Statistical Inference with Belief Functions and Possibility Measures : a discussion of basic assumptions

Didier Dubois[1] and Thierry Denœux[2]

Abstract This paper reconsiders the problem of statistical inference from the standpoint of evidence theory and possibility theory. The Generalized Bayes theorem due to Smets is described and illustrated on a small canonical example. Critiques addressed to this model are discussed as well as the robust Bayesian solution. Finally, the proposal made by Shafer to exploit likelihood information in terms of consonant belief function within the scope of possibility theory is reconsidered. A major objection to this approach, due to a lack of commutativity between combination and conditioning, is circumvented by assuming that the set of hypotheses or parameter values is rich enough.

1 Introduction

Let X be a space of observations. Given a probabilistic parametric model $P_\theta, \theta \in \Theta$, interpreted as a conditional probability $P(\cdot|\theta)$, a set of independent observations $x_1, \ldots, x_k$ obtained in the same conditions and a subjective prior probability $P_{sub}(\theta)$, Bayes' theorem in probability theory prescribes that a posterior probability on Θ can computed as $P(\theta|x_1, \ldots, x_k) \propto \prod_{i=1}^{k} P(x_i|\theta)P_{sub}(\theta)$, where $P(x_i|\theta)$ is the likelihood function. A recurring question in statistical inference is : what information does observed data provide about a probabilistic model when no prior probability is supplied and Bayes' theorem cannot be applied? what to say on the basis of observations, when only likelihood information is available?

In this paper, we review some statistical inference methods that can be proposed in the setting of belief functions and possibility theory. Both settings have the merit of not requiring prior knowledge when learning from data. We try to provide a clear presentation of Smets' Generalized Bayes theorem without prior, laying bare the

1. IRIT, CNRS and Université de Toulouse, France, `dubois@irit.fr` ·
2. HEUDIASYC, CNRS and Université de Technologie de Compiègne, France, `thierry.denoeux@hds.utc.fr`

assumptions. Then we discuss the related literature in probability theory. We study to what extent a similar approach makes sense in possibility theory.

2 The Generalized Bayes Theorem for belief functions

The problem of inferring knowledge from likelihood functions has been addressed by Philippe Smets in his 1978 thesis [10] in the setting of belief functions, given several observations forming a finite set X and non-binary parameter space Θ. The Generalized Bayes Theorem (GBT) computes a non-trivial uncertainty measure on the parameter space from parameterized belief functions on X even if no prior knowledge about the parameter is available. If there is some prior information, it can be used. Bayes' theorem is retrieved in the special case where belief functions are probability measures and a prior probability distribution on Θ is given. The GBT has been applied to classification problems [3]. It is interesting to study what are its underlying assumptions and under which conditions it can be applied to statistical inference; of interest is how it compares with other approaches.

Let X be a frame of discernment. An uncertain body of evidence is defined by means of a mass function m which is a probability distribution over the power set 2^X. In particular, $\sum_{E\subseteq X} m(E) = 1$. The mass $m(E)$ is the probability mass that could be allocated to some element of E but is not by lack of information. The quantity $m(\emptyset)$ represents a degree of internal conflict, and according to Smets, may suggest the idea that the truth may lie outside X (open world assumption). For simplicity, we assume $m(\emptyset) = 0$ (closed world assumption). The following notions are useful in the sequel:

- The degree of belief is $bel(A) = \sum_{E\subseteq A} m(E)$;
- The degree of plausibility is $pl(A) = \sum_{\neq E\cap A} m(E) = 1 - bel(\overline{A})$, where $\overline{A}$ is the complement of A;
- Standard (normalized) conditioning : $pl(A|B) = \frac{pl(A\cap B)}{pl(B)}; bel(A|B) = \frac{bel(A\cup\bar{B})-bel(\bar{B})}{1-bel(B)}$;
- Conjunctive merging $\cap\!\!\!\!\bigcirc$: $(m_1 \cap\!\!\!\!\bigcirc m_2)(C) = \sum_{A,B,A\cap B=C} m_1(A)m_2(B)$;
- Dempster rule of combination $\oplus$: It consists in renormalizing $m_1 \cap\!\!\!\!\bigcirc m_2$ dividing it by $1-(m_1 \cap\!\!\!\!\bigcirc m_2)(\emptyset)$, which makes sense under a closed-world assumption.

Given a family $\{bel_X(\cdot|\theta), \theta \in \Theta\}$ of belief functions (supposed to be normalized), parameterized by θ, the ballooning extension (or conditional embedding) of $bel_X(\cdot|\theta)$ into $X \times \Theta$ is the least committed belief function whose conditional on θ is $bel_X(\cdot|\theta)$. It consists in assigning each mass $m_X(E|\theta)$ to the subset $E \cup \overline{\{\theta\}} \subseteq X \times \Theta, \forall E \subseteq X$. On $X \times \Theta$, the ballooning extension is such that $bel^{\theta}(E \cup \overline{\{\theta\}}) = bel_X(E|\theta)$ (assuming $pl(E \cup \overline{\{\theta\}}) = 1, \forall \theta \in \Theta$).

The inference problem can then be stated as follows: Given a set of parametric belief functions $bel_X(\cdot|\theta)$, $\theta \in \Theta$, and some observation $x \in X$, compute $bel_{\Theta}(\cdot|x)$. It is assumed that for $T \subseteq \Theta, pl_X(x|T)$ is a function of elementary likelihoods $pl_X(x|\theta), pl_X(x|\overline{\{\theta\}}), \theta \in T$. Computing the posterior belief function $bel_{\Theta}(\cdot|x)$ goes as follows, given a **finite** parameter space Θ and a set of parametric belief functions $bel_X(\cdot|\theta), \theta \in \Theta$:

1. **Conditional embedding** of each $bel_X(\cdot|\theta)$ in $X \times \Theta$ (ballooning);
2. **Conjunctive merging** of the embedded belief functions $bel^\theta, \theta \in \Theta$ on $X \times \Theta$;
3. **Conditioning** of the result on the observation x;
4. **Marginalizing** on Θ.

The use of the conjunctive merging rule in step 2 assumes that the belief functions $bel_X(\cdot|\theta), \theta \in \Theta$ have been inferred from distinct sets of empirical data obtained from independent sources. Moreover, this step comes down to applying to $T = \Theta$ the disjunctive combination rule to the conditional belief functions $bel_X(\cdot|\theta)$: $bel_X(A|T) = \prod_{\theta \in T} bel_X(A|\theta), \forall A \subseteq X$. Finally, after marginalization, posterior plausibility functions $pl_\Theta(T|A)$ are proportional to $1 - \prod_{\theta \in T}(1 - pl_X(A|\theta)), \forall T \subseteq \Theta$.

The problem has been extended to n independent observations $x_1, \ldots, x_n$ in $\{x, \bar{x}\}^n$ [11]. The GBT has a nice commutativity property. One may compute $bel_{X^n}(x_1, \ldots x_n|\theta)$, conjunctively combining $bel_X(\cdot|\theta)$, perform a conditional embedding on $X^n \times \Theta$, then get the posterior belief function $bel_\Theta(\theta|x_1, \ldots, x_n)$. It is equivalent to computing n posterior belief functions $bel_\Theta(\theta|x_i)$ and get the same $bel_\Theta(\theta|x_1, \ldots, x_n)$ by Dempster's rule of combination of these $bel_\Theta(\theta|x_i)$. In other words the following identity holds: $bel_\Theta(\cdot|x_1, \ldots, x_n) = bel_\Theta(\cdot|x_1) \oplus \ldots \oplus bel_\Theta(\cdot|x_n)$.

3 Computing the posterior belief function from likelihoods

Suppose only a finite number of frequentist likelihood functions $\{P(\cdot|\theta_i), i = 1, \ldots, k\}$, are available, ***and each one comes from a different population***. The procedure then specializes as follows:

1. *Conditional embedding* of $P(\cdot|\theta_i)$ over $X \times \Theta$ into belief functions bel^i: the associated mass function is defined by $m^i(\overline{\theta_i} \cup \{x\}) = m^i(\{(\theta_i, x)\} \cup (\overline{\{\theta_i\}} \times X)) = P(\{x\}|\theta_i), x \in X$; bel^i on $X \times \Theta$ has a vacuous marginal on Θ and yields $P(\cdot|\theta_i)$ back when conditioned on θ_i.
2. *Conjunctive merging* of the bel^θ's on $X \times \Theta$. This step comes down to assigning mass $\prod_{i=1,\ldots,k} P(x_{j_i}|\theta_i)$ to the set $\bigcap_{i=1,\ldots,k}\{(\theta_i, x_{j_i})\} \cup (\overline{\{\theta_i\}} \times X) = \bigcup_{i=1,\ldots,k}\{(\theta_i, x_{j_i})\}$. Let ϕ be the mapping assigning observation x_{j_i} to each θ_i. We can write $m(\phi)$ for $m(\bigcup_{i=1,\ldots,k}\{(\theta_i, x_{j_i})\})$.
3. *Conditioning m on the observation x*. Then $pl_\Theta(\theta|x) = \frac{\sum_{\phi:\phi(\theta)=x} m(\phi)}{\sum_{\theta \in \Theta} \sum_{\phi \in X^\Theta:\phi(\theta)=x} m(\phi)}$.

The simplest example of the problem is a simple space $\mathscr{S} = \{x, \bar{x}\} \times \{\theta, \bar{\theta}\}$ with two possible mutually exclusive hypotheses $\Theta = \{\theta, \bar{\theta}\}$, and two possible mutually exclusive observations $\{x, \bar{x}\}$. The available knowledge consists in the two likelihood values $a = P(x|\theta) > b = P(x|\bar{\theta})$. And it is assumed that x is observed.

For this example (actually studied by Shafer [9]), conditional embedding comes down to defining $m_1(x \cup \bar{\theta}) = a, m_1(\bar{x} \cup \bar{\theta}) = 1 - a$, and likewise: $m_2(x \cup \theta) =$

$b, m_2(\overline{x} \cup \theta) = 1 - b$. Conjunctive merging yields $m(x) = ab; m(\overline{x}) = (1-a)(1-b)$; $m((x \cap \theta) \cup (\overline{x} \cap \overline{\theta})) = a(1-b); m((x \cap \overline{\theta}) \cup (\overline{x} \cap \theta)) = a(1-b)$.

The following results are obtained if x is observed:

$$bel_\Theta(\theta|x) = \frac{pl(x) - pl(x \cap \overline{\theta})}{pl(x)} = \frac{a(1-b)}{a+b-ab}; bel_\Theta(\overline{\theta}|x) = \frac{b(1-a)}{a+b-ab}. \quad (1)$$

It is natural that $bel_\Theta(\theta|x)$ should be all the higher as $P(x|\theta)$ is close to 1 and $P(x|\overline{\theta})$ is low. In particular

1. $bel_\Theta(\theta|x) = 1$ if and only if $P(x|\theta) = 1$ and $P(x|\overline{\theta}) = 0$;
2. $bel_\Theta(\theta|x) = 0 = bel_\Theta(\overline{\theta}|x)$ if and only if $P(x|\theta) = P(x|\overline{\theta}) = 0$ or $= 1$;
3. If $a = b$ then $0 \leq bel_\Theta(\theta|x) = bel_\Theta(\overline{\theta}|x) \leq 1/4$.

Shafer [9] extended this example to n observations of the form x or $\overline{x}$. He showed that for large values of $n, bel(\theta|x_1, \ldots, x_n) + bel(\overline{\theta}|x_1, \ldots, x_n) \approx 1$ and that the posterior beliefs agree at the limit with the Bayesian solution with uniform prior.

A different approach applies sensitivity analysis to Bayes rule, varying the unknown prior probability. This approach is popular in the robust Bayesian community where some prior information is supposed to be available in the form of a suitable family of probability functions (see Whitcomb[13] for a bibliography). The sensitivity analysis approach and the GBT presuppose different assumptions: In the former, no information on the dependence between the two items $a = P(x|\theta)$ and $b = P(x|\overline{\theta})$ is assumed; but in case of total ignorance on the prior, the resulting posterior is unknown and no information is gained from observing x. But the GBT assumes cognitive independence between two distinct populations or sources that provide each likelihood function. This is what makes the posterior belief function non-trivial. A number of other approaches to the *no prior* problem come down to selecting a "reasonable" probability measure on $\mathscr{S}$ in the set $\mathscr{P} = \{P, a = P(x|\theta) > b = P(x|\overline{\theta})\}$, induced by the likelihood values, for instance applying the maximum likelihood principle, i.e. maximizing $P(x)$ (which is not so good as it results in $P(\theta|x) = 1$). Several such approaches are reviewed by Dubois, Gilio and Kern-Isberner [6] : maximal entropy, Shapley value, uniform prior, etc.

Alternatively, one may keep the likelihood values upon observing x as $\lambda(\theta) = a, \lambda(\overline{\theta}) = b$ and view them as a measures of confidence, as strongly advocated by frequentist statisticians after Fisher and Edwards [7]; however, this approach may be considered as lacking formal foundations, all the more so as this school of thought never considers extending such uncertainty measures from elementary parameter values to disjunctions thereof.

4 Critiques of the GBT

There are several situations where the GBT is questionable as shown by Shafer [9]. Moreover, some authors like Walley [12] have criticized it as not satisfying the strong likelihood principle.

1. **The binomial example.** Consider the case of a coin such that $P(x|\theta) = \theta \in \Theta = [0,1]$ is the probability of getting a tail (x), to be learned from observations. We now have an uncountable infinite family of conditional belief functions such that $bel_X(x|\theta) = \theta, bel_X(\overline{x}|\theta) = 1-\theta$, $\theta \in [0,1]$. *The assumption that these belief functions have been obtained from distinct sets of data is no longer tenable, as this would imply an infinite quantity of information*! A way to circumvent this problem could be to discretize the domain Θ into $\Theta' = \{\theta_1, \ldots, \theta_k\}$, with $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$. However, the k belief functions $bel_X(\cdot|\theta_i)$ for $i = 1, \ldots, k$ are now linked by the following relations: $bel_X(x|\theta_i) \leq bel_X(x|\theta_j)$ whenever $\theta_i \leq \theta_j$. Consequently, they cannot be independent. As noted by Shafer [9], "the choice of a belief function analysis depends on the nature of the evidence for the model, not just on the model itself".
2. **The fiducial example.** Shafer also considers the case of a measuring instrument with errors. Let $\theta \in \Theta$ be the unknown quantity and $x \in X$ be the measured value. It is supposed that $\Theta = X = \mathbb{N}$ contains integers. Suppose we know the symmetric probability distribution P of errors $e = x - \theta$. This probability distribution can be viewed as a belief function on $X \times \Theta$, letting $m(\{(x,\theta) : e = |x-\theta|\}) = P(e)$. The projection of this belief function on X is $bel(x|\theta) = P(x-\theta)$, i.e., it is additive and coincides with $P(x|\theta)$. But the same holds for the projection of this belief function on Θ, since $P(x|\theta) = P(\theta|x) = P(x-\theta)$. If $\Theta = X = \{0,1\}$ and $x = \theta + e$ modulo 2, assuming $P(0) = a, P(1) = b = 1-a$, we find that $bel(\theta = 1|x = 1) = a$, which differs from the value obtained with the GBT if $b = 1-a$, that is $\frac{a^2}{1-a-a^2}$. Again in this case the two likelihood functions are related.
3. **The strong likelihood principle.** In the statistical literature, likelihood functions are considered to live on a ratio scale. Edwards [7] considers the likelihood function $\lambda(\theta)$ to be proportional to $P(x|\theta)$, the proportionality constant being arbitrary. In particular, no comparison of likelihood of hypotheses across data sets, say $\lambda_1(\theta) = P(x_1|\theta)$ and $\lambda_2(\theta) = P(x_2|\theta)$ is considered meaningful; only likelihood ratios $\frac{P(x_2|\theta)}{P(x_1|\theta)}$ make sense. Moreover the likelihood principle states that all the information that is provided by the data x concerning the relative merits of two hypotheses θ_1 and θ_2 is contained in the likelihood ratio of these hypotheses. Hence the invariance property, recalled by Walley [12], here stated in terms of belief functions: Let f be the function such that $bel_\Theta(\cdot|x_1, \ldots x_n) = f(P(x_i|\theta), i = 1, \ldots n, \theta \in \Theta)$. Then, for all real values $c > 0$, $f(P(x_i|\theta), i = 1, \ldots n, \theta \in \Theta) = f(c \cdot P(x_i|\theta), i = 1, \ldots n, \theta \in \Theta)$. It is clear that the GBT violates this property, as well as other inference techniques recalled in Section 2. However the Bayesian inference method does satisfy this strong likelihood principle. Walley essentially shows that, when the initial information takes the form of likelihood functions $P(x_i|\theta)$, enforcing the strong likelihood principle to the GBT leads to a probabilistic posterior belief function where the plausibility of each singleton $\theta \in \Theta$ is proportional to $P(x|\theta)^\alpha$ for some $\alpha > 0$. So it comes down to working with a Bayesian approach under uniform priors, up to a rescaling of the likelihood functions.

Is the strong likelihood principle a *sine qua non* condition for statistical inference? It can be questioned. First, there seems to be a clash of intuitions between this principle and the frequentist approach based on a fixed amount of observations N. Suppose $P(x|\theta)$ derives from the result of experiments that yield $N(x\theta) = n_1, N(\bar{x}\theta) = n_2, N(x\bar{\theta}) = n_3, N(\bar{x}\bar{\theta}) = n_4$ with $N = \sum_{i=1}^4 n_i$. Then $P(x|\theta) = a = \frac{n_1}{n_1+n_2}$ and $P(x|\bar{\theta}) = b = \frac{n_3}{n_3+n_4}$. Hence $\frac{n_1}{a} + \frac{n_3}{b} = N$, so that multiplying a and b by positive constant c clearly implies dividing N by c. In such a situation, claiming the invariance of the likelihood under positive scalar multiplication comes down to considering the statistical validity of the joint probability distribution on $X \times \Theta$ as not being affected by the number N of outcomes.

Another reason for questioning the strong likelihood principle is that if we extend the likelihood $\lambda(\theta) = cP(x|\theta)$ of elementary hypotheses, viewed as a representation of uncertainty about θ, to disjunctions of hypotheses, the corresponding set-function Λ should obey the laws of possibility measures [2, 5] in the absence of probabilistic prior, namely, the following properties look reasonable for such a set-function Λ:

- The properties of probability theory enforce $\forall T \subseteq \Theta, \Lambda(T) \leq \max_{\theta \in T} \lambda(\theta)$;
- A set-function representing likelihood should be monotonic with respect to inclusion: If $\theta \in T, \Lambda(T) \geq \lambda(\theta)$;
- Keeping the same scale as probability functions, we assume $\Lambda(\Theta) = 1$.

Then it is clear that $\lambda(\theta) = \frac{P(x|\theta)}{\max_{\theta \in \Theta} P(x|\theta)}$ and $\Lambda(T) = \max_{\theta \in T} \lambda(\theta)$, i.e., the extended likelihood function is a possibility measure, and the coefficient c is then fixed. We find Shafer [8] proposal of a consonant belief function induced by likelihood information.

5 Statistical inference in possibility theory

It is interesting to see if the same approach as the GBT can be carried out in the more restrictive setting of possibility theory, where only consonant belief functions are used. Suppose conditional possibility distributions $\{\pi(\cdot|\theta), \theta \in \Theta\}$ in the unit interval are available. The consonant conditional embedding consists in defining possibility distributions π_θ on $X \times \Theta$ as $\pi_\theta(x, \theta_i) = \pi(x|\theta)$ if $\theta_i = \theta$ and 1 otherwise. It is clear that the projection of π_θ on Θ is vacuous, i.e., $\max_{x \in X} \pi_\theta(x, \theta_i) = 1, \forall \theta_i \in \Theta$. Combining all these $\pi_\theta(\cdot, \cdot)$ conjunctively by means of any t-norm just yields the joint possibility distribution $\pi(x, \theta) = \pi(x|\theta)$. By conditioning on observation x, it yields $\pi_\Theta(\theta|x) = \frac{\pi(x|\theta)}{\max_{\theta' \in \Theta} \pi(x|\theta')}$.

In case of n observations x_i, we are faced again with two procedures to compute $\pi(\theta|x_1, \ldots, x_n)$: Either combine the resulting conditional possibilities $\pi_\Theta(\theta|x_i)$, using an appropriate t-norm $\star$, or combine first the possibilistic likelihoods as $\pi(x_1, \ldots, x_n|\theta)$ and condition next. It is clear that these two procedures are not equivalent since $\star_{i=1\ldots n} \frac{\pi(x_i|\theta)}{\max_{\theta' \in \Theta} \pi(x_i|\theta')} \neq \frac{\star_{i=1\ldots n} \pi(x_i|\theta)}{\max_{\theta' \in \Theta} \star_{i=1\ldots n} \pi(x_i|\theta')}$. This difficulty is the cause of the rejection of this technique by Shafer himself [9]. In fact it is easy to see that a sufficient condition for these two approaches coinciding is that

$$\max_{\theta' \in \Theta} \pi(x|\theta') = 1, \forall x \in X.$$

This property, previously laid bare in [4], can be called the Hypothesis Exhaustivity Assumption (HEA). It means that the distribution $\pi(x|\theta)$ is a normalized possibility distribution on Θ as much as it is on X. This situation is similar to the one for probabilistic likelihood functions in the fiducial case. This is an assumption about Θ stating that for any piece of evidence $x \in X$, at least one hypothesis θ is not in conflict with x, i.e., $\forall x, \exists \theta, \pi(x|\theta) = 1$. It will hold if Θ is large enough to explain all observations. Aickin [1] seems to have rediscovered it and calls $\pi(x|\theta)$ *committed to the model*.

An example where such an assumption is verified is the following: Suppose lower probability bounds $0 < a_{x\theta} \leq P(x|\theta)$ are available. They can be viewed as conditional necessity values $N(\{x\}|\theta) = a_{x\theta}, \theta \in \Theta$. Now, $N(\{x\}|\theta) = a_{x\theta} > 0$ is equivalent to $\pi(x|\theta) = 1$, $\pi(x'|\theta) = 1 - a_{x\theta}$ for $x' \neq x$. The HEA on Θ now means that for each $x \in X$ there is a constraint of the form $0 < a_{x\theta} \leq P(x|\theta)$ for some $\theta \in \Theta$, so that this observation is totally possible, under some assumption θ. Let $\Theta(x) = \{\theta \in \Theta, P(x|\theta) \geq a_{x\theta} > 0\}$ be the set of hypotheses that may tentatively explain x. The HEA says $\forall x \in X, \Theta(x) \neq \emptyset$. Note that $\Theta(x) = \{\theta \in \Theta, \pi(x|\theta) = 1\}$, so that $\forall x \in X, \max_{\theta' \in \Theta} \pi(x|\theta') = 1$ holds.

Let us now consider the properties of possibilistic inference in this case:

- If lower bounds on likelihoods are viewed as unrelated items of possibilistic information, we can combine possibility degrees via product in case of a sequence of observations $x_1, \dots x_n$: $\pi(\theta|x_1, \dots x_n) = \prod_{i=1,\dots,n} \pi(x_i|\theta) = \prod_{i:\theta \notin \Theta(x_i)} (1 - a_{x_i\theta})$. It means that we can all the more certainly rule out assumption θ as there are more observations for which θ is not a plausible explanation.
- $N(\theta|x_1, \dots x_n) = 1 - \max_{\theta' \neq \theta} \prod_{i=1,\dots,n} \pi(x_i|\theta') > 0$ only if $\forall \theta' \neq \theta, \exists x_i, \pi(x_i|\theta') < 1$, that is: $\forall \theta' \neq \theta, \exists x_i : \theta' \notin \Theta(x_i)$. It means that:
  - We become more and more certain about θ as long as all hypotheses other than θ fail to plausibly explain one of the observations.
  - We have no longer any certainty at all about θ, if $\theta' \in \bigcup_{i=1,\dots,n} \Theta(x_i)$, for some $\theta' \neq \theta$, i.e., some hypothesis other than θ can explain the whole set of observations.

In other words this form of statistical inference looks as reasonable as can be.

6 Conclusion

It is clearly interesting from both theoretical and practical points of view to reconsider the statistical inference methodology outside the Bayesian framework, beyond a mere sensitivity analysis method as done by robust statisticians, when only likelihood functions, or even only bounds on them are available and prior probabilities are not assigned. In particular, it is clear that the inference technique should depend on what kind of information is available and on the way it is acquired. One situation where likelihood functions can be exploited in a non-trivial way is when these likelihoods come from separate populations for each parameter values. More generally, some additional assumption is needed to complement the pure likelihood information. This paper has reviewed a number of techniques to that effect, whereby the notion of conditioning at work in learning schemes of probabilistic inference is extended to other theories of uncertainty. It seems that possibility theory may play a key role in the development of simple inference techniques under poor information, especially as an approximation of more complex methods, due to the close connections between likelihoods and possibility distributions. A more extensive account of the literature is needed so as to encompass alternative approaches based on imprecise probabilities such as the imprecise Dirichlet model. It is useful to re-examine, in the light of the GBT and the possibilistic inference scheme, Bayesian objections against classical likelihood-based inference techniques, which have often been developed in an ad hoc way with no relations to new uncertainty theories.

References

1. M. Aickin. Connecting Dempster-Shafer belief functions with likelihood-based inference, Synthese, 123(3): 347-364 (2000)
2. G. Coletti, R. Scozzafava. Coherent conditional probability as a measure of uncertainty of the relevant conditioning events. Proc. of ECSQARU03, LNAI 2711, pp. 407-418, Springer Verlag (2003)
3. T. Denœux, P. Smets. Classification using belief functions: the relationship between the case-based and model-based approaches, IEEE Transactions on Systems, Man and Cybernetics B, 36(6): 1395-1406 (2006)
4. D. Dubois, H. Prade. On the combination of evidence in various mathematical framework, In: Reliability Data Collection and Analysis (J. Flamm, T. Luisi, eds.), Kluwer Acad. Publ., Dordrecht, pp. 213-241 (1992)
5. D. Dubois. Possibility theory and statistical reasoning. Computational Statistics & Data Analysis 51(1): 47-69 (2006)
6. D. Dubois, A. Gilio, G. Kern-Isberner. Probabilistic abduction without priors. Int. J. Approx. Reasoning 47(3): 333-351 (2008)
7. W. F. Edwards, Likelihood, Cambridge University Press, Cambridge, U.K. (1972)
8. G. Shafer. A Mathematical Theory of Evidence. Princeton University Press (1976)
9. G Shafer. Belief Functions and Parametric Models, Journal of the Royal Statistical Society. Series B, 44: 322-352 (1982)
10. P. Smets. Un modèle mathématico-statistique simulant le processus du diagnostic médical, Université Libre de Bruxelles, Brussels, Belgium (1978)
11. P. Smets. Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approx. Reasoning 9(1): 1-35 (1993)
12. P. Walley. Belief Function Representations of Statistical Evidence, Ann. Statist. 15:1439-1465 (1987)
13. K. Whitcomb. Quasi-Bayesian analysis using imprecise probability assessments and the generalized Bayes rule. Theory and Decision, 58 : 209-238 (2005)