International Journal of Approximate Reasoning 72 (2016) 71–94

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Prediction of future observations using belief functions: A likelihood-based approach

Orakanya Kanjanatarakul [a,b,*], Thierry Denœux [b], Songsak Sriboonchitta [c]

[a] Faculty of Management Sciences, Chiang Mai Rajabhat University, Thailand
[b] Sorbonne Universités, Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, France
[c] Faculty of Economics, Chiang Mai University, Thailand

ARTICLE INFO

Article history:
Received 25 January 2015
Received in revised form 7 December 2015
Accepted 7 December 2015
Available online 14 December 2015

Keywords:
Dempster–Shafer theory
Evidence theory
Statistical inference
Forecasting

ABSTRACT

We study a new approach to statistical prediction in the Dempster–Shafer framework. Given a parametric model, the random variable to be predicted is expressed as a function of the parameter and a pivotal random variable. A consonant belief function in the parameter space is constructed from the likelihood function, and combined with the pivotal distribution to yield a predictive belief function that quantifies the uncertainty about the future data. The method boils down to Bayesian prediction when a probabilistic prior is available. The asymptotic consistency of the method is established in the iid case, under some assumptions. The predictive belief function can be approximated to any desired accuracy using Monte Carlo simulation and nonlinear optimization. As an illustration, the method is applied to multiple linear regression.

1. Introduction

The Dempster–Shafer theory of belief functions [11,12,58] is now a well established formal framework for reasoning with uncertainty. It has been successfully applied to many problems, including classification [21], function approximation [54,64], clustering [20,47], image segmentation [40], scene perception [39], multiple-attribute decision making [10], machine diagnosis and prognosis [56,57], etc. To further extend the application of Dempster–Shafer theory to new problems, we need well-founded and computationally tractable methods to model different kinds of evidence in the belief function framework. The purpose of this paper, which builds on previous work by the authors [18,19,35,36], is to present such methods for statistical inference and prediction.

Although statistical inference provided the first motivation for introducing belief functions in the 1960s [11–13], applications in this area have remained limited. The reason might be that the approach initially introduced by Dempster [15], and further elaborated in recent years [41,45,46] under the name of the "weak belief" model, is computationally demanding and it cannot be applied easily to the complex statistical models encountered in many areas, such as machine learning or econometrics. For this reason, frequentist and Bayesian methods have remained by far the most popular. Yet, these approaches are not without defect. It is well known that frequentist methods provide pre-experimental measures of the accuracy of statistical evidence, which are not conditioned on specific data [8]. For instance, a 95% confidence interval contains the parameter of interest for 95% of the samples, but the 95% value is just an average, and the interval may certainly (or certainly not) contain the parameter for some specific samples [8, page 5]. For this reason, a confidence level or a p-value are not appropriate measures of the strength of statistical evidence (see more discussion on this point in [8]). Bayesian methods do

* Corresponding author.

http://dx.doi.org/10.1016/j.ijar.2015.12.004

implement some form of post-experimental reasoning. However, they require the statistician to provide a prior probability distribution, which is problematic when no prior knowledge, or only weak information, is available. These shortcomings of traditional methods of inference have motivated the development of alternative approaches up to these days. The theory of belief functions, which focuses on the concept of evidence [58], seems particularly well-suited as a model of statistical evidence. However, statistical methods based on belief functions will not gain widespread acceptance unless they are conceptually simple and easily applicable to a wide range of problems and models.

In this paper, we advocate another approach to statistical inference using belief functions, based on the concept of likelihood. This approach was initially introduced by Shafer in [58, Chapter 11] and was later studied by some authors [1,65]. It was recently derived axiomatically from three principles: the likelihood principle, compatibility with Bayesian inference and the principle of maximum uncertainty [18,19]. This approach is in line with likelihood-based inference as advocated by Fisher in his later work [28] and, later, by Birnbaum [9], Barnard [5], and Edwards [27], among others. It retains the idea that "all we need to know about the result of a random experiment is contained in the likelihood function", but reinterprets it as defining a consonant belief function. Combining this belief function by Dempster's rule with a Bayesian prior yields the Bayesian posterior distribution, which ensures compatibility with Bayesian inference. An important advantage of the belief function approach, however, is that it allows the statistician to use either a weaker form of prior information,[1] as a general belief function, or even no prior information at all (which corresponds to providing a vacuous belief function as prior information).

In recent work [36], we have extended the likelihood-based approach to prediction problems. Prediction can be defined as the task of making statements about data that have not yet been observed. Assume, for instance, that we have drawn y balls out of n draws with replacement from an urn contain an unknown proportion θ of black balls, and a proportion $1-\theta$ of white balls. Let z be a binary variable defined by $z=1$ if the next ball to be drawn is black, and $z=0$ otherwise. Guessing the value of z is a prediction problem. The general model for such problems involves a pair (y,z) of random quantities whose joint distribution depends on some parameter θ, where y is observed but z is not yet observed. In [36], we proposed a solution to this problem, using the likelihood-based approach outlined above, and we applied it to a very specific model in the field of marketing econometrics. The same approach was used in [67] to calibrate a certain kind of binary classifiers. In this paper, we further explore this method by proving that, under some mild assumptions, the predictive belief function converges, in some sense, to the true probability distribution of the not-yet observed data. We also address describe several simulation and approximation techniques to estimate the predictive belief function or an outer approximation thereof. Finally, we illustrate the practical application of the method using multiple linear regression. In particular, we show that the *ex ante* forecasting problem has a natural and simple solution using our approach.

The rest of this paper is organized as follows. Some background on belief functions will first be given in Section 2. The estimation and prediction methods will then be presented, respectively, in Sections 3 and 4. The application to linear regression will then be studied in Section 5. Finally, Section 6 will conclude the paper.

2. Background on belief functions

Most applications of Dempster–Shafer theory use belief functions defined on finite sets [58]. However, in statistical models, the parameter and sample spaces are often infinite. To make the paper self-contained, we will recall some basic definitions and results on belief functions defined on arbitrary spaces (finite or not).

2.1. Belief function induced by a source

Let $(\Omega, \mathcal{B})$ be a measurable space. A *belief function* on $\mathcal{B}$ is a mapping $Bel : \mathcal{B} \rightarrow [0,1]$ verifying the following three conditions:

1. $Bel(\emptyset) = 0$;
2. $Bel(\Omega) = 1$;
3. For any $k \geq 2$ and any collection $B_1, \ldots, B_k$ of elements of $\mathcal{B}$,

$$Bel\left(\bigcup_{i=1}^{k} B_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} B_i\right). \tag{1}$$

Similarly, a *plausibility function* can be defined as a function $Pl : \mathcal{B} \rightarrow [0,1]$ such that:

1. $Pl(\emptyset) = 0$;
2. $Pl(\Omega) = 1$;
3. For any $k \geq 2$ and any collection $B_1, \ldots, B_k$ of elements of $\mathcal{B}$,

[1] A similar goal is pursued by robust Bayes [7] and imprecise probability approaches (see, e.g., [42,48]), which attempt to represent weak prior information by sets of probability measures. A comparison with these alternative approaches is beyond the scope of this paper.