ARTICLE IN PRESS
JID:IJA AID:8133 /FLA [m3G; v1.224; Prn:19/10/2017; 15:07] P.1 (1-23)
International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Frequency-calibrated belief functions: Review and new insights

Thierry Denœux [a,b,*], Shoumei Li [a]

[a] Beijing University of Technology College of Applied Sciences, Department of Statistics, Beijing, China
[b] Sorbonne Universités, Université de Technologie de Compiègne, CNRS UMR 7253 Heudiasyc, Compiègne, France

ARTICLE INFO

Article history:
Received 13 June 2017
Received in revised form 24 September 2017
Accepted 16 October 2017
Available online xxxx

Keywords:
Dempster–Shafer theory
Evidence theory
Statistical inference
Estimation
Prediction

ABSTRACT

Starting with Dempster's seminal work, several approaches to statistical inference based on belief functions have been proposed. Some of these approaches can be seen as implementing some form of prior-free Bayesian inference, while some others put the emphasis on long-run frequency properties and are more related to classical frequentist methods. This paper focusses on the latter class of techniques, which have been developed independently and had not been put in perspective until now. Existing definitions for frequency-calibrated belief functions as well as corresponding construction methods are reviewed, and some new notions and techniques are introduced. The connections with other frequentist notions such as confidence distributions and confidence curves are also explored. The different construction techniques are illustrated on simple inference problems, with a focus on interpretation and implementation issues.

1. Introduction

The theory of belief functions, or Dempster–Shafer (DS) theory [1–3], is a general framework for reasoning under uncertainty. Its success in applications [4,5] owes much to its flexibility and its ability to represent and combine elementary items of evidence in a wide range of problems. The validity and cogency of the inferences and decisions performed within this theory thus crucially depend on the validity of the operational methods used for expressing uncertain and partial evidence in the formalism of belief functions.

The first category of problems to which belief functions have been applied is parametric statistical inference. Dempster's approach [6,7] extends Fisher's fiducial inference by making use of a structural equation $X = \varphi(\theta, U)$, which relates the observable data X, an unknown parameter $\theta \in \Theta$ and an auxiliary random variable U with known distribution. After observing $X = x$, the random set $\Gamma(U, x) = \{\theta \in \Theta \mid x = \varphi(\theta, U)\}$ defines a belief function on Θ. Although conceptually simple and elegant, this method leads to intricate computations for most but very simple inference problems. An alternative approach, introduced by Shafer [2], is based on the construction of a consonant belief function directly from the likelihood function. This approach, in line with likelihood inference [8–10], is much easier to implement than Dempster's method [11–13] and it can be justified from basic axiomatic requirements [14,15]. Both Dempster's method and the likelihood-based approach are compatible with Bayesian inference, in the sense that combining the data-conditional belief function with a prior probability distribution using Dempster's rule of combination [1,2] yields the Bayesian posterior distribution. These methods can

* Corresponding author at: Sorbonne Universités, Université de Technologie de Compiègne, CNRS UMR 7253 Heudiasyc, Compiègne, France.
E-mail addresses: thierry.denoeux@utc.fr (T. Denœux), lisma@bjut.edu.cn (S. Li).

https://doi.org/10.1016/j.ijar.2017.10.013
0888-613X/© 2017 Published by Elsevier Inc.

Table 1
Summary of the different definitions of frequency-calibrated belief functions.

	Estimation	Prediction
$100(1-\alpha)\%$ confidence belief function	*Ref.* This paper *Def.* $\forall\theta$, $\mathbb{P}_{\boldsymbol{X}\mid\theta}\left(pl_{\theta\mid\boldsymbol{X}}(\theta)=1\right)\geq 1-\alpha$	*Ref.* [16] [17] *Def.* $\forall\theta\in\Theta$, $$\mathbb{P}_{\boldsymbol{X}\mid\theta}\left(Bel_{Y\mid\boldsymbol{X}}(A)\leq\mathbb{P}_{Y\mid\boldsymbol{X},\theta}(A),\ \forall A\in\mathcal{B}_Y\right)\geq 1-\alpha$$
Confidence structure	*Ref.* [20][21] *Def.* $Bel_{\theta\mid\boldsymbol{x}}(H)=\mathbb{P}_U(\Gamma(U,\boldsymbol{x})\subseteq H)$, with $\Gamma(U,\boldsymbol{X})$ such that $\forall\theta$, $\forall A\in\mathcal{B}_U$, $$\mathbb{P}_{\boldsymbol{X}\mid\theta}\left\{\theta\in\bigcup_{u\in A}\Gamma(u,\boldsymbol{X})\right\}\geq\mathbb{P}_U(A)$$	*Ref.* This paper *Def.* $Bel_{Y\mid\boldsymbol{x}}(B)=\mathbb{P}_U(\Gamma(U,\boldsymbol{x})\subseteq B)$, with $\Gamma(U,\boldsymbol{X})$ such that $\forall\theta$, $\forall A\in\mathcal{B}_U$, $$\mathbb{P}_{\boldsymbol{X},Y\mid\theta}\left\{Y\in\bigcup_{u\in A}\Gamma(u,\boldsymbol{X})\right\}\geq\mathbb{P}_U(A)$$
Valid belief function	*Ref.* [22,23,18,19] *Def.* $\forall\alpha\in(0,1)$, $\forall\theta$, $$\mathbb{P}_{\boldsymbol{X}\mid\theta}\left\{pl_{\theta\mid\boldsymbol{X}}(\theta)\leq\alpha\right\}\leq\alpha$$	*Ref.* [24], [19, Chapter 9] *Def.* $\forall\alpha\in(0,1)$, $\forall\theta\in\Theta$, $$\mathbb{P}_{\boldsymbol{X},Y\mid\theta}\left\{pl_{Y\mid\boldsymbol{X}}(Y)\leq\alpha\right\}\leq\alpha$$

thus be seen as implementing a form of prior-free generalization of Bayesian inference. They lack, however, the frequency calibration properties expected by many statisticians.

In recent years, several attempts have been made to blend belief function inference with frequentist ideas. In [16,17], the first author proposed a notion of predictive belief function, which under repeated sampling is less committed than the true probability distribution of interest with some prescribed probability. Using different ideas, Liu and Martin developed the Inferential Model (IM) approach, which can be seen as a modification of Dempster's model that produces *credible*, or *valid* belief functions with well defined frequentist properties [18,19]. Yet another notion is the theory of Confidence Structures proposed by Balch [20,21] as an extension of confidence distributions. The common idea underlying these three distinct approaches is to constrain degrees of belief to meet some properties in a repeated sampling framework. Each of them allows one to construct a different kind of "frequency-calibrated" belief function, i.e., a belief function based on observed data, which assigns degrees of belief to which a frequentist meaning can be attached. These approaches seem to have been developed independently and they have not been compared from a conceptual or practical point of view. The objective of this paper is to fill this gap by reviewing the different notions of frequency-calibrated belief functions, relating them to statistical notions developed in other contexts such as confidence distributions or confidence curves, and describing some simple procedures (including some original ones) for generating such belief functions in realistic statistical inference situations. We also introduce some new notions, namely: the extension of predictive belief functions at a given confidence level to estimation, and the extension of confidence structures to prediction. These new notions allow us to provide the global picture shown in Table 1, in which three different principles are applied both to estimation and to prediction. The rest of this paper will be devoted to detailed explanation and in-depth discussion of the notions summarized in this table. The emphasis will be on underlying principles, with the objective of bringing recent results and ideas to the attention of a large audience of researchers interested in belief functions. Accordingly, technicalities will be avoided by considering only simple statistical models and inference problems.

We will assume that the reader already has some familiarity with the theory of belief functions. A concise exposition of the main relevant notions can be found in [25], for instance. The three approaches mentioned will then be described sequentially. Section 2 will be devoted to the frequentist notion of predictive belief functions introduced in [16]. Confidence Structures and valid belief functions will then be reviewed, respectively, in Section 3 and 4. A summary and some conclusions will be provided in Section 5.

Notations and terminology. Before entering into the description of different notions and methods related to frequency-calibrated belief functions, let us first clarify the notations and terminology. Throughout this paper, we will denote by $\boldsymbol{x}$ the observed data, assumed to be a realization of a random vector $\boldsymbol{X}$ with sample space $\Omega_{\boldsymbol{X}}$. The σ-algebra $\mathcal{B}_{\boldsymbol{X}}$ on $\Omega_{\boldsymbol{X}}$ will be the power set when $\Omega_{\boldsymbol{X}}$ is finite, and the Borel σ-algebra on $\Omega_{\boldsymbol{X}}$ when $\Omega_{\boldsymbol{X}}=\mathbb{R}^n$. In general, random variables and their realizations will be denoted by uppercase and lowercase letters, respectively. We will consider a parametric model $\boldsymbol{X}\sim\mathbb{P}_{\boldsymbol{X}\mid\theta}$, where $\theta\in\Theta$ is a fixed but unknown parameter. An *estimative belief function* $Bel_{\theta\mid\boldsymbol{x}}$ is a data-conditional belief function on Θ, defined after observing the data $\boldsymbol{x}$. It basically encodes statistical evidence about θ. Given a measurable subset $H\subset\Theta$, the quantity $Bel_{\theta\mid\boldsymbol{x}}(H)$ is interpreted as one's degree of belief in the proposition $\theta\in H$, based on the evidence $\boldsymbol{X}=\boldsymbol{x}$. It is a function $h(\boldsymbol{x})$ of $\boldsymbol{x}$. The notation $Bel_{\theta\mid\boldsymbol{X}}(H)$ stands for the random variable $h(\boldsymbol{X})$.

As opposed to estimation, *prediction* is concerned with the determination of a random quantity. Typically, we have a pair of random variables $(\boldsymbol{X},\boldsymbol{Y})$, where $\boldsymbol{X}$ is the (past) observed data and $\boldsymbol{Y}$ is the (future) not-yet observed data taking values in the probability space $(\Omega_{\boldsymbol{Y}},\mathcal{B}_{\boldsymbol{Y}})$. As before, $\mathcal{B}_{\boldsymbol{Y}}$ will be $2^{\Omega_{\boldsymbol{Y}}}$ in the finite case and the Borel σ-algebra in the continuous

Please cite this article in press as: T. Denœux, S. Li, Frequency-calibrated belief functions: Review and new insights, Int. J. Approx. Reason. (2017), https://doi.org/10.1016/j.ijar.2017.10.013