International Journal of Approximate Reasoning 55 (2014) 1614–1617

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Discussion

Rejoinder on "Likelihood-based belief function: Justification and some extensions to low-quality data"

Thierry Denœux

Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, France

ARTICLE INFO

Article history:
Received 23 February 2014
Received in revised form 3 April 2014
Accepted 7 April 2014
Available online 1 May 2014

Keywords:
Statistical inference
Dempster–Shafer theory
Evidence theory
Likelihood principle
Uncertain data

ABSTRACT

This note is a rejoinder to comments by Dubois and Moral about my paper "Likelihood-based belief function: justification and some extensions to low-quality data" published in this issue. The main comments concern (1) the axiomatic justification for defining a consonant belief function in the parameter space from the likelihood function and (2) the Bayesian treatment of statistical inference from uncertain observations, when uncertainty is quantified by belief functions. Both issues are discussed in this note, in response to the discussants' comments.

I thank the discussants, Dr. Dubois and Prof. Moral, for there insightful and thought-provoking comments about my paper "Likelihood-based belief function: justification and some extensions to low-quality data" published in this issue [6,7,11]. Both discussants questioned some of the axioms I proposed in order to justify the method for defining a consonant belief function from the likelihood function introduced by Shafer [13]. Additionally, Moral [11] argued for a Bayesian analysis of statistical inference from uncertain observations, different from the approach presented in [6]. These two issues are further examined below, in response to the discussants' comments.

1. Axiomatic justification of likelihood-based belief functions

In [13], Shafer introduced a method for modeling statistical evidence using a consonant belief function. In [6], I attempted to justify this method axiomatically by reasoning as follows. Let $X \in \mathbb{X}$ denote the observable data, $\theta \in \Theta$ the parameter of interest and $f(x;\theta)$ the probability mass or density function describing the data-generating mechanism. Having observed a realization x of X, we wish to quantify the uncertainty on Θ using a belief function $Bel_\Theta(\cdot;x)$. Which properties should $Bel_\Theta(\cdot;x)$ have? In [6], I proposed the following three requirements:

1. Likelihood principle: $Bel_\Theta(\cdot;x)$ should only depend on the likelihood function, defined by $L(\theta;x) = \alpha f(x;\theta)$ for all $\theta \in \Theta$, where α is any positive multiplicative constant.
2. Compatibility with Bayesian inference: if a Bayesian prior $\pi(\theta)$ is available, combining it with $Bel_\Theta(\cdot;x)$ using Dempster's rule should yield the Bayesian posterior.
3. Least Commitment Principle: $Bel_\Theta(\cdot;x)$ should be the least committed belief function, among all those satisfying the previous two requirements.

DOI of original article: http://dx.doi.org/10.1016/j.ijar.2013.06.007.
DOIs of comments: http://dx.doi.org/10.1016/j.ijar.2014.04.002, http://dx.doi.org/10.1016/j.ijar.2014.04.005.

http://dx.doi.org/10.1016/j.ijar.2014.04.010
0888-613X/

The first two requirements jointly imply that the contour function $pl(\theta; x)$ associated to $Bel_\Theta(\cdot; x)$ should be proportional to the likelihood function:

$$pl(\theta; x) \propto L(\theta; x). \tag{1}$$

Assuming the commonality ordering [8] to be relevant for comparing the informational content of belief functions, the third requirement, together with the other two, then yields a unique solution, which is the consonant belief function whose contour function is the relative likelihood function:

$$pl(\theta; x) = \frac{L(\theta; x)}{\sup_{\theta' \in \Theta} L(\theta'; x)}. \tag{2}$$

Dubois [7] and Moral [11] both acknowledge that this formal result provides a justification for viewing the normalized likelihood function as defining a consonant belief function, a view that has been shared by several authors (see, e.g., [17,21,1], among others). However, they also claim that the axioms that support this solution (in particular, requirements 2 and 3 above) are questionable and that, consequently, they are not sufficiently compelling to rule out alternative methods of statistical inference in the belief function framework.

I can only agree with the discussants on this point. Indeed, I never claimed that the approach advocated in [6] is the only acceptable one for statistical inference in the belief function framework. As a matter of fact, formal requirements justifying a solution to some problem are rarely unquestionable or self-evident (see, e.g., the long-lasting debates on Savage's or Cox's axioms [15,19,16]). The axiomatic approach only lays bare some basic principles, from which the solution can be derived. These principles should be as "natural" or "reasonable" as possible. However, accepting or rejecting them is a matter of personal judgement.

Dubois [7] is, of course, right to point out that the informational ordering of belief functions based on their commonalities is not the only "reasonable" one. Other examples are the ordering based on plausibilities, the specialization ordering, as well as orderings based on the canonical decomposition of belief functions, as introduced in [5]. None of these orderings seems to qualify as the unique way of comparing the information content of belief functions. Similarly, there is no unique equivalent of Shannon's entropy for measuring the degree of uncertainty of belief functions (see, e.g., [12,10]). Consequently, any application of the Least Commitment Principle [9,18] can be criticized as being based on a more or less arbitrary choice of an ordering or an uncertainty measure. Some arguments in favor of one informational ordering or another might be found in the future. It seems unlikely, though, that such arguments will be strong enough to discard other solutions. I am more inclined to accept the coexistence of several notions of information content for belief functions, just as there exist several notions of independence (see, e.g., [2–4]).

In his comments, Moral [11] focuses his criticism on the second requirement, namely, that combining the belief function induced by observations with a Bayesian prior *using Dempster's rule* yields the Bayesian posterior. Indeed, the belief function induced by two independent samples is not equal to the orthogonal sum of the belief functions induced by each of the samples. As noted in [6], this remark suggests that Dempster's rule is not the only reasonable mechanism for pooling statistical evidence, which leaves open the possibility of defining other ways of combining a Bayesian prior with a belief function induced by the data. Starting from this observation, Moral [11] proposes two alternatives to the consonant belief function induced by (2). However, while these two alternative solutions make sense in the imprecise probability framework, when interpreting upper probabilities as lower selling prices [20], they appear to be weakly justified in the belief function framework. Actually, one of the proposed solutions (denoted by $\overline{P}_4$) is not a plausibility function, and it is not known how consistency of the other solution (denoted by pl_3) with Bayesian inference could be achieved.

To conclude this section, we might observe that, while the discussants' criticisms are partly well-founded, they do not point to any better justified alternative to the method advocated in [6] for representing statistical evidence in the belief function framework.

2. Bayesian analysis of inference from uncertain data

In [6], I considered the situation where the data x are imperfectly observed, and uncertainty about x is described by a mass function $m_{\mathbb{X}}$ induced by a finite random set $(\Omega, 2^\Omega, P_\Omega, \Gamma)$ (to simplify the discussion, the sample space $\mathbb{X}$ is assumed to be finite). I then proposed an extension of the likelihood function to such uncertain data, as:

$$L(\theta; m_{\mathbb{X}}) = \sum_{x \in \mathbb{X}} f(x; \theta) pl(x), \tag{3}$$

where $pl(x)$ is the contour function associated to $m_{\mathbb{X}}$. This corresponding contour function in the parameter space is then

$$pl(\theta; m_{\mathbb{X}}) = \frac{L(\theta; m_{\mathbb{X}})}{\sup_{\theta \in \Theta} L(\theta; m_{\mathbb{X}})}, \tag{4}$$

which is a proper generalization of (2). In [6], I also proposed a Bayesian analysis of this problem, and arrived at the conclusion that the posterior probability function $f(\theta | m_{\mathbb{X}})$ given $m_{\mathbb{X}}$ is not proportional to $pl(\theta; m_{\mathbb{X}})$, which means that the Dempster–Shafer and Bayesian analyses of this problem yield different solutions.