Models of Consensus for Knowledge Acquisition

Daniel E. O'Leary

3660 Trousdale Pkwy
University of Southern California
Los Angeles, CA 90089-1421
213-740-4856
oleary@rcf.usc.edu

Abstract

Models of consensus are used in the knowledge acquisition process to determine what knowledge is built into a system. Unfortunately, in some cases, the consensus position is incorrect.

This paper develops two analytic models of consensus that can be useful in the knowledge acquisition process. The first employs the binomial model to study the probability that using the consensus judgment for knowledge acquisition is correct or incorrect. That basic model is extended to account for both different levels of expertise and unequal prior odds. Conditions are found to indicate when the consensus model should be used in knowledge acquisition.

The second is a Bayesian model of the use of consensus judgment and knowledge acquisition. The Bayesian approach also finds that, in some cases, consensus judgment is not correct. Conditions similar to those for the binomial model are found to be appropriate for determining when the probability of consensus is correct is greater than the probability that consensus is incorrect.

1. Introduction

The process of knowledge acquisition often involves multiple experts. In some cases, groups of experts have different opinions or solutions. In those situations, one approach to dealing with the different solutions is to choose the consensus position as the basis for the knowledge used in the system. Unfortunately, in some cases the consensus judgment is incorrect.

Thus, the purpose of this paper is to summarize and extend analytic models of consensus in order to study conditions under which consensus provides a higher probability of being correct than the non consensus position. Those models allow us to structure the use of consensus in knowledge acquisition, e.g., choosing the appropriate number of experts to assure a given probability of consensus being correct. As a result, this paper provides theoretic foundations of consensus, and a basis for the use of consensus in knowledge acquisition.

1.1 Consensus as a Basis of Knowledge Acquisition

Lack of consensus among a group of experts implies that some of the experts are not correct (or are employing a different model). However, even complete agreement among the experts does not guarantee a correct solution.

Unfortunately, there is little evidence on the relationship between consensus and correctness. Empirically, researchers have found that experts in some domains have been correct only 40% to 60% of the time (e.g., Sorenson et al. [1]). Thus, it is reasonable to assume that in some situations, consensus judgments will not be correct. Thus, there is interest in analyzing some of the characteristics of consensus judgment, e.g., under what conditions should we expect it to be used in knowledge acquisition.

1.2 Outline of This Paper

This paper proceeds as follows. Section 2 develops a basic model of the correctness of consensus judgments. That section summarizes some classic research as applied to the consensus problem. Section 3 and 4 investigates some extensions of that model, by relaxing assumptions inherent in that model. Some of the results of section 3 and 4 are new, such as the conditions for use of consensus in the situation of unequal prior odds. Section 5 studies when the consensus judgment is correct in the context of a Bayesian model. Section 6 reviews
some of the implications of these models and briefly discusses issues in their implementation. Section 7 provides a brief summary and some extensions.

2. A Constant Probability Model

Throughout this paper, it is assumed we are concerned with consensus about dichotomous decisions or recommendations, although, as noted later in the paper, these results can be extended. There are many real world situations meeting this description. For example, a loan system must decide whether loan applicants will default or not default, and whether to grant or not grant the loan.

The knowledge acquisition process is assumed to employ \( n \) experts, each with an equal probability of being correct. The probability of success is assumed constant for each knowledge acquisition judgment. The experts are assumed to arrive at their decisions independently. The expert's decisions are then summarized by some unbiased source to determine the consensus judgment.

2.1 Background and Assumptions

Condorcet [2] first recognized that Bernoulli's [3] work on the binomial could be used to model the probability of reaching correct decisions under different voting systems. Condorcet's [2] work become the basis of modern research in voting (e.g., Black [4]) and jury decision making (Gorfman and Owen [5]). One of the common themes of that research is to determine the probability that the consensus position is correct.

The binomial consists of \( n \) independent trials, where each dichotomous choice decision has a probability \( p \) of success and a probability \((1-p)\) of failure. In a knowledge acquisition setting

the use of the binomial would assume that each of the experts from whom knowledge is being solicited would have equal competence. In addition, it is also assumed that each of the two alternatives is equally likely to be correct. This assumption of equal prior odds creates a special case of the binomial, analogous to the case of using a fair coin. Each of these assumptions will be relaxed later in the paper.

### 2.2 A Model of the Consensus-Correctness Relationship

Let \( n \) be the number of experts in the knowledge acquisition process. Let \( M \) be the minimum number of experts necessary to establish a majority. When \( n \) is odd, \( M=(n+1)/2 \), when \( n \) is even \( M=(n/2)+1 \). Let \( m \) be the number of number of experts for a given consensus, where \( m = M, \ldots, n \). Let \( P_C \) be the probability that consensus judgment is correct.

Given the two assumptions from the previous section,

\[
P_C = \sum_{m=0}^{n} \binom{n}{m} p^m (1-p)^{n-m},
\]

where, \( \binom{n}{m} \) is \( n \) things taken \( m \) at a time. A set of binomial table values for \( P_C \) for some values of \( p \) and \( n \) is given in table 1.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Probability of Consensus Being Correct Assumes Equal Prior Odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>( p = .10 )</td>
</tr>
<tr>
<td>3</td>
<td>.028</td>
</tr>
<tr>
<td>5</td>
<td>.009</td>
</tr>
<tr>
<td>7</td>
<td>.003</td>
</tr>
</tbody>
</table>

### 2.2 Some Results from the Model

Condorcet [2] found a number of important results from the use of the binomial as a model of consensus. Assume that \( n \) is odd and \( n \geq 3 \).

**Result 1**
If \( p > .5 \) then \( P_C > p \).

**Result 2**
If \( p > .5 \) then \( P_C \) is monotonically increasing in \( n \) with a limit of 1.

**Result 3**
If \( p = .5 \) then \( P_C = .5 \).
Result 4
If p < .5 then P_C is monotonically decreasing in n with a limit of 0.

Result 5
If p < .5 then P_C < p.

Result 1 indicates that if p > .5 then the probability that the consensus decision is correct, is greater than the probability that a single decision is correct. In this situation, the consensus judgment should be used in the system.

Result 2 suggests that, if p > .5 then the more the number in the panel of experts evaluating the system, the higher that the probability of consensus is correct. Thus, if p > .5, it is beneficial to have as many experts as possible involved in the knowledge acquisition process.

Result 3 finds that in this specific case nothing is gained by going from individual judgments to consensus judgments. If the probability of all decision makers being correct is .5 then the probability that consensus of those decision makers is correct is also .5. In that case the individual judgment has the same probability of being correct.

Result 4 indicates that, if p < .5 then the more the number in the panel of experts evaluating the system, the lower that the probability of consensus is correct. In this case, we do not want a large number of experts involved, if we are planning on using the consensus decision.

Finally, Result 5 finds that if p < .5 then the probability that the consensus decision is correct, is less than the probability that a single decision is correct. In that situation, consensus actually results in a lower probability of correctness. In this case, the consensus judgment is not a good basis for knowledge acquisition.

3. Relaxation of Equal Competence Assumption

This section extends the model of the previous section by relaxing the equal competence assumption. It is reasonable to assume that different experts will have a different probability of providing the correct decision. For example, experts are often delineated as having different titles indicating gradation in expertise. Thus, it is reasonable to assume that the experts come from a number of different classes, where within each class, each expert is equally competent, yet there is an ordering of the competence of the different classes.

Assume there are two different groups of experts, A and B (this assumption could be extended to more than two groups). It is assumed that within either of those two groups the probability of correctness is equal. Let p_i be the probability that an individual expert in group i is correct, i = A or B. Assume that .5 < p_A ≤ 1 and that p_B < p_A. Let P_C(i) be the probability that a consensus decision of group i is correct, i = A, B or, A and B (written as A,B).

Margolis [6] examined the model with this revised assumption and developed the following three results.

Result 6
If p_B ≤ .5 then P_C(A,B) < P_C(A).

Result 7
If p_B > .5, then there exists some cardinality of group B, referred to as a critical value B*, such that P_C(A,B) > P_C(A).

Result 8
There exists some value p_B* < p_A, such that if p_B > p_B* then P_C(A,B) > P_C(A).

Result 6 indicates that if the value of p_B is low enough then it does not make sense to aggregate the experts of the two classes in order to develop a consensus value. Result 7 indicates that for p_B of an appropriate level, if group B is large enough then it makes sense to integrate the members into one large group of A and B, that will make the consensus decision. Result 8 indicates that if p_B is large enough then group B should be integrated with group A, regardless of the size of group B. These results are surprising to a certain extent, since they indicate that, in some situations, lower quality decision makers should be integrated in with higher quality decision makers for consensus judgments.

Result 7 may lead to the requirement that group B be quite large, so as to be impractical in the case of acquiring knowledge. If there are 30 members of A, p_A = .7 and p_B = .51 then B* would be several hundred, and thus beyond the scope of virtually all knowledge acquisition projects.

Using results from Margolis [6], the critical point nature of Result 8 can be exemplified as follows. If p_A = .9
then $p_B^* = .82$. If $p_A = .8$ then $p_B^* = .70$. If $p_A = .7$ then $p_B^* = .62$. If $p_A = .6$ then $p_B^* = .55$.

These results can be extended. For example, the following result indicates that if a subset of some set of experts is being used to develop a consensus judgment, then it is always better to add more of those same equal experts to the set of experts from which consensus is being developed.

Result 9
Let $A^*$ be a subset of $A$. $P_{C(A)} > P_{C(A^*)}$ for all $A^*$, not equal to $A$.

3.1 Limitation of Equal Class Behavior

The limitation of this approach is that it is assumed that within a class, all experts are equally likely to be correct. This limitation can be addressed by using some approximations to the binomial distribution.

3.2 Normal Approximation

The normal distribution can be used as an approximation to the binomial (Feller [7]). Using that approximation, an alternative approach has been developed by Grofman [1978] and Grofman et al. [8] that employs this result. Rather than multiple distinct sets of experts, they treat the set of experts as a single class, with competency normally distributed with a particular mean value of $p^*$ and a variance of $p^*(1-p^*)/n$. Thus, expert correctness has a distribution. In that case the conclusions of the equal competence model will hold, with $p^*$ substituting for $p$.

3.3 Poisson Approximation

The poisson distribution also can be used to approximate the binomial (Feller [7]), where the poisson is defined as $p(k;L) = e^{-L} \frac{L^k}{k!}$. In the same sense that the normal approximation to the binomial can be used to develop an alternative approach to the multiple classes, so can the poisson distribution. In the approximation of the poisson distribution, the parameter $L$ is equal to $n*(1-p)$. With $L$ specified as $n^*(1-p)$ the same results as in section 2 hold. Unlike the binomial, the only constraint on $L$, is that $L$ reflects the density of correct judgments in the group of experts.

4. Relaxation of the Equal Prior Odds Assumption

The model in section 2 also assumes that there are equal prior odds as to which of the alternatives is correct. This assumption is equivalent to the "fair coin" assumption that both a head and a tail are equally likely on each toss. However, in most decision making situations it is unlikely that the relevant states of nature are equally likely. For example, in the case of the prediction of bankruptcy, roughly 3% of the firms in the United States go bankrupt each year. If the choice is between predicting bankrupt or not bankrupt, then the prior odds are, respectively, .03 and .97.

Let $p_S$ be the probability of one state of the dichotomous decision occurring. Let $p_S' = (1 - p_S)$, be the probability of the other. In the case of equal prior odds, $p_S = p_S' = .5$. If the prior odds are not equal then there is no longer interest in $p$, instead the concern is with a revised probability that captures the difference in the prior odds. Let $p_R$ be the probability of the expert making the correct decision, given the prior odds for the state of nature $S$, assuming all experts are of equal competence. Let $p_R'$ be the probability of the expert choosing alternative $R'$, making the correct decision, given the prior odds for the state of nature $S'$ and assuming equal competence. Using Bayes' theorem, we have $p_R' = \frac{(p^*p_S' + (1-p)^*p_S)}{(p^*p_S' + (1-p)^*p_S)}$ and $p_R = \frac{(p^*p_S + (1-p)^*p_S')}{(p^*p_S + (1-p)^*p_S')}$ Some example values are given in table 2.

Table 2
Probability of an Individual Decision Being Correct for Equal Competencies When Prior Odds are Equal

<table>
<thead>
<tr>
<th>Prior Competency (p)</th>
<th>Odds</th>
<th>.10</th>
<th>.20</th>
<th>.30</th>
<th>.40</th>
<th>.50</th>
<th>.60</th>
<th>.70</th>
<th>.90</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.10</td>
<td>.012</td>
<td>.027</td>
<td>.045</td>
<td>.069</td>
<td>.100</td>
<td>.222</td>
<td>.300</td>
<td>.500</td>
</tr>
<tr>
<td></td>
<td>.20</td>
<td>.045</td>
<td>.097</td>
<td>.155</td>
<td>.222</td>
<td>.300</td>
<td>.500</td>
<td>.700</td>
<td>.900</td>
</tr>
<tr>
<td></td>
<td>.30</td>
<td>.100</td>
<td>.206</td>
<td>.368</td>
<td>.500</td>
<td>.692</td>
<td>.857</td>
<td>.900</td>
<td>.97</td>
</tr>
<tr>
<td></td>
<td>.40</td>
<td>.012</td>
<td>.045</td>
<td>.100</td>
<td>.206</td>
<td>.368</td>
<td>.500</td>
<td>.692</td>
<td>.857</td>
</tr>
<tr>
<td></td>
<td>.50</td>
<td>.012</td>
<td>.045</td>
<td>.100</td>
<td>.206</td>
<td>.368</td>
<td>.500</td>
<td>.700</td>
<td>.900</td>
</tr>
</tbody>
</table>
4.1 Relationship Between p and p_S, and p_R

There are a number of relationships between p, p_S and p_R that map the revised model into results obtained for the basic model, as discussed in results 1-5.

Result 10
If p + p_S > 1 then p_R > .5.

Proof of Result 10

\[ p_R = \frac{p*p_S}{2p*p_S + (1 - p - p_S)} \]

Since (1 - p - p_S) is less than 0, p_R > .5

Result 11
If p + p_S < 1 then p_R < .5.

Result 12
If p + p_S = 1 then p_R = .5.

Proof of Result 12

\[ p_R = \frac{p*p_S}{2p*p_S + (1 - p - p_S)} \]

Since p + p_S = 1, p_R = .5

How is the probability that a consensus judgment is correct impacted by the unequal prior odds? Results 1-5 when combined with results 10-12 provide us with the answer. We should use consensus only if p + p_S > 1. Thus, the quality of consensus judgments is a function of both those probabilities.

4.1 Monotonicity Result for Revised Model

In addition, we can establish a monotonicity result for p_R. In particular, the following result indicates that p_R is monotonically increasing as the prior odds increase.

Result 13
p_R is monotonically increasing in p_S.

Proof
Let p_S > p_S", then

\[ \frac{p*p_S}{p*p_S + (1-p)*(1-p_S)} > \frac{p*p_S"}{p*p_S" + (1-p)*(1-p_S")} \]

\[ \frac{p*p_S}{2p*p_S + 1 - p - p_S} > \frac{p*p_S"}{2p*p_S + 1 - p - p_S"} \]

p*p_S - p*p_S" > p*p_S" - p*p_S"

Since p_S > p_S", the inequality holds and p_R is monotonically increasing in p_S.

This can be a useful result. For example, we can make the following two statements. First, if we know p and have a conservative estimate of p_S, such that p + p_S > 1, then we know that we should use consensus. We do not need to know p_S exactly. We may be able to use simply a lower bound. Second, if the prior odds are greater than .5, we know that the simplified equal odds model underestimates p_R. Thus, in some cases the equal prior odds model helps bound the case where the prior odds are not equal.

5. A Bayesian Model of Correct Consensus Judgments

Sections 2, 3 and 4 have focused on single knowledge acquisition judgments. This section examines what happens when we have multiple knowledge acquisition judgments decided using consensus.
There are at least two cases to consider. If each of the judgments is independent, then the binomial can be used as a model, as noted earlier. Each decision can be treated separately. In the case of the basic binomial model, with equal prior odds and competence, if \( p > .5 \), \( P_C > .5 \), and the consensus solution should be used for each knowledge acquisition judgment.

However, that assumes independence of the individual consensus decisions. Instead, it is reasonable to assume that if the consensus decision is correct, then that would lead us to conclude that there is a higher probability that the next consensus judgment will be correct: success breeds success. This is consistent with a Bayesian approach where prior probabilities are updated with experience. If the consensus judgment is correct for knowledge acquisition judgment, then that might give us additional confidence that the consensus judgment is correct.

Thus, the model used in the remainder of this section is as follows. A knowledge acquisition judgment is made and consensus is used to determine if the knowledge should be included in the system. The chosen knowledge is either correct or not correct. Then another knowledge acquisition judgment is investigated, and the knowledge is either correct or not correct, etc.

The concern faced by the knowledge acquisition team is, what is the probability that the consensus judgment is correct? If the consensus judgment has a higher probability of correctness than then non consensus judgment, then we would expect the adoption of the consensus approach for knowledge acquisition purposes.

### 5.1 Prior Probability

Bayesian analysis (for the mathematical relationship see Raiffa and Schlaifer [9, pp. 50-51]) of a binomial often assumes that the decision maker’s prior probability distribution is distributed according a beta distribution with parameters \( k \) and \( n \). In this paper it is assumed that \( p \) represents the underlying probability that the consensus position is correct. In addition, it assumed that \( p \) is distributed according to a beta distribution, with \( k \) and \( n \) chosen to represent prior feelings and information about the knowledge acquired. The choice of the parameters \( k \) and \( n \) is important, since the expected value of \( p \) is \( k/n \).

As a result, assume that the prior distribution for \( p \) is 
\[
Pr(p|k,n) = \frac{[(n-1)!/(k-1)!(n-k-1)!]}{(n+k-1)!} p^{k-1}(1-p)^{n-k-1}, \quad \text{where} \quad 0 < p < 1,\ n,k > 0
\]

### 5.2 Probability Revision

There are some useful properties associated with treating the beta as the prior for \( p \). If there \( n\# \) knowledge acquisition judgments to be made and the consensus judgment is correct \( k\# \) times, then the parameters of the revised distribution, the posterior, are written as the sum of the parameters in the prior, plus these changes, resulting in the revised parameters \( n+n\# \) and \( k+k\# \). The expected value of the posterior is \( (k+k\#)/(n+n\#) \). Thus, \( p = k/n \), becomes \( p\# = (k+k\#)/(n+n\#) \). Thus, this distribution allows us to capture the success of consensus using the test data in updating the parameters of the distribution.

### 5.3 Example

Suppose that, in general, three out of five consensus position judgments provide correct knowledge acquisition judgments, i.e., \( k = 3 \) and \( n = 5 \). In that situation, for the first knowledge acquisition judgment, the probability that the consensus judgment is correct is \( 3/5 = .6 \) and the probability that the consensus judgment is incorrect is \( 2/5 = .4 \). In the case of a system with two knowledge acquisition judgments there are four branches in the decision tree. If the first consensus judgment is correct then the Bayesian model finds that the probability that the second consensus judgment is correct is \( 4/6 \) and the probability that the second consensus judgment is incorrect is \( 2/6 \). If the first consensus judgment is not correct, then the probability that the second consensus judgment is either correct or incorrect is \( 3/6 \). Note that even though the probability of consensus increases with each success, and the original probability of consensus being correct is greater than \(.5\), the probability of consensus being correct for the second consensus judgment is at \(.5\), complete uncertainty. This result is contrary to the finding for the basic binomial model.

### 5.4 General Case

The entire tree of possibilities of this Bayesian model would require infinite enumeration for any pair of \( k \) and \( n \), for all branches in any decision tree enumerating the probabilities [10]. In addition, if the decision tree is drawn to accommodate a large enough \( n\# \), there then will always be at least one branch in the decision tree where the probability of consensus being correct is less than the probability of consensus being incorrect. However, we can show the following result for
determining whether the probability of consensus being correct exceeds that of the probability of consensus being incorrect. An example is provided in table 3.

Table 3 -- Example

Probability that System Agrees with Consensus and Not Consensus, with j Cases

<table>
<thead>
<tr>
<th>k</th>
<th>n=2</th>
<th>n=3</th>
<th>n=4</th>
<th>n=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>j</td>
<td>j/2</td>
<td>j/3</td>
<td>j/4</td>
</tr>
<tr>
<td>2</td>
<td>j+1</td>
<td>(j+1)/2</td>
<td>(j+1)/3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(j+2)</td>
<td>(j+2)/2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Result 14:
The probability that consensus is correct is greater than or equal to the probability that consensus is incorrect at that branch if and only if (k + k#) ≥ (n + n#)/2 for that branch in the decision tree.

6. Implementation

This section discusses some of the implications of the models in this paper and their implementation.

6.1 Structuring Consensus for Knowledge Acquisition

The basic model and its extensions, discussed in sections 2 and 3, has a number of implications. First, the model indicates that the decision on whether or not consensus should be used is a function of the sum of two parameters: p and pS. Consensus should not be used indiscriminately. Second, in the consensus decision in the basic model, where p > .5, it is always beneficial for the use of a complete set of the best experts. If all the top experts cannot participate then it is likely that the next highest class of experts should also be used in the development of the consensus judgment. Third, the models imply some stopping criteria or design criteria for consensus analysis. For example, we can see from table 1, that if there are equal prior odds and if p = .70 and we wish a pC ≥ .8 then we must use at least 5 experts to develop the consensus judgment.

The findings of the Bayesian model have a similar set of implications. If the probability the consensus judgment is correct is greater than or equal to .5 then it is better to use the consensus judgment.

6.2 Some Implementation Considerations

In order to implement the models in this paper, basic knowledge of the underlying parameters is required. In the first model of consensus, the probabilities p and pS were necessary. The competency levels p could be difficult to obtain. Prior odds of events, pS could be obtained from experience. However, there is little in the literature about the quality of expertise in even broad categories of events. In a similar manner, the parameters k and n for the Bayesian model could be obtained using an analysis of empirical data or similar experimental approaches. Future research could address these issues.

7. Summary and Extensions

This paper has developed two basic models that can be useful in the knowledge acquisition process. The first model was based on the binomial, but was extended to include multiple levels of expertise and unequal prior odds. The results presented here summarized some classic results and presented new results.

The second model was a Bayesian model, designed to study the correctness of the consensus judgment and the knowledge acquisition process. That model was important since it investigated a situation where the probability that consensus judgment was correct changed with different successes and failures of group consensus performance.

7.1 General Extensions

Each of the two models assumed a dichotomous decision. Thus, rather than the basic binomial models presented in this paper, more general multinomial models could be developed.

The models presented in this paper ignore the impact of multiple systems. In effect, each system is treated separately. Thus, another extension would account for such portfolio effects.

7.2 Extensions of the Basic Binomial Model

The basic model was limited to simple majorities as the means of the definition of consensus. Alternative approaches used by other organizations may include a two-thirds majority. These alternative definitions of
consensus could be accounted for in the model developed above.

7.3 Extensions of the Bayesian Model

In the development of the Bayesian model, a beta prior was assumed. Other alternative prior distributions could have been developed. For example, Raiffa and Schlaifer [9] discuss a normal distribution with similar updating properties to the beta distribution. Other extensions might also be generated for the Bayesian model [11].

References


