

On the Robustness of Imprecise Probability Methods

Marco E. G. V. Cattaneo
Department of Statistics, LMU Munich
cattaneo@stat.uni-muenchen.de

Abstract

Imprecise probability methods are often claimed to be robust, or more robust than conventional methods. In particular, the higher robustness of the resulting methods seems to be the principal argument supporting the imprecise probability approach to statistics over the Bayesian one. The goal of the present paper is to investigate the robustness of imprecise probability methods, and in particular to clarify the terminology used to describe this fundamental issue of the imprecise probability approach.

Keywords. Robustness, imprecise probabilities, Bayesian analysis, credibility, decision making, indecision, sensitivity analysis, imprecise Dirichlet model.

1 Introduction

The theories of imprecise probability replace probability measures by more general mathematical objects, which can often be identified with particular sets of probability measures. Such sets appear naturally also in Bayesian sensitivity analysis (also called robust Bayesian analysis) [6, 27] and robust statistics [4, 20]. Hence, there is a strong connection between imprecise probability and robustness. In fact, methods resulting from the imprecise probability approaches to inference and decision making are often claimed to be “robust” (or “more robust” than alternative methods) [1, 14, 36], usually without specifying the meaning of “robust”. The goal of the present paper is to investigate the robustness of imprecise probability methods. We will focus in particular on the most developed theory of imprecise probability: the theory of lower and upper previsions [33, 35].

The question of the robustness of imprecise probability methods is particularly important in statistics, where the imprecise probability approach can be seen as an alternative to the Bayesian approach. In fact, when comparing these two approaches to statistics,

the latter has clear advantages in terms of technical and conceptual simplicity [12, 13], also thanks to important invariances [3, 18, 21]. On the other hand, the (higher) robustness of the resulting methods seems to be one of the few general advantages claimed by the proponents of the imprecise probability approach. That is, the alleged (higher) robustness of the imprecise probability methods seems to be the principal argument for preferring the imprecise probability approach to statistics over the Bayesian one.

The present paper examines various aspects of the question of the robustness of imprecise probability methods, and in particular tries to clarify the terminology used to describe this fundamental issue of the imprecise probability approach. The paper is organized as follows. In the next section the concept of robustness is introduced. The robustness of imprecise probability methods is then investigated in Section 3, which is the core of the paper. In particular, in Subsection 3.1 the higher credibility of imprecise probability analyses over Bayesian analyses is discussed. These two kinds of analyses are then compared with regard to decision making: Subsection 3.2 considers the case when a decision has to be made, while the case when indecision is allowed is studied in Subsection 3.3. The final section summarizes the results.

2 Robustness

Robustness means “insensitivity to small deviations from the assumptions” [19, p. 2]. In the Bayesian approach to inference and decision making it mainly refers to “possible misspecification of the prior distribution” [7, p. 195]. Hence, the conclusions of a Bayesian analysis are not robust if there are several reasonable choices for the prior distribution and the conclusions depend on which prior is actually chosen, as in the following example.

Example 1 *In the Bayesian framework, given an exchangeable sequence of Bernoulli random variables*

X_1, X_2, \dots , de Finetti's theorem [16, § 11.4] implies that they are independent and $\text{Ber}(\theta)$ -distributed conditional on the success probability $\theta \in [0, 1]$. That is, to complete the Bayesian model we must choose a (prior) probability distribution for θ . Suppose that we have (almost) no prior information about θ : several prior probability distributions have been suggested in this situation. In particular, Bayes [5] and Jeffreys [24] proposed the prior uniform distribution of θ on $[0, 1]$ and of $\arcsin \sqrt{\theta}$ on $[0, \pi/2]$, respectively. Using Walley's (s, t) -parametrization of the beta distribution [33, 34], these two proposals correspond to the priors $\theta \sim \text{Beta}(2, 1/2)$ and $\theta \sim \text{Beta}(1, 1/2)$, respectively.

Assume now that we observe $X_1 + \dots + X_7 = 6$. That is, of the first seven Bernoulli trials, six were successes and one was a failure. In general, on the basis of these data, the conjugate prior distribution $\text{Beta}(s, t)$ is updated to the posterior distribution

$$\text{Beta}\left(s + 7, \frac{s + t + 6}{s + 7}\right). \quad (1)$$

In particular, Bayes' and Jeffreys' priors are updated to the posteriors $\theta \sim \text{Beta}(9, 7/9)$ and $\theta \sim \text{Beta}(8, 13/16)$, respectively.

Finally, suppose that we must choose between two courses of action with uncertain payoffs $A = 5X_8 - 4$ and $B = 4 - 5X_8$, respectively, expressed in a linear utility scale. This can be interpreted as choosing the side of a bet with odds of 4 to 1 on a success in the next Bernoulli trial, where the total stake is a fixed small amount of money. In general, the conjugate prior distribution $\text{Beta}(s, t)$ leads to the posterior expected utilities

$$E(A) = \frac{s}{s+7}(5t-4) + \frac{7}{s+7}\left(5\frac{6}{7}-4\right) \quad (2)$$

and $E(B) = -E(A)$. These are plotted in Figure 1 as functions of $s \in (0, 3]$, in the case $t = 1/2$ and in the limit cases $t \rightarrow 1$ and $t \rightarrow 0$. In particular, Jeffreys' prior would lead to the choice of the first course of action (that is, betting on success), since $E(A) > E(B)$ when $(s, t) = (1, 1/2)$, while Bayes' prior would lead to the choice of the second course of action (that is, betting on failure), since $E(B) > E(A)$ when $(s, t) = (2, 1/2)$.

Therefore, in this situation the decision resulting from the Bayesian approach is not robust, if both Bayes' and Jeffreys' priors are considered as reasonable choices in the case of (almost) no prior information about θ . The Bayesian answer to this non-robustness issue would be to give more careful consideration to the prior information about θ , in order to be able to identify more precisely the prior probability distribution for θ .

Exactly as for the Bayesian approach, the conclusions resulting from the imprecise probability approach to inference and decision making are robust if they are not too sensitive to small deviations from the assumptions in general, and to possible misspecification of the prior (imprecise) probability distribution in particular. More precise definitions of robustness would be possible, but would have a high degree of arbitrariness, while the above informal definition is sufficient for the scope of the present paper.

3 Imprecise Probability Methods

The robustness of some kinds of conclusions resulting from an imprecise probability analysis has been studied in [32], with comforting results. However, this study did not consider the robustness of the conclusions when the imprecise probabilities have been updated in the light of new data. In this situation, which is obviously very important for the imprecise probability approach to statistics, the conclusions resulting from an imprecise probability analysis are in general not robust (and not more robust than the ones resulting from a Bayesian analysis), as shown in the following example.

Example 2 Let X be a random variable taking value in the set $\{1, 2, 3\}$. Assume that our prior imprecise probabilities are determined by the unique assessment $\underline{P}(X) = x$, where $x \in [1, 3]$ is a real number. Suppose now that we learn that the value of X is not 2. That is, we observe the event $X \in \{1, 3\}$. If we update our prior imprecise probabilities by regular extension [33, Appx. J], then the posterior lower prevision of X is

$$\underline{P}(X) = \begin{cases} 1 & \text{if } x < 2, \\ x & \text{if } x \geq 2, \end{cases} \quad (3)$$

while if we update them by natural extension, then it is

$$\underline{P}(X) = \begin{cases} 1 & \text{if } x \leq 2, \\ x & \text{if } x > 2, \end{cases} \quad (4)$$

since the prior lower probability of the observed event is 0 if and only if $x \leq 2$. In both cases (3) and (4), the posterior lower prevision of X , as a function of $x \in [1, 3]$, has a discontinuity at $x = 2$.

Therefore, the posterior lower prevision of X is not robust, if for example both values $x = 1.99$ and $x = 2.01$ are considered as reasonable choices for the prior lower prevision. By contrast, in a Bayesian analysis of this situation, the posterior expectation of X would be a continuous function of the prior probability values, although it would be very sensitive to these values if the prior probability of the observed event were very small. Anyway, in this situation the posterior

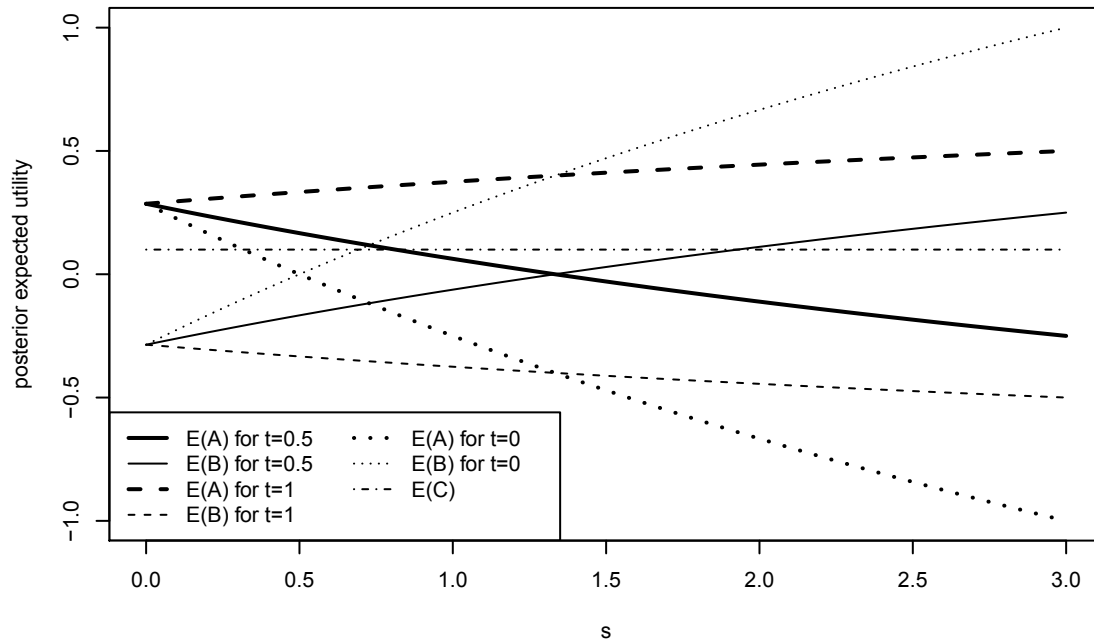


Figure 1: Expected utilities according to the posterior distribution (1), as functions of $s \in (0, 3]$, in the case $t = 1/2$ and in the limit cases $t \rightarrow 1$ and $t \rightarrow 0$.

distribution of the imprecise probability analysis is in general not more robust than the one of a Bayesian analysis.

However, the situation analyzed in Example 2 is artificial, and consequently its importance for the imprecise probability methods suggested in the literature is not clear. For this reason, in the remainder of the present section we shall consider further the situation of Example 1, focusing on the imprecise probability model that seems to be by far the most studied and used: the imprecise Dirichlet model [8, 34], in the special case of Bernoulli random variables [33, § 5.3].

The imprecise Dirichlet model satisfies some important invariance properties, and in particular the representation invariance principle [34]. This principle describes a particular kind of robustness with respect to assumptions about the statistical model, and it cannot be satisfied by objective Bayesian analyses. However, it can be satisfied by subjective Bayesian analyses, and its appropriateness is questionable anyway [34, p. 52]. On the other hand, the imprecise Dirichlet model is highly non-robust with respect to other aspects of the statistical model [28, 29]. Therefore, to keep things simple, in the remainder of this section we shall consider only the robustness with respect to the choice of the prior distribution.

3.1 Credibility

From the standpoint of the theory of lower and upper previsions, a Bayesian analysis corresponds to the special case of an imprecise probability analysis in which we have so much prior information that the previsions are linear. Hence, from this standpoint, a lower prevision can be interpreted as being based on less information (or assumptions) than a linear prevision dominating it. In this case, the Bayesian analysis can thus be considered as less credible than the imprecise probability analysis, according to a “law of decreasing credibility” [26, p. 1], stating that the credibility of the conclusions decreases when additional assumptions are made.

Such a law seems reasonable when inferences such as confidence or credible regions are considered as conclusions, but it does not necessarily seem reasonable when decisions or point estimates are considered. Anyway, for the sake of argument, let’s agree that imprecise probability analyses are more credible than Bayesian analyses (when the linear previsions dominate the lower previsions). Does this imply that they are also more robust?

Example 3 *In the imprecise probability framework, given an exchangeable sequence of Bernoulli random variables X_1, X_2, \dots , a generalization of de Finetti’s theorem [15] implies that they are independent and*

Ber(θ)-distributed conditional on the success probability $\theta \in [0, 1]$. That is, to complete the imprecise probability model we must choose a (prior) imprecise probability distribution for θ . The usual choice of the prior imprecise probability distribution in the case of (almost) no prior information about θ is the imprecise Dirichlet model, which corresponds to the set of all *Beta*(s, t) distributions with $t \in (0, 1)$. That is, the parameter s must still be chosen: the most popular choices appear to be $s = 2$ and $s = 1$ [8, 34, 36]. In this context, it is important to note that the imprecise previsions resulting from different choices of s are nested, the more imprecise corresponding to the larger values of s .

When observing $X_1 + \dots + X_7 = 6$, the imprecise Dirichlet model is updated by regular extension to the posterior imprecise probability distribution corresponding to the set of all distributions (1) with $t \in (0, 1)$. The posterior lower and upper previsions, $\underline{P}(A)$ and $\overline{P}(A)$, of the utility of the first course of action described in Example 1 are the limits of (2) as $t \rightarrow 0$ and as $t \rightarrow 1$, respectively. By contrast, the posterior lower and upper previsions, $\underline{P}(B) = -\overline{P}(A)$ and $\overline{P}(B) = -\underline{P}(A)$, of the utility of the second course of action are the limits of $E(B) = -E(A)$ as $t \rightarrow 1$ and as $t \rightarrow 0$, respectively. These two pairs of posterior lower and upper previsions are plotted in Figure 1 as functions of $s \in (0, 3]$.

The posterior imprecise previsions with $s = 1$ are thus more credible (in the sense considered above) than the posterior expectations resulting from Jeffreys' prior, and the posterior imprecise previsions with $s = 2$ are more credible than the posterior expectations resulting from both Bayes' and Jeffreys' priors. However, it is not clear why these posterior imprecise previsions should be more robust than the posterior expectations of Example 1, since they too depend strongly on the choice of s .

The question of the alleged higher robustness of imprecise probability analyses compared to Bayesian analyses can perhaps be better clarified by considering the choice of a probability distribution as consisting of two steps. First we choose a lower prevision \underline{P} , and then we select a linear prevision P dominating it. The second step can be seen as an additional assumption, and therefore the imprecise probability analysis based on \underline{P} is more credible than the Bayesian analysis based on P . Moreover, since there is certainly some arbitrariness in the second step, the imprecise probability analysis can appear to be more robust than the Bayesian analysis. However, once P has been selected, it does not depend on the choice of \underline{P} anymore. That is, the robustness of the imprecise probability analysis is relative to the arbitrariness in

the choice of \underline{P} , while the robustness of the Bayesian analysis is relative to the arbitrariness in the choice of P (and not in both choices of \underline{P} and P). So it is not clear that in general the Bayesian analysis is less robust than the imprecise probability analysis, even when the latter is more credible (in the above sense).

Of course, the imprecise probability analysis would be more robust than the Bayesian analysis, if there were no arbitrariness in the choice of the lower prevision. In this case, "conclusions drawn from the imprecise model are automatically robust, because they do not rely on arbitrary or doubtful assumptions" [33, p. 5]. Unfortunately, this is never the case, because there is always some arbitrariness in the choice of a model, even when we choose the vacuous model. In fact, if the vacuous prevision is a reasonable choice, then probably also a slightly more determined imprecise prevision would be reasonable.

In particular, the choice of the prior distribution in the imprecise probability analysis of Example 3 does not seem to be less arbitrary than the choice of the prior distribution in the Bayesian analysis of Example 1. In fact, thanks to symmetry arguments, in the Bayesian analysis the choice of $t = 1/2$ is less problematic than the choice of s , which must be chosen also in the imprecise probability analysis. In analogy to the discussion above, we could see the choice of the prior probability distribution in Example 1 as consisting of two steps. First we choose to restrict attention to the beta distributions and we select the value of s , while in a second step we also choose the value of t . With this description, it appears that the imprecise Dirichlet model (corresponding to the choices in the first step) has one assumption less than the Bayesian beta prior (the assumption of a particular value for t). However, this appearance is misleading, because in the imprecise Dirichlet model we also make a choice about t : we choose to let it vary in the whole interval $(0, 1)$. In fact, replacing this interval for instance with the interval $[\varepsilon, 1 - \varepsilon]$, for some small positive ε , could also be a reasonable choice [11].

An important difference between the choices of s in Examples 1 and 3 is that in the latter case the imprecise previsions resulting from different values of s are nested, and this could make the choice "less crucial" than in the former case [34, p. 12]. The importance of this property of the imprecise Dirichlet model for the question of the robustness of the imprecise probability analysis of Example 3 depends on how the imprecise previsions are used. Therefore, in the following subsections we shall consider the decision problem of Example 1 in the imprecise probability framework of Example 3.

3.2 Decision

Several decision criteria have been suggested in the literature on imprecise probabilities [2, 17, 31]. Some of these criteria, like Γ -maximin, induce a total preorder on the possible decisions, and usually identify a single optimal decision. When such criteria are used in an imprecise probability analysis, the resulting conclusions are in general not more robust than those resulting from a Bayesian analysis, as shown in the following example.

Example 4 Consider the decision problem of Example 1 in the imprecise probability framework of Example 3. In particular, Figure 1 shows that $\underline{P}(A) > \underline{P}(B)$ when $s = 1$, while $\underline{P}(B) > \underline{P}(A)$ when $s = 2$. Hence, the Γ -maximin decision would correspond to the first course of action (that is, betting on success) when $s = 1$, and to the second course of action (that is, betting on failure) when $s = 2$. We would obtain the same decisions if we used the Γ -maximax, Hurwicz [2, 22], or interval bound dominance [17] criteria instead of Γ -maximin.

Therefore, in this situation the decision resulting from the imprecise probability approach is not robust, if one of these criteria is used and both $s = 1$ and $s = 2$ are considered as reasonable choices for the parameter s of the imprecise Dirichlet model in the case of (almost) no prior information about θ . In complete analogy with the Bayesian analysis of Example 1, an answer to this non-robustness issue would be to give more careful consideration to the prior information about θ , in order to be able to identify more precisely the prior imprecise probability distribution for θ .

Other decision criteria, like maximality, E-admissibility, or interval dominance, often do not identify a unique optimal decision, and are perhaps more in keeping with the spirit of imprecise probabilities. When such criteria are used, imprecise probability analyses can be seen as descriptions of the robustness or non-robustness of Bayesian analyses. In fact, if one of these criteria identifies a single optimal decision in an imprecise probability analysis based on a lower prevision \underline{P} , then this decision is the unique optimal one in each Bayesian analysis based on a linear prevision P dominating \underline{P} (assuming that in these Bayesian analyses there are optimal decisions). By contrast, the two approaches diverge when the Bayesian analysis is not robust, in the sense that different linear previsions P dominating \underline{P} lead to different optimal decisions. In this case, all these decisions are optimal in the imprecise probability analysis based on \underline{P} , when one of the above criteria is used. However, this situation has very different meanings for the two approaches to decision making. In the Bayes-

ian approach the non-robustness issue can be tackled by identifying more precisely the linear prevision P , while in the imprecise probability approach there is not necessarily a more precise lower prevision \underline{P} that would still be a reasonable choice.

Therefore, since the goal of decision making is to select one of the possible decisions, in the imprecise probability approach we often still have to choose one of the optimal decisions, when one of the above criteria is used. This choice can be based on a second decision criterion selected among the ones usually identifying a single optimal decision, like Γ -maximin [25]. However, when such two-stage decision procedures are used in an imprecise probability analysis, the resulting conclusions are in general not more robust than those resulting from a Bayesian analysis, as shown in the following example.

Example 5 Figure 1 shows that in the decision problem of Example 1, when $s = 1$ we have $E(A) > E(B)$ if $t \in (0, 1)$ is sufficiently large, and $E(B) > E(A)$ if $t \in (0, 1)$ is sufficiently small. That is, the decision resulting from the Bayesian approach is not robust, if all $\text{Beta}(1, t)$ distributions with $t \in (0, 1)$ are considered as reasonable choices for the prior probability distribution. Therefore, in the imprecise probability framework of Example 3, when $s = 1$ both courses of action would correspond to optimal decisions according to the criteria of maximality, E-admissibility, or interval dominance. Exactly the same holds in the case with $s = 2$. By contrast, when $s = 1/3$ these criteria would lead to a single optimal decision, corresponding to the first course of action (that is, betting on success), since in this case $\underline{P}(A) > \underline{P}(B)$, as can be seen in Figure 1. That is, the decision resulting from the Bayesian approach is robust, if only the $\text{Beta}(1/3, t)$ distributions with $t \in (0, 1)$ are considered as reasonable choices for the prior probability distribution.

However, if the goal of the imprecise probability analysis is decision making (and not the study of the robustness or non-robustness of Bayesian analyses), then when $s = 1$ or $s = 2$ we still have to select one of the two possible decisions. If we choose one of the four criteria considered in Example 4 as the second decision criterion in a two-stage decision procedure, then we obviously obtain the same conclusions as in Example 4.

Another possibility (besides a second criterion in a two-stage procedure) for choosing a decision when there are multiple optimal decisions, is to select it arbitrarily. Of course, there is no real hope that the resulting decisions can be robust, since arbitrariness is antithetical to robustness. However, one could main-

tain that such an arbitrary choice cannot be non-robust, because from the point of view of the decision criterion all optimal decisions are in a certain sense “equivalent”. But even from this point of view the decisions resulting from the imprecise probability approach are not robust in general, as shown in the following example.

Example 6 *Consider again the decision problem of Example 1 in the imprecise probability framework of Example 3, with as decision criterion maximality, E-admissibility, or interval dominance. In Example 5 we have seen that in this case the first course of action (that is, betting on success) would correspond to the unique optimal decision when $s = 1/3$, while both courses of action would correspond to optimal decisions when $s = 1$. Hence, if we would choose one of the two optimal decisions arbitrarily when $s = 1$, then we could choose the second course of action (that is, betting on failure), which does not correspond to the single optimal decision when $s = 1/3$.*

Therefore, in this situation the decision resulting from the imprecise probability approach is not robust, if both $s = 1/3$ and $s = 1$ are considered as reasonable choices for the parameter s of the imprecise Dirichlet model. Of course, $s = 1/3$ is not a usual choice for this parameter, but it would suffice to slightly modify the decision problem, in order to obtain that the difference in the decisions is between the cases $s = 1$ and $s = 2$ (instead of $s = 1/3$ and $s = 1$). For instance, it would suffice to consider the decision problem corresponding to choosing the side of a bet with odds of 5 to 2 (instead of 4 to 1) on a success in the next Bernoulli trial, where the total stake is a fixed small amount of money (in this situation, the decision resulting from the Bayesian approach would be the same for both Bayes’ and Jeffreys’ priors: betting on success).

Hence, in this subsection we have seen that when a decision has to be made, the imprecise probability approach is in general not more robust than the Bayesian one. In particular, the choice of s in the imprecise probability analyses of Examples 4, 5, and 6 does not appear to be “less crucial” than in the Bayesian analysis of Example 1. In this context, it is important to note that the results would remain substantially unchanged if randomized decisions were allowed in these examples. In this case, we would have infinitely many possible decisions, but the (sets of) randomization probabilities of the optimal decisions would still change in a discontinuous way at either $s = 4/3$ or $s = 1/2$ (depending on the example being considered).

3.3 Indecision

As discussed in Subsection 3.2, decision criteria like maximality, E-admissibility, or interval dominance often do not identify a unique optimal decision, when used in an imprecise probability analysis. Instead of choosing a decision from the set of all optimal decisions, the set itself is sometimes considered as the conclusion resulting from the imprecise probability approach [1, 14, 36]. That is, (partial) indecision is sometimes allowed.

In this case, the set of all possible decisions of the original decision problem is practically replaced by its power set (without the empty set). The resulting new decision problem is in a certain sense smoother than the original one, because the indecision about two (originally) possible decisions can be seen as a middle course between them. Therefore, non-robustness issues regarding the new decision problem can be less serious than those regarding the original one. However, the Bayesian approach too can be applied to the new decision problem, as shown in the following example.

Example 7 *In Example 5 we have considered the decision criteria of maximality, E-admissibility, and interval dominance for the decision problem of Example 1, in the imprecise probability framework of Example 3. We have seen that the first course of action (that is, betting on success) would correspond to the unique optimal decision when $s = 1/3$, while both courses of action would correspond to optimal decisions when $s = 1$ or $s = 2$. Hence, if indecision is allowed, then we would stick to the first course of action when $s = 1/3$, but we would have indecision between the two courses of action when $s = 1$ or $s = 2$.*

In order to apply the Bayesian approach when indecision is allowed, we can define the utility C of the indecision between the two courses of action. Assuming risk aversion, this utility must be larger than the utility of choosing one of the two courses of action at random (by tossing a fair coin) [37]: that is, $C > 0$. The choice $C = 1/10$ is plotted in Figure 1: we can see that in this case the decision resulting from the Bayesian approach would still be the first course of action (that is, betting on success) when $(s, t) = (1/3, 1/2)$, and the second course of action (that is, betting on failure) when $(s, t) = (2, 1/2)$, but it would be the indecision between the two courses of action when $(s, t) = (1, 1/2)$.

The new decision problem in Example 7 can be considered as smoother than the original one in Example 1, because in a certain sense there is a new possible choice (the indecision) somewhere in between the two courses of action. In particular, with the new decision

problem the choice of s is perhaps “less crucial” than with the original one, but this holds for the Bayesian analysis as well as for the imprecise probability analysis.

Apparently, the imprecise probability approach has the advantage of not needing to define the utilities of the cases of (partial) indecision. However, this appearance can be misleading. First, the definition of these utilities can be avoided in the Bayesian approach too, for instance by replacing the posterior expectations of the utilities of the original decisions with their highest posterior density intervals (for a given probability level), and using interval dominance as a decision criterion. Second, and most important, the definition of the utilities for the cases of (partial) indecision is necessary anyway to evaluate and compare the resulting imprecise probability methods: much work has recently been done in this direction [37]. The trouble is that the imprecise probability methods are obtained on the basis of one decision problem (without utilities for the cases of indecision), and are then evaluated on the basis of another (with utilities for the cases of indecision).

The difficulty in evaluating and comparing imprecise probability methods is strictly related to a fundamental issue in the imprecise probability approach to inference and decision making: the difficulty in comparing models with different degrees of imprecision [30]. The discussion of this issue goes far beyond the scope of the present paper, but it is important to note the connection with the difficulty in the choice of the parameter s of the imprecise Dirichlet model of Example 3, since the degree of imprecision of this model increases with s .

4 Conclusion

Imprecise probability methods are often claimed to be robust, or more robust than Bayesian methods. Sometimes the expression “more robust” is simply used as a synonym for “more imprecise” or “less determinate” [23]. However, this use is misleading, if not wrong. In fact, “more robust” has a positive connotation, which “more imprecise” or “less determinate” do not have, and which derives from its usual interpretation in science and engineering as meaning something like “less sensitive to small changes in the conditions or in the assumptions”.

In particular, in the Bayesian approach to inference and decision making, robustness mainly refers to changes in the choice of prior probability distribution. A Bayesian sensitivity analysis (also called robust Bayesian analysis) is the study of the robustness of the conclusions of a Bayesian analysis. The

fact that Bayesian sensitivity analyses are often performed by letting the prior vary in a set of probability distributions can suggest the idea that imprecise probability analyses are robust (since imprecise probability measures can be identified with particular sets of probability measures). In fact, as discussed in Subsection 3.1, imprecise probability analyses can perhaps be considered as more credible than Bayesian ones, and as noted in Subsection 3.2, they can be seen as descriptions of the robustness or non-robustness of Bayesian analyses, when decision criteria like maximality, E-admissibility, or interval dominance are used. However, the robustness of imprecise probability analyses does not refer to the variability of a (precise) prior in a set of probability distributions, but rather to the variability of the (imprecise) prior in a set of imprecise probability distributions.

Another source of confusion about the robustness of imprecise probability methods (besides the meaning of “robust” in the expression “robust Bayesian analysis”) seems to be the idea that they are allowed to be inconclusive, while Bayesian methods are not. In fact, the Bayesian approach to a particular decision problem is sometimes compared to the imprecise probability approach to a modified version of the decision problem, in which (partial) indecision is allowed. As discussed in Subsection 3.3, the new decision problem is in a certain sense smoother than the original one, and so robustness can be less of an issue. However, both approaches can be applied to both decision problems, and a fair comparison is possible only if they are applied to the same one.

In conclusion, imprecise probability methods are in general not robust, and not more robust than Bayesian methods. The robustness of the imprecise probability approach to inference and decision making can be increased by introducing a second-order possibility distribution, allowing a smoother and more efficient updating rule [9, 10], but this goes beyond the scope of the present paper, and will be the subject of future work.

Acknowledgments

The author wishes to thank Alessio Benavoli and other participants in the session “Imprecise Probability Theories for Statistical Problems: Why? How? When?” at CFE-ERCIM 2012 for stimulating discussions, and the anonymous referees for their valuable comments and suggestions.

References

- [1] Antonucci, A., Cattaneo, M., and Corani, G. (2012). Likelihood-based robust classification with Bayesian networks. In *Advances in Computational Intelligence, Part 3*. Springer, 491–500.
- [2] Augustin, T. (2002). Expected utility within a generalized concept of probability – a comprehensive framework for decision making under ambiguity. *Stat. Pap.* 43, 5–22.
- [3] Augustin, T. (2003). On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view: A cautionary note on updating imprecise priors. In *ISIPTA '03*. Carleton Scientific, 31–45.
- [4] Augustin, T., and Hable, R. (2010). On the impact of robust statistics on imprecise probability models: A review. *Struct. Safety* 32, 358–365.
- [5] Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc. Lond.* 53, 370–418.
- [6] Berger, J. O. (1984). The robust Bayesian viewpoint. In *Robustness of Bayesian Analyses*. North-Holland, 63–124.
- [7] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd edn. Springer.
- [8] Bernard, J.-M. (2005). An introduction to the imprecise Dirichlet model for multinomial data. *Int. J. Approx. Reasoning* 39, 123–150.
- [9] Cattaneo, M. (2008). Fuzzy probabilities based on the likelihood function. In *Soft Methods for Handling Variability and Imprecision*. Springer, 43–50.
- [10] Cattaneo, M. (2009). A generalization of credal networks. In *ISIPTA '09*. SIPTA, 79–88.
- [11] Corani, G., and Benavoli, A. (2010). Restricting the IDM for classification. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Methods*. Springer, 328–337.
- [12] Couso, I., Moral, S., and Walley, P. (2000). A survey of concepts of independence for imprecise probabilities. *Risk Decis. Policy* 5, 165–181.
- [13] Cozman, F. G. (2005). Graphical models for imprecise probabilities. *Int. J. Approx. Reasoning* 39, 167–184.
- [14] De Bock, J., and De Cooman, G. (2011). State sequence prediction in imprecise hidden Markov models. In *ISIPTA '11*. SIPTA, 159–168.
- [15] De Cooman, G., Quaeghebeur, E., and Miranda, E. (2009). Exchangeable lower previsions. *Bernoulli* 15, 721–735.
- [16] De Finetti, B. (1975). *Theory of Probability*. Vol. 2. Wiley.
- [17] Destercke, S. (2010). A decision rule for imprecise probabilities based on pair-wise comparison of expectation bounds. In *Combining Soft Computing and Statistical Methods in Data Analysis*. Springer, 189–197.
- [18] Elga, A. (2010). Subjective probabilities should be sharp. *Philos. Imprint* 10.
- [19] Huber, P. J., and Ronchetti, E. M. (2009). *Robust Statistics*. 2nd edn. Wiley.
- [20] Huber, P. J., and Strassen, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Stat.* 1, 251–263.
- [21] Huntley, N., and Troffaes, M. C. M. (2009). Characterizing factuality in normal form sequential decision making. In *ISIPTA '09*. SIPTA, 239–248.
- [22] Hurwicz, L. (1951). Some specification problems and applications to econometric models. *Econometrica* 19, 343–344.
- [23] Hutter, M. (2003). Robust estimators under the imprecise Dirichlet model. In *ISIPTA '03*. Carleton Scientific, 274–289.
- [24] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond., Ser. A* 186, 453–461.
- [25] Levi, I. (1974). On indeterminate probabilities. *J. Philos.* 71, 391–418.
- [26] Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer.
- [27] Pericchi, L. R., and Walley, P. (1991). Robust Bayesian credible intervals and prior ignorance. *Int. Stat. Rev.* 58, 1–23.
- [28] Piatti, A., Zaffalon, M., and Trojani, F. (2005). Limits of learning from imperfect observations under prior ignorance: the case of the imprecise Dirichlet model. In *ISIPTA '05*. SIPTA, 276–286.

-
- [29] Piatti, A., Zaffalon, M., Trojani, F., and Hutter, M. (2009). Limits of learning about a categorical latent variable under prior near-ignorance. *Int. J. Approx. Reasoning* 50, 597–611.
- [30] Seidenfeld, T., Schervish, M. J., and Kadane, J. B. (2011). Forecasting with imprecise probabilities. In *ISIPTA '11*. SIPTA, 317–326.
- [31] Troffaes, M. C. M. (2007). Decision making under uncertainty using imprecise probabilities. *Int. J. Approx. Reasoning* 45, 17–29.
- [32] Troffaes, M. C. M., and Hable, R. (2011). Robustness of natural extension. In *ISIPTA '11*. SIPTA, 361–370.
- [33] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.
- [34] Walley, P. (1996). Inferences from multinomial data: Learning about a bag of marbles. *J. R. Stat. Soc., Ser. B* 58, 3–57.
- [35] Williams, P. (2007). Notes on conditional previsions. *Int. J. Approx. Reasoning* 44, 366–383.
- [36] Zaffalon, M. (2001). Statistical inference of the naive credal classifier. In *ISIPTA '01*. Shaker, 384–393.
- [37] Zaffalon, M., Corani, G., and Mauá, D. (2012). Evaluating credal classifiers by utility-discounted predictive accuracy. *Int. J. Approx. Reasoning* 53, 1282–1301.