Memorandum COSOR 91-11

On some misconceptions about subjective probability and Bayesian inference

F.P.A. Coolen
ON SOME MISCONCEPTIONS ABOUT SUBJECTIVE PROBABILITY
AND BAYESIAN INFERENCE

Frank Coolen

abstract
Some misconceptions about subjective probability and Bayesian inference exist, that can be obstacles for acceptance and use of these concepts in practice. The statement that 'humans are not natural Bayesian personalists' is discussed, together with a formal mistake inherent to preposterior analysis. The most important problem that has to be solved for practical acceptance is correct measurement of subjective probability, where the random variable of interest must be easily interpretable.

1. Introduction
In many practical decision problems where uncertainty is involved one has to rely on human knowledge. To make communication possible the level of uncertainty needs to be quantified. It has been shown (see e.g. French, 1986, chapter 6) that logical assumptions of rational behaviour lead directly to the use of the concept of probability to quantify uncertainty. With 'concept of probability' we mean acceptance of the axioms of Kolmogoroff (the discussion of countable or finite additivity, as suggested by DeFinetti (1970), is beyond the scope of this note). We state that the correct and consistent manner to deal with probability is by accepting the fact that probability is subjective by its very nature (DeFinetti, 1970). It is important to realize that the subjective nature of probability does not lead to rejection of the mathematics of probability and statistics. Realization of the nature of probability is necessary in practice, before any mathematical treatment is necessary.
Consider \( P(A) \), the probability of the truth of some proposition \( A \), and \( P(B) \) analogously (both positive). Let \( P(A \land B) \) be the probability of both
propositions being true, and \( P(A|B) \) the probability of the truth of \( A \), given \( B \) is true (\( P(B|A) \) analogously).

In all systems of probability Bayes' theorem holds, that is
\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)},
\]
where the probabilities can be personal (subjective). Thus, for a rational decision maker (see French (1986) for an extensive description of rationality) his or her personal probabilities also obey Bayes' theorem.

We want to examine the statement "humans are not natural Bayesian personalists" (Callen, 1991) in the light of these observations.

In the above form of Bayes' theorem, \( P(A) \) is the prior and \( P(A|B) \) the posterior probability, resulting from this theorem. The likelihood usually plays the role of \( P(B|A) \). The probability that quantifies someone's subjective beliefs after learning that \( B \) is true will be denoted by \( \tilde{P}(A|B) \). In many situations these two probabilities differ, also when a parametric model is used to give a likelihood.

We want to look at the problems associated with the practical use of Bayesian methods, where there are two sources of difficulty, one the prior \( P(A) \), and the other the model and as a consequence \( P(B|A) \).

We also examine the statement (Kaye and Koehler, 1991): "Although Bayes' theorem provides a normative rule for revising probabilistic beliefs in the light of new evidence, many studies indicate that within certain domains people do not extract as much information from new evidence as the data warrant, that they are slow to revise incorrect probabilistic hypotheses, that they attribute probative value to diagnostically worthless information, that they underutilize statistical base rates and that they confuse likelihoods and posteriors".

We indicate how these comments should not be used to reject the practical use of Bayes' theorem, but indicate problems in the area of probability measurement and the correct interpretation of data. Thus the problems lie in the area of elicitation, problem presentation and training.

Lastly we discuss preposterior analysis (Martz and Waller, 1982, par. 5.3.2), a technique that must be rejected on formal grounds. The fact that this has not been remarked by Martz and Waller shows once more the existence of misconceptions.
Lindley (1991) states that "there is a forceful argument, that is being increasingly accepted, which concludes that the only sensible way to handle uncertainty is by means of probability". However, to enable practical acceptance and use of probability within decision problems, suitable methods for the measurement of probability must be developed. This is an aspect that has long been neglected by mathematicians. While the methods are mathematically consistent, they are difficult to apply correctly in practice because people have to state their uncertainty about the parameters of a model. Unfortunately, model parameters are frequently not directly observable or measurable, and thus have little meaning for the decision maker. This shortcoming is seen in the elaborate construction of conjugate priors, the forms of which are determined by the model. The fact that people may not understand the meaning of parameters, implies that they can not give accurate probability statements about them. Recent books on Bayesian statistics, such as Lee (1989) and Press (1989) pay little attention to practical methods for the measurement of probability (in these books it is also assumed that a person can state probabilities for the parameters of a model). This drawback has also been mentioned by Wooff (1991).

DeFinetti (1970) devoted a chapter (ch. 5) to the evaluation of probability, and Lindley (1991) also discussed the measurement of probability (section 3). O'Hagan (1988) makes a step in the right direction by discussing measurement of probability in an introduction to the entire concept, but there is still much to be done to make the concept really useful in practice.

We discuss the reasons why \( P(A|B) \) and \( \tilde{P}(A|B) \) might differ. Firstly, we look at the prior distribution, and secondly we study the likelihood. There are many more possible reasons, but we restrict ourselves to those that cause the most important misconceptions.
2. Prior Distribution

Suppose we are interested in determining the uncertainty of someone about a proposition A, for example because we need information about A to make a decision and we regard this person to be an expert. Assuming rational behaviour, the expert’s subjective beliefs about A are expressed as a prior probability P(A). It is obvious that P(A) should be measured by asking the person to provide some statements about A, because this is the problem under discussion and relates to the expert’s knowledge. However, if a parametric model is assumed that describes the probability of B, and A is a proposition about the parameter of this model (so the model provides the likelihood P(B|A)), then it can be difficult for the expert to express beliefs about A, because the parametric model is not the usual way of handling the subject. This approach places unreasonable demands on the expert and, moreover, the demands may be impossible if the parameter cannot be interpreted within the expert’s frame of reference. Although the practical problem in such an approach is obvious, this is exactly the way most mathematicians expect prior probabilities to be expressed. For the expert under consideration this approach will certainly not encourage the further use of probability, because of a lack of understanding. This is a serious impediment to the practical acceptance of probability when a decision has to be taken under uncertainty.

The obvious way to avoid this obstacle is by analysing first which aspects of uncertainty can be discussed with experts, such that they can think of these aspects within their usual reference frame. Asking questions about model parameters is, therefore, frequently unacceptable. How to measure a prior remains an open question. It is clear that methods must be developed that are usable in practice.

3. Likelihood

Although the achievement of a prior distribution might lead to problems, the misconceptions mentioned in the introduction mostly result from problems with the likelihood. We will discuss these
problems firstly when a parametric model is assumed, and secondly when a parametric model is not assumed.

Suppose that we assume a parametric model to describe the probability of $B$, with $A$ a proposition about the parameter.

Suppose we have the expert’s uncertainty about $A$, represented as a prior $P(A)$. We ignore for the moment the difficulty raised in section 2. New evidence (additional knowledge about proposition $B$) will modify the expert’s uncertainty about $A$.

In Bayes’ theorem the likelihood is based on the assumed model, which, however, may not be the model for the relationship between $A$ and $B$ that the expert has in mind. In this case the posterior from Bayes’ theorem, $P(A|B)$, will therefore not be identical to the expert’s probability $\hat{P}(A|B)$, for the parameters of the model in the light of new evidence. That is to say $P(A|B)$ and $\hat{P}(A|B)$ are based on different models and therefore it is not surprising that they do not agree. It should be understood that for one person a Bayesian posterior is not really subjective, if a parametric model is assumed (unless this person relates $A$ and $B$ exactly according to the model, which is unlikely).

At this point preposterior analysis, as presented e.g. by Martz and Waller (1982), must be rejected on formal grounds. To summarize, they let a person state prior beliefs about a parameter, and then use hypothetical data $B^*$ to determine the ‘would-be-posterior’ $P(A|B^*)$. The person is asked if he would agree, or, if this posterior would really present his beliefs if the hypothetical data would be gathered as new evidence. If not, and this is where a fatal mistake is made, the person is asked to change his prior.

Realizing what probability is (the real beliefs of a person), it must be clear that it is not the prior that must be changed (formally this is an absurd treatment of probability), but disagreement of $P(A|B^*)$ and $\hat{P}(A|B^*)$ is due either to the fact that the person does not understand the total concept of probability well enough, or to the fact that the chosen parametrical model is not in agreement with the model with which the person thinks about the relation between the uncertain proposition (prior) and the new evidence (likelihood).

If a technique as preposterior analysis is to be used at all, it can
only be used to explore the assumed parametric model. Here we must emphasize the difficulties arising from the ignorance of probability most people have, resulting from the fact that they are not taught how to use it, and, in particular, the use of conditional probabilities.

A more logical approach would be to ignore parametric models, and ask the expert who provides the prior on a proposition $A$ also the likelihood $P(B|A)$ (however, in practice this may be impossible because of a very large or even infinite number of possible propositions on $A$ of interest). In this case $P(B|A)$ is also a subjective probability. Here the possible problems have a more psychological nature than above. Because the likelihood is also the probability of the expert, after real new evidence $B$ has come available $P(A|B)$ should be exactly equal to $P(A|B^*)$, if the hypothetical data $B^*$ is exactly the same as $B$, and no other factors of interest have changed. As shown by Kaye and Koehler (1991), this equality does not occur in practice. This may not be considered as an argument against practical use of the concept of probability (including Bayes’ theorem), but must be used to give direction to future research on the subject.

The observed differences, described by Kaye and Koehler (1991) can result from several sources. The most obvious are:

- People are not used to express their beliefs as probabilities.
- People actually confuse probabilities such as $P(B|A)$, $P(A|B)$ and $P(AnB)$.
- One person’s likelihood $P(B^*|A)$ might not be equal to $P(B|A)$ after the evidence has really come available, even if $B^*$ is equal to $B$. It is possible that also other factors have provided new evidence together with $B$, while these factors where not considered within the hypothetical analysis.

This list of sources that might lead to the misconceptions mentioned in the introduction, could without doubt be extended. The three sources stated here lead to the conclusion mentioned by O’Hagan (1988): People should be taught to express their uncertainty by means of probabilities. Methods to enable this need a lot of research, especially by actually working with people in practice. Here
mathematicians should cooperate with researchers from other disciplines, e.g. psychologists. Methods for measurement of probability must be developed that can really be used, and that do not discourage the use of probability.

As Lindley (1991) states: 'How is probability to be measured? In the same way that anything is measured: by comparison with a standard'. Attention to this point might lead to more acceptance of the concept of probability in future, where, of course, sound theories must be the base for all methods.

4. Conclusion

We have discussed some misconceptions about subjective probability and Bayesian inference. These discourage the use of probability and need to be solved.

When assuming rational behaviour, the use of probability to quantify uncertainty cannot be avoided. Researchers should, at this point, first study what probability really is (personal degrees of belief), for example by studying DeFinetti's work. It is not possible to work on this subject without sound knowledge of the meaning and interpretation of probability itself.

After one has correct knowledge of the total concept, one should try to find suitable methods to apply the concept in practice. The main points for consideration are:

- Let people state their uncertainty only about propositions concerning subjects or variables they understand, because they are directly observable or measurable. Parameters are almost never to be used in direct conversation.

- People should be taught to quantify their uncertainty by means of probability. Measuring probability must be done very accurately, where for example the notion of conditional probability must be understood entirely.

- Mathematicians should not go on using mathematical models and techniques that cannot be interpreted by persons who are responsible for decisions to be taken, or by experts who have to provide
information needed for these decisions (by means of probability).
- Preposterior analysis may not lead to changes of prior probabilities, because this is fundamentally wrong. If it leads to rejecting a certain likelihood this would be correct, but this would probably lead to practical problems. However, it is better to try to solve new problems correctly then to use wrong methods.
- If \( P(A|B) \) and \( \bar{P}(A|B) \) are not equal, this should not lead to rejection of the use of Bayesian methods in practice, but to proper consideration of what went wrong in applying the theorem.

Finally, we must be cautious about expressions like 'humans are not natural Bayesian personalists'. Such are indeed rash generalisations based on the symptoms and not supported by a proper consideration of the nature of the problem. Such superficial analysis cannot indicate the correct direction for future research in the application of Bayesian methods.

References


<table>
<thead>
<tr>
<th>Number</th>
<th>Month</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>91-01</td>
<td>January</td>
<td>M.W.I. van Kraaij, W.Z. Venema, J. Wessels</td>
<td>The construction of a strategy for manpower planning problems.</td>
</tr>
<tr>
<td>91-03</td>
<td>January</td>
<td>M.W.P. Savelsbergh</td>
<td>The vehicle routing problem with time windows: minimizing route duration.</td>
</tr>
<tr>
<td>91-04</td>
<td>January</td>
<td>M.W.I. van Kraaij</td>
<td>Some considerations concerning the problem interpreter of the new manpower planning system formasy.</td>
</tr>
<tr>
<td>91-06</td>
<td>March</td>
<td>R.J.G. Wilms</td>
<td>Properties of Fourier-Stieltjes sequences of distribution with support in [0, 1).</td>
</tr>
<tr>
<td>91-07</td>
<td>March</td>
<td>F. Coolen, R. Dekker, A. Smit</td>
<td>Analysis of a two-phase inspection model with competing risks.</td>
</tr>
<tr>
<td>91-08</td>
<td>April</td>
<td>P.J. Zwietering, E.H.L. Aarts, J. Wessels</td>
<td>The Design and Complexity of Exact Multi-Layered Perceptrons.</td>
</tr>
<tr>
<td>91-09</td>
<td>May</td>
<td>P.J. Zwietering, E.H.L. Aarts, J. Wessels</td>
<td>The Classification capabilities of Exact Two-Layered Perceptrons.</td>
</tr>
<tr>
<td>91-10</td>
<td>May</td>
<td>P.J. Zwietering, E.H.L. Aarts, J. Wessels</td>
<td>Sorting With A Neural Net.</td>
</tr>
<tr>
<td>91-11</td>
<td>May</td>
<td>F. Coolen</td>
<td>On some misconceptions about subjective probability and Bayesian inference.</td>
</tr>
</tbody>
</table>