$.**
follows that for all the $k$ experts the distribution of $\pi$ is $B(\Sigma r_i, \Sigma n_i)$.

As our estimate of the unknown parameter $\pi$ we take the mean of the pooled distribution which is

$$\pi^* = \frac{\Sigma r_i}{\Sigma n_i}$$

(4)

If we applied (1) in the present example our estimate $\bar{\pi}$ would be

$$\bar{\pi} = \frac{1}{k} \sum_{i=1}^{k} \frac{r_i}{n_i} = \frac{\Sigma p_i}{k}$$

(5)

The difference between (4) and (5) is that in (5) all the experts opinions are equally weighted, whereas in (4) experts whose estimates are based on larger number of experiences receive a greater weight. The estimate $\pi^*$ can be thought of as

$$\pi^* = \sum_{i=1}^{k} W_i p_i$$

(6)

where

$$W_i = \frac{n_i}{\Sigma n_i}$$

(7)

(4) indicates why we have called it Pseudo-Bayesian while (6) and (7) indicate why the scheme may also be regarded as the weight of experience method.
4. Interaction Between the Experts

When several experts differ in their assessment, some of the differences may be reduced by allowing the experts to interact with each other. Several schemes have been suggested, the most widely known being the Delphi method (Dalkay [5], Linstone and Turoff [6]). In the Delphi method, the experts after their initial estimates are provided with group summaries such as the mean, median and the quantiles of the pooled estimates. In the light of all the estimates, the experts are asked to make a second set of estimates. This is continued until a consensus is reached or no change occurs in the estimates; the experts have hardened positions. It has been found that the gap between the estimates often have been narrowed, so that some of the methods described in the previous section can be applied. The criticism of the Delphi method has been that it gives only a rigid quantitative feed back, and may have built into it movements towards consensus. Another variation of the Delphi method -- controlled feedback -- is the Shang Inquirer (Ford [7]). This method also has been criticized for similar reasons. An interactive procedure with a more qualitative feedback has been considered by DeGroot [8], Chatterjee and Seneta [10], and Press [11]. In [10], the experts are provided with their several initial estimates, along with the biographical details of the experts, which enables the participants to determine subjectively the weights which one might assign to the various estimates. Several models are proposed for describing the generation of successive estimates. These are essentially theoretical schemes, which have been suggested from behavioral considerations. Conditions for reaching consensus when the weights attached by each of the experts to the other expert's
estimates undergo change in the successive stages of the estimate exchanging process is obtained. For a group of open minded participants (i.e., experts who believe that with successive round of estimates the individual information bases are becoming a common pooled information base) it is shown that consensus is always reached. Consensus may also be reached in this "information pool" set up, even when there is a hardening of position among the participants (each participant putting more and more weight on their own estimate) provided the hardening does not take place too rapidly. The mathematical and other details are found in [9], [10]. Press [11] has proposed a very elaborate qualitative feedback procedure, and has suggested methods for combining the several estimates to arrive at a consensus estimate. The method is highly theoretical based on numerous assumptions, and does not seem to have much practical value.

When several experts are used, if it is physically possible some attempt should always be made to allow for possible interactions. This operation has several advantages. It might often produce a consensus, narrow the difference of gap, and may unearth factors which were not considered before.

The above discussion can be summarized in the following diagram.
5. Combining Expert Opinions in Presence of Scanty or Patchy Data

There are many situations where besides expert estimates, there is some real life data. The amount of data however is very small and cannot be entirely relied upon. The question then is how to incorporate the expert opinions, with the available data. We will assume that among the k experts there is nothing to distinguish one from the other i.e., the decision maker regards all of them equally knowledgeable.

Let the several estimates, as before, be denoted by $\theta_1, \theta_2, \ldots, \theta_k$. We will denote the likelihood of the observed sample by $L(\theta)$. Two methods for combining the expert estimates, and incorporating the sample data suggests itself. These are

i) Likelihood Weighted Estimates

ii) Combined Expert-Data Estimates.

a. Likelihood Weighted Estimates

Let $L(\theta_i)$ be the likelihood of the observed sample for the $i^{th}$ experts estimate $\theta_i$ of the unknown parameter. The likelihood weighted estimator

$$\theta_{LW} = \frac{\sum_{i=1}^{k} \theta_i L(\theta_i)}{\sum_{i=1}^{k} L(\theta_i)}$$

$\theta_{LW}$ is the weighted mean of the several expert estimates the weights being proportional to the likelihood. The estimator has the following properties

i) it is the posterior mean of the experts estimates when an uniform prior is placed on each of the experts;
ii) experts whose estimates explain the observed data better get higher weights (i.e., estimates with higher likelihoods).

iii) In situations where the likelihoods are the same, $\theta_{LW}$ reduce to the mean of the $k$ experts estimates.

The criticism against $\theta_{LW}$ is that it does not take account of the variability or the observed differences between the experts, nor does it take into account directly the estimate of $\theta$ provided by the sample data, meagre though it may be. These two criticisms are taken into account in the construction of the next estimate.

b. Combined Expert-Data Estimates

In a situation where $k$ experts are close to each other, it seems reasonable to place more weight or credence to the expert estimates. On the other hand if the experts differ widely from each other, we should put less weight on the expert estimates, and more on the estimate derived from the observed data. Let $\bar{\theta}$ denote the mean of the $k$ expert estimates,

$$\bar{\theta} = \frac{1}{k} \sum_{i=1}^{k} \theta_i$$

and denote by $\sigma^2_{\bar{\theta}}$ the variance of the $k$ estimates.

$$\sigma^2_{\bar{\theta}} = \frac{1}{k-1} \sum_{i=1}^{k} (\theta_i - \bar{\theta})^2.$$

Let $\hat{\theta}$ be the maximum likelihood estimate of $\theta$ obtained from the likelihood $L(\theta)$ i.e.,
\[ \hat{\theta} = \max_{\theta} L(\theta) \] .

The variance \( \hat{\sigma}^2 \) of \( \hat{\theta} \) is

\[
\hat{\sigma}^2 = \frac{1}{\theta} \cdot \frac{1}{\frac{d^2 \log L}{d\theta^2}}
\]

The combined expert-data estimator \( \theta_{ED} \) is defined by

\[
\theta_{ED} = \frac{\sigma_\hat{\theta}^2}{\sigma_\theta^2 + \sigma_\hat{\theta}^2} \hat{\theta} + \frac{\sigma_\hat{\theta}^2}{\sigma_\theta^2 + \sigma_\hat{\theta}^2} \bar{\theta}
\]

\( \theta_{ED} \) has some attractive properties. They are

i) gives more weight to expert opinions if there is more agreement between the experts;

ii) gives more weight to the data estimate as the sample size increases.

We illustrate these two estimates in a small example.

**Example.** Six experts are asked to estimate the proportion of people who have a certain condition in a large population. In a sample of size 8, it was found that 3 people had that specific condition. The data generating process can be thought of as a Bernoulli process. The table below shows the several experts estimates and the likelihood weights.
\[
\text{Expert} \quad \begin{array}{c}
\text{Estimate } (\theta_i) \\
1 \quad .10 \\
2 \quad .20 \\
3 \quad .25 \\
4 \quad .30 \\
5 \quad .35 \\
6 \quad .70 \\
\end{array} \\
\text{W}_i = \frac{\theta_i^3(1-\theta_i)^5}{\Sigma \theta_i^3(1-\theta_i)^5} \\
\begin{array}{c}
1 \quad .025 \\
2 \quad .109 \\
3 \quad .154 \\
4 \quad .188 \\
5 \quad .489 \\
6 \quad .035 \\
\end{array}
\]

\[
\bar{\theta} = \frac{1}{6} \Sigma \theta_i = .317.
\]

\[
\hat{\theta} = \frac{3}{8} = .375
\]

\[
\sigma^2_{\bar{\theta}} = \frac{1}{5} \Sigma (\theta_i - \bar{\theta})^2 = .0428.
\]

\[
\sigma^2_{\hat{\theta}} = \frac{\hat{\pi}(1-\hat{\pi})}{n} = .0293.
\]

\[
\theta_{\text{LW}} = \Sigma W_i \theta_i = 0.325.
\]

\[
\theta_{\text{ED}} = \left( \frac{0.0428}{0.0721} \right) \times .375 + \left( \frac{0.0293}{0.0721} \right) \times .317 = 0.351.
\]

Since there is considerable variation in the estimates by the experts, it is seen that \( \theta_{\text{ED}} \) puts more weight on the m.l.e. than on the average expert estimate. Similarly \( \theta_{\text{LW}} \) puts more weight on those expert estimates which are more consistent with the observed data, than those experts who differ widely from the observed value.
References


5. Dalkey, N. C. The Delphi Method: An Experimental Study of Group 

   Applications*. Addison Wesley, 1975.

7. Ford, D. A. Shang Inquiry as an Alternative to Delphi: Some 
   Experimental Findings. *Technological Forecasting and Social 
   Change* 7, 139-164, 1975.


11. Press, S. J. Qualitative Controlled Feedback for Forming Group 