Frequency-Calibrated Belief Functions: Review and New Insights

Thierry Denœux[a,*], Shoumei Li[b]

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

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 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.

Keywords: Dempster-Shafer Theory, Evidence Theory, Statistical inference, Estimation, Prediction.

1. Introduction

The theory of belief functions, or Dempster-Shafer (DS) theory [1, 2, 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

[*]Corresponding author.
Email addresses: `thierry.denoeux@utc.fr` (Thierry Denœux), `lisma@bjut.edu.cn` (Shoumei Li)

of a consonant belief function directly from the likelihood function. This approach, in line with likelihood inference [8, 9, 10], is much easier to implement than Dempster's method [11, 12, 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 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. he 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

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.*. $\forall \alpha \in (0,1)$, $\forall \theta \in \Theta$, $$\mathbb{P}_{\boldsymbol{X},Y\mid\theta}\left\{pl_{Y\mid\boldsymbol{X}}(Y) \leq \alpha\right\} \leq \alpha$$

a parametric model $\boldsymbol{X} \sim \mathbb{P}_{\boldsymbol{X}|\theta}$, where $\theta \in \Theta$ is a fixed but unknown parameter. An *estimative belief function* $Bel_{\theta|\boldsymbol{x}}$ 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 $Bel_{\theta|\boldsymbol{x}}(H)$ 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 $Bel_{\theta|\boldsymbol{X}}(H)$ 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 case. The joint distribution $\mathbb{P}_{\boldsymbol{X},\boldsymbol{Y}|\theta}$ depends on parameter θ. After observing $\boldsymbol{X} = \boldsymbol{x}$, we wish to make statements about a future realization of $\boldsymbol{Y}$. A *predictive belief function* $Bel_{\boldsymbol{Y}|\boldsymbol{x}}$ is a data-conditional belief function from $\mathcal{B}_{\boldsymbol{Y}}$ to $[0,1]$ quantifying the uncertainty on $\boldsymbol{Y}$ after observing the evidence $\boldsymbol{x}$. To simplify the exposition, we will consider throughout this paper the special case where the future data is a real random variable and we will denote it by Y.

In the next sections, we will assume that we repeatedly draw realizations $\boldsymbol{x}$ of the data and compute estimative or predictive belief functions using some procedures. We will then consider various requirements that can be imposed on such procedures, so that the resulting belief assessments based on any given realization $\boldsymbol{x}$ can be trusted.

2. Belief functions at a given confidence level

In this section, the definition of a predictive belief function at a given confidence level will first be recalled in Section 2.1. Techniques to generate such belief functions will be reviewed in Section 2.2 and a counterpart for estimation problems will be introduced in Section 2.3.

2.1. Predictive belief functions

The notion of predictive belief function introduced in [16] is based on the following idea. If we knew the conditional distribution $\mathbb{P}_{Y|\boldsymbol{x},\theta}$ of Y given $\boldsymbol{X} = \boldsymbol{x}$, then it would be natural to equate our degrees of belief $Bel_{Y|\boldsymbol{x}}(A)$ with degrees of chance $\mathbb{P}_{Y|\boldsymbol{x},\theta}(A)$ for any event A in Ω_Y, i.e., we would impose

$$Bel_{Y|\boldsymbol{x}} = \mathbb{P}_{Y|\boldsymbol{x},\theta}.$$

In real situations, however, we only have limited information about $\mathbb{P}_{Y|\boldsymbol{x},\theta}$ in the form of the observed data $\boldsymbol{x}$. Our predictive belief function should thus be less informative, or *less committed* [26] than $\mathbb{P}_{Y|\boldsymbol{x},\theta}$, which can be expressed by the following inequalities

$$Bel_{Y|\boldsymbol{x}}(A) \leq \mathbb{P}_{Y|\boldsymbol{x},\theta}(A) \tag{1}$$

for all measurable event $A \in \mathcal{B}_Y$. Property (1) can be equivalently expressed using the dual plausibility function $Pl_{Y|\boldsymbol{x}}(A) = 1 - Bel_{Y|\boldsymbol{x}}(\overline{A})$ as

$$Pl_{Y|\boldsymbol{x}}(A) \geq \mathbb{P}_{Y|\boldsymbol{x},\theta}(A) \tag{2}$$

for all event A. However, conditions (1) and (2) are generally too strict to be of any practical value, as they can be guaranteed only for the vacuous belief function verifying $Bel_{Y|\boldsymbol{x}}(A) = 0$ for all $A \subset \Omega_Y$ and $Pl_{Y|\boldsymbol{x}}(A) = 1$ for all $A \neq \emptyset$. For instance, consider the case where X has a binomial distribution $\mathcal{B}(n, \theta)$ and Y has a Bernoulli distribution $\mathcal{B}(1, \theta)$, with X and Y independent. Having observed $X = x$, no value of θ in $(0, 1)$ can be ruled out. Consequently (1) implies $Bel(\{0\}) \leq 1-\theta$ and $Bel(\{1\}) \leq \theta$ for any $\theta \in (0, 1)$, a condition only verified for the vacuous belief function defined by $Bel(\{0\}) = Bel(\{1\}) = 0$.

The solution proposed in [16] is to weaken condition (1) by imposing only that it hold for at least a proportion $1 - \alpha \in (0, 1)$ of the samples $\boldsymbol{x}$, under repeated sampling. We then have the following requirement,

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left(Bel_{Y|\boldsymbol{X}}(A) \leq \mathbb{P}_{Y|\boldsymbol{X},\theta}(A),\ \forall A \in \mathcal{B}_Y\right) \geq 1 - \alpha, \tag{3}$$

for all $\theta \in \Theta$. In (3), both $Bel_{Y|\boldsymbol{X}}(A)$ and $\mathbb{P}_{Y|\boldsymbol{X},\theta}(A)$ are random variables, defined as functions of $\boldsymbol{X}$. A belief function verifying (3) is called a *predictive belief function at confidence level* $1 - \alpha$. It is an approximate $1 - \alpha$-level predictive belief function if Property (3) holds only in the limit as the sample size tends to infinity.

Example 1. *Consider again the binomial case, with* X *and* Y *independently distributed according to binomial* $\mathcal{B}(n, \theta)$ *and Bernoulli* $\mathcal{B}(1, \theta)$ *distributions, respectively. The Clopper-Pearson confidence interval [27] on* θ *at level* $1 - \alpha$ *is*

$$\mathrm{qBeta}_{\mathrm{x,n-x+1}}\left(\frac{\alpha}{2}\right) \leq \theta \leq \mathrm{qBeta}_{\mathrm{x+1,n-x}}\left(1 - \frac{\alpha}{2}\right), \tag{4}$$

where $\mathrm{qBeta}_{\mathrm{a,b}}(\mathrm{p})$ *is the p-th quantile from a beta distribution with shape parameters a and b. This interval is known to be conservative. With probability at least* $1 - \alpha$*, we thus have simultaneously* $\mathrm{qBeta}_{\mathrm{x,n-x+1}}(\alpha/2) \leq \theta$ *and* $1 - \mathrm{qBeta}_{\mathrm{x+1,n-x}}(1 - \alpha/2) \leq 1 - \theta$*. Consider the following mass function on* $\Omega_Y = \{0, 1\}$*,*

$$m_{Y|x}(\{1\}) = \mathrm{qBeta}_{\mathrm{x,n-x+1}}(\alpha/2) \tag{5a}$$

$$m_{Y|x}(\{0\}) = 1 - \mathrm{qBeta}_{\mathrm{x+1,n-x}}(1 - \alpha/2) \tag{5b}$$

$$m_{Y|x}(\{0, 1\}) = \mathrm{qBeta}_{\mathrm{x+1,n-x}}(1 - \alpha/2) - \mathrm{qBeta}_{\mathrm{x,n-x+1}}(\alpha/2). \tag{5c}$$

We have $Bel_{Y|x}(\{1\}) = m_{Y|x}(\{1\})$ *and* $Bel_{Y|x}(\{0\}) = m_{Y|x}(\{0\})$*. Consequently, the condition*

$$Bel_{Y|x}(\{1\}) \leq \theta \text{ and } Bel_{Y|x}(\{0\}) \leq 1 - \theta$$

holds with probability at least $1 - \alpha$*, and* $Bel_{Y|x}$ *is a predictive belief function for* Y *at level* $1 - \alpha$*. For* $\alpha = 0.05$*,* $n = 20$ *and* $x = 5$*, we get* $Bel_{Y|x}(\{1\}) = m_{Y|x}(\{1\}) \approx 0.0866$ *and* $Bel_{Y|x}(\{0\}) = m_{Y|x}(\{0\}) \approx 0.509$*.* □

2.2. Practical construction

Discrete case. In [16], the approach of Example 1 is generalized to compute predictive belief functions in the case where $X_1, \ldots, X_n, Y$ are i.i.d. according to a *discrete distribution* with finite support $\Omega_Y = \{\xi_1, \ldots, \xi_K\}$. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $\theta = (\theta_1, \ldots, \theta_K)$ with $\theta_k = \mathbb{P}(Y = \xi_k)$, $k = 1, \ldots, K$. The method is based on simultaneous confidence intervals for multinomial proportions θ_k [28, 29], which define a lower probability measure $\underline{P}$ such that

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left(\underline{P}(A) \leq \mathbb{P}_{Y|\boldsymbol{X},\theta}(A),\ \forall A \in \mathcal{B}_Y\right) \geq 1 - \alpha.$$

Usually, $\underline{P}$ is not a belief function (except for $K = 2$ and $K = 3$), but we can construct the best approximating belief function $Bel_{Y|\boldsymbol{x}}$ such that $Bel_{Y|\boldsymbol{x}} \leq \underline{P}$ using linear optimization. Analytical formula are given for the case where the elements of Ω_Y are ordered and the focal sets of $Bel_{Y|\boldsymbol{x}}$ are restricted to be intervals. A similar approach is proposed in [30] to construct a predictive *possibility distribution* (equivalent to a consonant belief function).

Confidence bands. In [17], a method was proposed to construct predictive belief functions at a given confidence level based on confidence bands. The method can be used in the particular situation where the observed data is an n-sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ and $X_1, \ldots, X_n, Y$ are i.i.d. according to a *continuous distribution*. A *confidence band* for Y at level $\alpha \in (0,1)$ [31, page 334] is a pair of cumulative distribution functions (cdfs) $\underline{F}(\cdot; \boldsymbol{x})$ and $\overline{F}(\cdot; \boldsymbol{x})$ depending on the observed data $\boldsymbol{x}$, such that

$$\mathbb{P}_{\boldsymbol{X}} \left\{ \underline{F}(y; \boldsymbol{X}) \leq F_Y(y) \leq \overline{F}(y; \boldsymbol{X}), \ \forall y \in \Omega_Y \right\} \geq 1 - \alpha.$$

A pair of cdfs $(\underline{F}, \overline{F})$ such that $\underline{F} \leq \overline{F}$, called a *probability box* or "p-box" [32], is a common representation of a set of probability distributions. It was shown in [33] that the lower envelope of the family of probability distributions represented by a p-box is a belief function Bel. This belief function is induced by a random interval $\Pi(U)$ with bounds $\overline{F}^{-1}(U)$ and $\underline{F}^{-1}(U)$, where $\overline{F}^{-1}$ and $\underline{F}^{-1}$ are, respectively, the generalized inverses of $\underline{F}$ and $\overline{F}$, and U has a standard uniform distribution. The random interval is closed if $\underline{F}$ is left-continuous and $\overline{F}$ is right-continuous. The following equalities hold for any event A:

$$Bel(A) = \mathbb{P}_U \big(\Pi(U) \subseteq A\big) \tag{6a}$$

$$Pl(A) = \mathbb{P}_U \big(\Pi(U) \cap A \neq \emptyset\big). \tag{6b}$$

In particular,

$$Bel((-\infty, y]) = \underline{F}(y), \quad Pl((-\infty, y]) = \overline{F}(y)$$

for any $y \in \mathbb{R}$ and

$$Bel([y_1, y_2]) = \Big(\underline{F}(y_2) - \overline{F}(y_1)\Big)_+ \tag{7a}$$

$$Pl([y_1, y_2]) = \Big(\overline{F}(y_2) - \underline{F}(y_1)\Big) \tag{7b}$$

for any $y_1 \leq y_2$, where $(\cdot)_+$ denotes the positive part.

An immediate consequence of the above results is that the belief function induced by a confidence band on Y at level $1 - \alpha$ is a predictive belief function with the same confidence level [17]. Any method to construct a confidence band thus yields a predictive belief function. One such method in a nonparametric setting is based on Kolmogorov's statistic [34]:

$$D_n = \sup_y |\widehat{F}(y; \boldsymbol{X}) - F_Y(y)|,$$

where $\widehat{F}(\cdot; \boldsymbol{X})$ is the empirical cdf of the sample $\boldsymbol{X} = (X_1, \ldots, X_n)$. The resulting confidence band [35, page 481] is

$$\underline{F}(y; \boldsymbol{X}) = \max(0, \widehat{F}(y; \boldsymbol{X}) - d_{n,1-\alpha}), \tag{8a}$$

$$\overline{F}(y; \boldsymbol{X}) = \min(1, \widehat{F}(y; \boldsymbol{X}) + d_{n,1-\alpha}).,\tag{8b}$$

where $d_{n,1-\alpha}$ is the $1 - \alpha$ quantile of D_n. We note that, in this case, both $\underline{F}$ and $\overline{F}$ are right-continuous steps functions. The upper cdf $\overline{F}$ can be replaced by the left-continuous function taking the same values everywhere except at sample points.

In a parametric setting, it is possible to compute a confidence band by determining lower and upper bounds of the cdf $F_{Y|\theta}$ when θ varies in a confidence region. For instance, Cheng and Iles

[36] give closed-form expressions for the upper and lower cdfs $\underline{F}$ and $\overline{F}$ in the case of general location-scale parametric model of the form:

$$F_{Y|\theta}(y) = G\left(\frac{y-\mu}{\sigma}\right),$$

where G is a fixed cdf, μ and σ are the unknown location and scale parameters, and $\theta = (\mu, \sigma)$. The confidence band is based on the Maximum Likelihood Estimators (MLEs) of μ and σ and on the Fisher information matrix. It thus has approximately the prescribed confidence level for fixed sample size n. Cheng and Iles [36] give explicit formula for the cases of the normal, lognormal, Gumbel and Weibull distributions.

Use of a structural equation. In this section, we propose a new general method to compute a predictive belief function at a given confidence level, by adapting the method described in [25]. For simplicity, we assume here that $\boldsymbol{X}$ and Y are independent, but the approach can easily be extended to relax this assumption. The method is based on a structural equation of the form

$$Y = \varphi(\theta, U), \tag{9}$$

where U is a pivotal random variable with known distribution [37, 18, 25]. Equation (9) can be obtained by inverting the cdf of Y. More precisely, let us first assume that Y is continuous; we can then observe that $U = F_{Y|\theta}(Y)$ has a standard uniform distribution. Denoting by $F_{Y|\theta}^{-1}$ the inverse of the cdf $F_{Y|\theta}$, we get

$$Y = F_{Y|\theta}^{-1}(U), \tag{10}$$

with $U \sim \mathcal{U}[0,1]$, which has the same form as (9). When Y is discrete, (10) is still valid if $F_{Y|\theta}^{-1}$ now denotes the generalized inverse of F_Y.

Let $C_\alpha(\boldsymbol{X})$ be a confidence region for θ at level $1-\alpha$, and consider the following random set,

$$\Pi(U; \boldsymbol{x}) = \varphi(C_\alpha(\boldsymbol{x}), U). \tag{11}$$

The following theorem states that the belief function induced by the random set (11) is a predictive belief function at level $1-\alpha$.

Theorem 1. *Let $Y = \varphi(\theta, U)$ be a random variable, and $C_\alpha(\boldsymbol{X})$ a confidence region for θ at level $1-\alpha$. Then, the belief function $Bel_{Y|\boldsymbol{x}}$ induced by the random set $\Pi(U; \boldsymbol{x}) = \varphi(C_\alpha(\boldsymbol{x}), U)$ is a predictive belief function at level $1-\alpha$.*

Proof. If $\theta \in C_\alpha(\boldsymbol{x})$, then $\varphi(\theta, U) \in \varphi(C_\alpha(\boldsymbol{x}), U)$ for any U. Consequently, the following implication holds for any measurable subset $A \in \mathcal{B}_Y$, and any $\boldsymbol{x} \in \Omega_{\boldsymbol{X}}$,

$$\varphi(C_\alpha(\boldsymbol{x}), U) \subseteq A \Rightarrow \varphi(\theta, U) \in A.$$

Hence,

$$\mathbb{P}_U(\varphi(C_\alpha(\boldsymbol{x}), U) \subseteq A) \leq \mathbb{P}_U(\varphi(\theta, U) \in A),$$

or, equivalently,

$$Bel_{Y|\boldsymbol{x}}(A) \leq \mathbb{P}_{Y|\theta}(A). \tag{12}$$

Table 2: Likelihood levels $c_{0.05}$ defining approximate 95% confidence regions.

p	1	2	5	10	15
c	0.15	0.05	3.9e-03	1.1e-04	3.7e-06

As (12) holds whenever $\theta \in C_\alpha(\boldsymbol{x})$, and $\mathbb{P}_{\boldsymbol{X}|\theta}(C_\alpha(\boldsymbol{X}) \ni \theta) \geq 1-\alpha$ for all $\theta \in \Theta$, it follows that (12) holds for any measurable event A with probability at least $1-\alpha$, i.e.,

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left(Bel_{Y|\boldsymbol{X}}(A) \leq \mathbb{P}_{Y|\theta}(A),\ \forall A \in \mathcal{B}_Y\right) \geq 1-\alpha, \tag{13}$$

for all $\theta \in \Theta$. □

Belief values $Bel_{Y|\boldsymbol{x}}(A) = \mathbb{P}_U(\varphi(C(\boldsymbol{x}), U) \subseteq A)$ can be estimated by Monte Carlo simulation using a pseudo-random sample $u_1, \ldots, u_N$ of U. The degree of belief $Bel_{Y|\boldsymbol{x}}(A)$ in A can then be estimated by the proportion of u_i such that $\varphi(C_\alpha(\boldsymbol{x}), u_i) \subseteq A$ (see [25] for more details on this method).

In the case where $\boldsymbol{X} = (X_1, \ldots, X_n)$ is iid, the likelihood function $L(\theta; \boldsymbol{x})$ will often provide us with a convenient means to obtain a confidence region on θ [38, 57]. Let

$$pl_{\theta|\boldsymbol{x}}(\theta) = \frac{L(\theta; \boldsymbol{x})}{L(\widehat{\theta}; \boldsymbol{x})} \tag{14}$$

be the relative likelihood function, where $\widehat{\theta}$ is a MLE of θ and it is assumed that $L(\widehat{\theta}; \boldsymbol{x}) < \infty$. From Wilks' theorem [39, 40], we know that, under regularity conditions[1], $-2\log pl_{\theta|\boldsymbol{X}}(\theta_0)$ converges in distribution to a chi square distribution with p degrees of freedom, where p is the dimension of θ, under the hypothesis $\theta = \theta_0$. Consequently, the sets

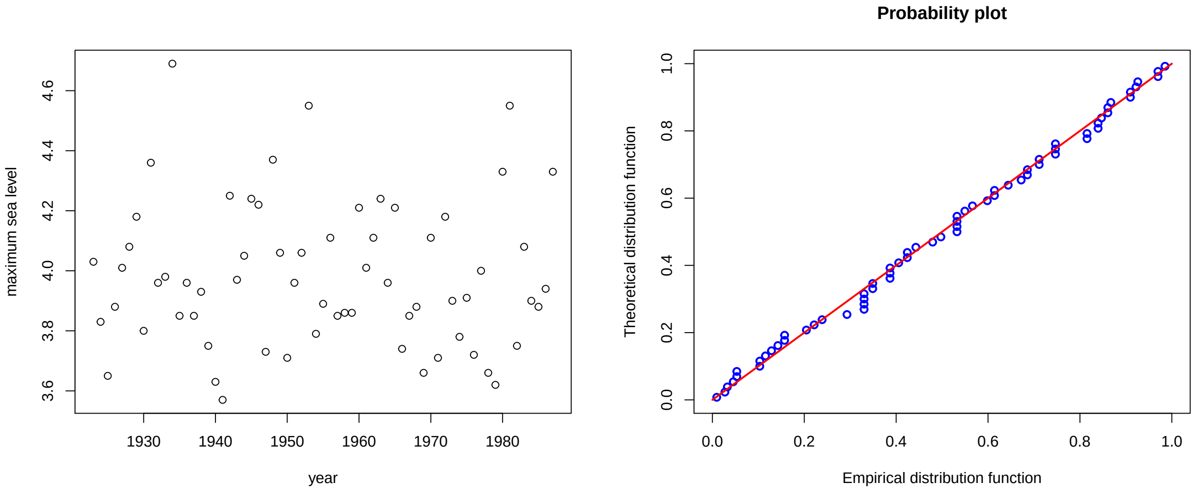
$$C_\alpha(\boldsymbol{X}) = \{\theta \in \Theta | pl_{\theta|\boldsymbol{X}}(\theta) \geq c_\alpha\}, \tag{15}$$

with $c_\alpha = \exp(-0.5\chi^2_{p;1-\alpha})$, are approximate $100(1-\alpha)$% confidence regions, i.e., their coverage probability is approximately equal to $1-\alpha$ for large n. This way of defining a predictive belief function is similar to the one described in [41, 25], except that the relative likelihood function is cut at a fixed level c_α. Table 2 gives the values of $c_{0.05}$ for different values of p. We can see that $c_{0.05}$ decreases quickly with p, which means that the likelihood-based confidence regions and, consequently, the corresponding predictive belief functions will become increasing imprecise as p increases.

Example 2. *The data shown in Figure 1(a) are annual maximum sea-levels recorded at Port Pirie, a location just north of Adelaide, South Australia, over the period 1923-1987 [42]. The probability plot in Figure 1(b) shows a good fit with the Gumbel distribution, with cdf*

$$F_{X|\theta}(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), \tag{16}$$

[1] These conditions are [40]: (1) $\log L(\theta; \boldsymbol{x})$ is regular with respect to its first-order and second-order θ-derivatives in an open set Θ_0 containing the true value θ_0 of θ, and (2) the MLE $\widehat{\theta}$ of θ is unique for $n \geq n_0$, for some $n_0 \in \mathbb{N}^*$.



(a) (b)

Figure 1: Annual maximum sea-levels recorded at Port Pirie over the period 1923-1987 (a), and probability plot for the Gumbel fit to the data (b).

where μ is the mode of the distribution, σ a scale parameter, and $\theta = (\mu, \sigma)$. Suppose that, based on these data, we want to predict the maximum sea level Y in the next $m = 10$ years. Assuming that the sea level distribution will remain unchanged in the near future (i.e., neglecting, for instance, the effect of sea level rise due to climate change), Y is the maximum $X_{(m)}$ of an iid sample $X_1, \ldots, X_m$ from X. Its cdf is, thus,

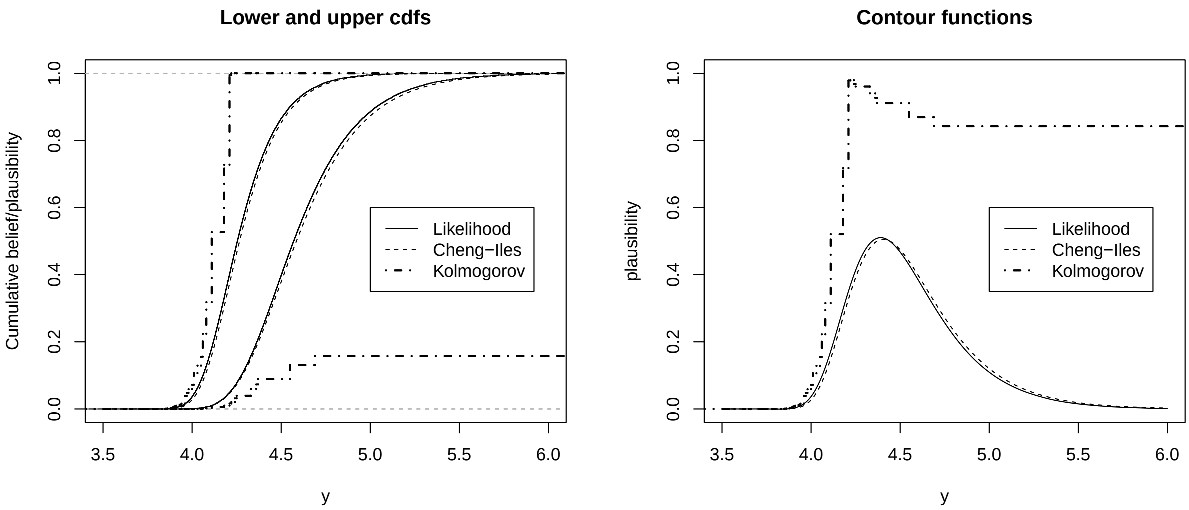
$$F_{Y|\theta}(y) = F_{X_{(m)}|\theta}(y) = F_{X|\theta}(y)^m = \exp\left(-m \exp\left(-\frac{y-\mu}{\sigma}\right)\right). \tag{17}$$

To construct a predictive belief function on Y, we may construct a confidence band $(\underline{F}, \overline{F})$ on X at level $1-\alpha$, and then consider the p-box defined by the lower and upper cdfs $(\underline{F}^m, \overline{F}^m)$. The structural equation (10) becomes, in that case,

$$Y = \mu - \sigma \log\log(U^{-1/m}), \tag{18}$$

with $U \sim \mathcal{U}[0,1]$.

Figures 2(a) and 2(b) show, respectively, the lower and upper cdfs and the contour functions of predictive belief functions at level $1-\alpha = 0.95$, constructed from (1) the Kolmogorov confidence band, (2) Cheng and Iles' confidence band, and (3) the structural equation (18) with the likelihood-based confidence region (15). For this last method, we used Monte Carlo simulation with $N = 1000$ draws. The bounds of each interval $\varphi(C(\boldsymbol{x}), u_i)$ were computed as the minimum and the maximum of $\varphi(\theta, u_i) = \mu - \sigma \log\log(u_i^{-1/m})$ subject to the constraint $pl_{\theta|\boldsymbol{x}}(\theta) \geq c_\alpha$. We can see that the Cheng-Iles and likelihood-based solutions are almost indiscernible. The small difference is due to the use of different confidence regions, which happen to be very similar for sample size $n = 65$. In contrast, the belief function based on the Kolmogorov band is much more imprecise, and probably too little informative to be of any practical use. Given the very good fit of the data with the Gumbel



(a) (b)

Figure 2: Port Pirie data: lower and upper cdfs (a) and contour functions (b) of the predictive belief functions at confidence level 95% computed by the three methods.

distribution as shown in Figure 1(b), the belief functions derived from parametric confidence bands are definitely be preferable in this case.

□

To conclude this section, we can mention a related, but distinct approach to the construction of predictive belief functions proposed in [43]. This approach is based on the notion of *pignistic probability* distribution, a probability distribution derived from a belief function for decision-making [44, 45, 46]. In [43], the authors propose to construct the least committed belief function whose pignistic probability distribution belongs to a confidence set of probability measures containing $\mathbb{P}_Y$ with some confidence level. This belief function can be shown to be consonant. Although this procedure is often simple to implement and has a well-defined justification, it does not produce belief or plausibility statements that are strictly calibrated in term of frequencies. For instance, for any $A \in \mathcal{B}_Y$, it is not possible to directly relate statements like $Bel_{Y|\boldsymbol{x}}(A) = 0.5$ or $Pl_{Y|\boldsymbol{x}}(A) = 0.8$ to frequencies. For this reason, we now favor the approach described previously in this section, which produces belief statements that have a clearer interpretation in terms of frequencies.

2.3. Estimation

The idea underlying the notion of predictive belief function as defined in Section 2.1 can be applied to estimation problems, resulting in a new notion of *estimative belief function*. Estimation can actually be seen as a special case of prediction, in which the predicted quantity is a constant. If we know for sure that $\theta = \theta_0$, then our belief function on θ is the certain belief function verifying

$$Bel^*_{\theta_0}(A) = I(\theta_0 \in A)$$

for all $A \subseteq \Theta$, where $I(\cdot)$ is the indicator function. An *estimative belief function at confidence level* $1 - \alpha$ can be defined as a data-conditional belief function $Bel_{\theta|\boldsymbol{x}}$ on Θ that is less committed than

$Bel^*_{\theta_0}$, for a proportion $1-\alpha$ of the samples:

$$\mathbb{P}_{\boldsymbol{X}|\theta_0}\left(Bel_{\theta|\boldsymbol{X}}(A) \leq Bel^*_{\theta_0}(A),\ \forall A \subseteq \Theta\right) \geq 1-\alpha, \tag{19}$$

for any $\theta_0 \in \Theta$.

Because $Bel^*_{\theta_0}$ can only takes values 0 and 1, the condition $(Bel_{\theta|\boldsymbol{X}}(A) \leq Bel^*_{\theta_0}(A), \forall A \subseteq \Theta)$ is equivalent to $Bel_{\theta|\boldsymbol{X}}(A) = 0$ for all $A \subset \Theta$ such that $\theta_0 \notin A$. From the equality $Pl_{\theta|\boldsymbol{X}}(A) = 1 - Bel_{\theta|\boldsymbol{X}}(\overline{A})$, this condition can be expressed as $Pl_{\theta|\boldsymbol{X}}(A) = 1$ for all $A \subseteq \Theta$ such that $\theta_0 \in A$ or, equivalently $pl_{\theta|\boldsymbol{X}}(\theta_0) = 1$. Condition (19) is thus equivalent to

$$\mathbb{P}_{\boldsymbol{X}|\theta_0}\left(pl_{\theta|\boldsymbol{X}}(\theta_0) = 1\right) \geq 1-\alpha, \tag{20}$$

for any $\theta_0 \in \Theta$. Let $C_\alpha(\boldsymbol{X})$ be the set

$$C_\alpha(\boldsymbol{X}) = \{\theta \in \Theta, pl_{\theta|\boldsymbol{X}}(\theta) = 1\}.$$

Equation (20) expresses that $C_\alpha(\boldsymbol{X})$ is a confidence region at level $1-\alpha$. Conversely, given a confidence region $C_\alpha(\boldsymbol{X})$ at level $1-\alpha$, any belief function verifying $pl_{\theta|\boldsymbol{X}}(\theta) = 1$ for all $\theta \in C_\alpha(\boldsymbol{X})$ is an estimative belief function at confidence level $1-\alpha$. Among all belief functions with this property, the least committed one is the logical belief function focussed on $C_\alpha(\boldsymbol{X})$, defined by

$$Bel_{\theta|\boldsymbol{X}}(A) = I\left(C_\alpha(\boldsymbol{X}) \subseteq A\right),$$

for all $A \subseteq \Theta$.

From these considerations, we can conclude that estimative belief functions at confidence level $1-\alpha$ are belief functions whose contour functions take value one inside some $1-\alpha$-level confidence region $C_\alpha(\boldsymbol{X})$. For any such belief function, there is always a logical belief function that is less committed: it is the belief function that assigns zero plausibilities outside $C_\alpha(\boldsymbol{X})$. However, such a logical belief function does not adequately represent the statistical evidence, as it declares all values of θ outside $C_\alpha(\boldsymbol{X})$ as impossible. A better alternative may be to assign a mass $1-\alpha$ to $C_\alpha(\boldsymbol{X})$ and a mass α to Θ. The corresponding contour function is given by

$$pl_{\theta|\boldsymbol{X}}(\theta) = \begin{cases} 1 & \text{if } \theta \in C_\alpha(\boldsymbol{X}), \\ \alpha & \text{otherwise.} \end{cases} \tag{21}$$

3. Confidence structures

The predictive and estimative belief functions defined in Section 2 have a major drawback: they require the user to specify a confidence level. Following common statistical practice, one may choose standard confidence values such as 95% or 99%, but these values are arbitrary. Confidence structures introduced by Balch [20] overcome this limitation by encoding confidence regions at all levels. This notion will be discussed in Section 3.1 with a focus on parameter estimation, which is the category of problems studied in [20]. An extension to prediction will then be introduced in Section 3.2.

3.1. Confidence Structures for estimation

Definition. A *confidence structure* as defined by Balch [20] is a data-conditional random set that encodes confidence regions at all levels. For any realization $\boldsymbol{x}$ of $\boldsymbol{X}$, this random set defines a belief function $Bel_{\theta|\boldsymbol{x}}$ with well-defined frequentist properties, distinct from those considered in Section 2. More precisely, let $(\Omega_U, \mathcal{B}_U, \mathbb{P}_U)$ be a probability space, $\mathcal{B}_\Theta$ an algebra of subsets of Θ, and Γ a mapping from $\Omega_U \times \Omega_{\boldsymbol{X}}$ to $\mathcal{B}_\Theta$, such that, for any $\boldsymbol{x} \in \Omega_{\boldsymbol{X}}$, the mapping $\Gamma(U, \boldsymbol{x})$ defines a random subset of Θ. The corresponding belief function is defined as

$$Bel_{\theta|\boldsymbol{x}}(B) = \mathbb{P}_U\left(\Gamma(U, \boldsymbol{x}) \subseteq B\right).$$

for all $B \in \mathcal{B}_\Theta$. Mapping Γ defines a confidence structure if the following inequality holds for all $\theta \in \Theta$ and all $A \in \mathcal{B}_U$,

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left\{\theta \in \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})\right\} \geq \mathbb{P}_U(A). \tag{22}$$

Condition (22) expresses that, for any measurable subset A of Ω_U, the random set $C(A, \boldsymbol{X}) = \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})$ is a confidence region for θ at confidence level $\mathbb{P}_U(A)$. Having observed a realization $\boldsymbol{x}$ of $\boldsymbol{X}$, let H be a subset of Θ, and let

$$A = \{u \in \Omega_U, \Gamma(u, \boldsymbol{x}) \subseteq H\}.$$

The degree of belief in H is $Bel_{\theta|\boldsymbol{x}}(H) = \mathbb{P}_U(A)$, and $C(A, \boldsymbol{x}) = \bigcup_{u \in A} \Gamma(u, \boldsymbol{x})$ is included in H. Consequently, H contains a realization of a confidence region with confidence level larger than, or equal to $Bel_{\theta|\boldsymbol{x}}(H)$. Degrees of belief are thus related to confidence levels for a family $\{C(A, \boldsymbol{x})\}$ of confidence regions. In terms of plausibilities, $Pl_{\theta|\boldsymbol{x}}(H) = \alpha$, for instance, means that the complement of H contains a realization of a confidence region at level at least equal to $1 - \alpha$.

Relation with confidence distributions. Confidence structures obviously include confidence distributions as a special case. Following [47, 48] a confidence distribution for a scalar parameter θ is a mapping $F : \Theta \times \Omega_{\boldsymbol{X}} \to [0, 1]$ such that

1. $F(\cdot, \boldsymbol{x})$ is a continuous cdf, for any $\boldsymbol{x} \in \Omega_{\boldsymbol{X}}$;
2. At the true parameter value $\theta = \theta_0$, $F(\theta_0, \boldsymbol{X})$ has a standard uniform distribution $\mathcal{U}[0, 1]$.

Denoting by $F^{-1}(\cdot, \boldsymbol{x})$ the inverse of F with respect to its first argument, it is easy to see that the random set $(-\infty, F^{-1}(\alpha, \boldsymbol{X})]$ is a lower-side confidence interval for θ at level α. Indeed,

$$\mathbb{P}_{\boldsymbol{X}|\theta_0}\left(\theta_0 \leq F^{-1}(\alpha, \boldsymbol{X})\right) = \mathbb{P}_{\boldsymbol{X}|\theta_0}\left(F(\theta_0, \boldsymbol{X}) \leq \alpha\right) = \alpha.$$

Conversely, a confidence distribution is often constructed from a family of one sided confidence intervals $(-\infty, L(\alpha, \boldsymbol{X})]$ at level α: we then set $F(\theta, \boldsymbol{x}) = L^{-1}(\theta, \boldsymbol{x})$.

It is clear that any confidence distribution corresponds to a confidence structure (Figure 3). Indeed, let $\Gamma(u, \boldsymbol{x}) = \{F^{-1}(u, \boldsymbol{x})\}$ and $U \sim \mathcal{U}[0, 1]$. For any measurable subset A of $[0, 1]$, we have

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left(\theta \in \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})\right) = \mathbb{P}_{\boldsymbol{X}|\theta}\left(F(\theta, \boldsymbol{X}) \in A\right) = \mathbb{P}_U(A).$$

The usefulness of the notion of confidence structure becomes apparent when we consider problems for which the notion of confidence distribution is not easily applicable, i.e., those involving discrete observations or a multidimensional parameter. Balch [20] proposed some methods to construct confidence structures for such problems. They are briefly reviewed below.

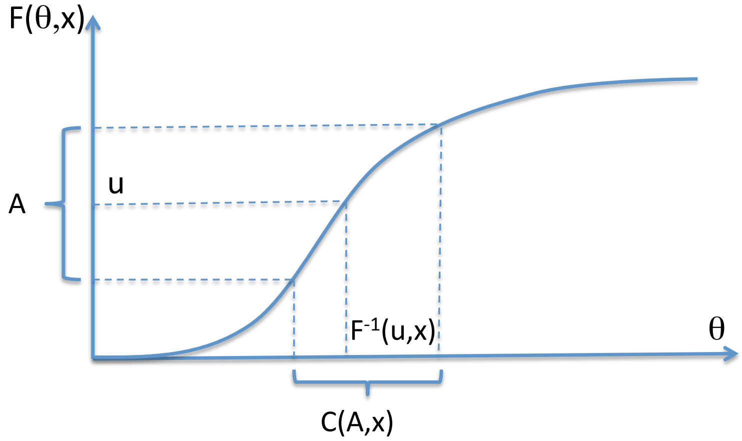


Figure 3: Confidence distribution and relation with the notion of confidence structure.

Pivoting the cdf. In the discrete case, there is no exact confidence distribution [48], but a confidence structure can be constructed using a p-box, resulting in a "C-box" [21]. Balch [20] describes a general method based on a technique for "pivoting the cdf" [49, page 434]. This method can be used when θ is a scalar parameter and there is a discrete statistic T such that $F_{T|\theta}(t)$ is monotone in θ for all t. If $F_{T|\theta}(t)$ is increasing in θ, then we can consider the following mapping,

$$\Gamma(u,t) = \left\{\theta \in \Theta \mid 1 - G_{T|\theta}(t) \leq u \leq F_{T|\theta}(t)\right\}, \tag{23}$$

where $G_{T|\theta}(t) = \mathbb{P}_{T|\theta}(T \geq t)$, and $U \sim \mathcal{U}[0,1]$. The induced confidence structure corresponds to a p-box with $\underline{F}(\theta) = 1 - G_{T|\theta}(t)$ and $\overline{F}(\theta) = F_{T|\theta}(t)$. If $F_{T|\theta}(t)$ is decreasing in θ, then we can choose the following mapping,

$$\Gamma(u,t) = \left\{\theta \in \Theta \mid 1 - F_{T|\theta}(t) \leq u \leq G_{T|\theta}(t)\right\}, \tag{24}$$

which again defines a p-box.

To show that the mappings (23) or (24) generate a confidence structure, let us consider an interval $A = [\alpha_1, \alpha_2]$, with $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ (see Figure 4). Consider, for instance, the mapping (24). We have

$$C(A,t) = \bigcup_{u \in A} \Gamma(u,t) = \left\{\theta \in \Theta \mid 1 - F_{T|\theta}(t) \leq \alpha_2\right\} \setminus \left\{\theta \in \Theta \mid G_{T|\theta}(t) \leq \alpha_1\right\}.$$

Now, both $F_{T|\theta}(T)$ and $G_{T|\theta}(T)$ are stochastically greater than $\mathcal{U}[0,1]$, i.e., $\mathbb{P}_{T|\theta}(F_{T|\theta}(T) \leq \alpha) \leq \alpha$ and $\mathbb{P}_{T|\theta}(G_{T|\theta}(T) \leq \alpha) \leq \alpha$ for all $\alpha \in [0,1]$ (see [49, page 434]). Hence,

$$\mathbb{P}_{T|\theta}\left(1 - F_{T|\theta}(t) \leq \alpha_2\right) \geq \alpha_2$$

and

$$\mathbb{P}_{T|\theta}\left(G_{T|\theta}(t) \leq \alpha_1\right) \leq \alpha_1$$

Consequently, we have

$$\mathbb{P}_{T|\theta}\left\{\theta \in \bigcup_{u \in A} \Gamma(u,T)\right\} \geq \alpha_2 - \alpha_1 = \mathbb{P}_U(A).$$

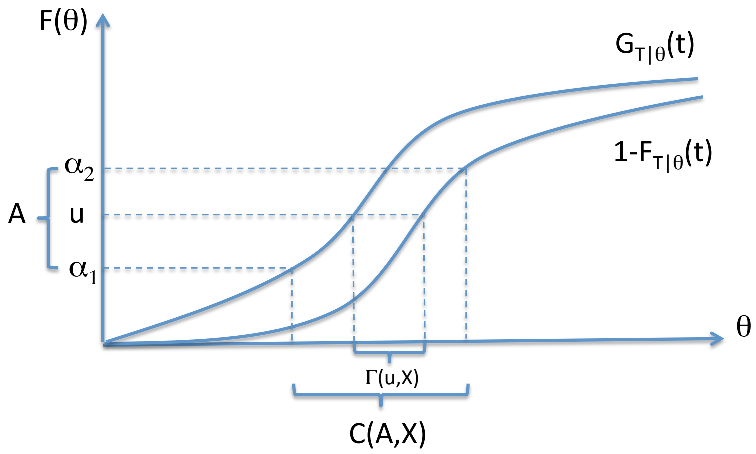


Figure 4: Constructing a confidence structure by the "pivoting the cdf" method.

Example 3. *Consider the case where X has a binomial distribution $\mathcal{B}(n, \theta)$, and take $T = X$. Here,*

$$F_{X|\theta}(x) = 1 - \mathrm{pBeta}_{\mathrm{x+1,n-x}}(\theta), \tag{25}$$

where $\mathrm{pBeta}_{\mathrm{x+1,n-x}}$ *is the beta cdf with shape parameters $x+1$ and $n-x$. It is decreasing in θ, so we select mapping (24). We thus get a C-box with bounding functions*

$$\underline{F}_{\theta|x}(\theta) = 1 - F_{X|\theta}(x) = \mathrm{pBeta}_{\mathrm{x+1,n-x}}(\theta) \tag{26a}$$

and

$$\overline{F}_{\theta|x}(\theta) = G_{X|\theta}(x) = 1 - F_{X|\theta}(x-1) = \mathrm{pBeta}_{\mathrm{x,n-x+1}}(\theta). \tag{26b}$$

We note that the confidence interval $C(A, x)$ for $A = [\alpha/2, 1-\alpha/2]$ is the Clopper-Pearson interval (4). Figure 5 shows an example of a C-box for $n = 10$ and $x = 3$.

Confidence regions. Balch [20] describes a method to construct confidence structures from p-values, which can be equivalently described in terms of confidence regions (Figure 6). Let $C_\alpha(\boldsymbol{X})$, $\alpha \in [0, 1]$ be a *nested* family of confidence regions, such that

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left(\theta \in C_\alpha(\boldsymbol{X})\right) \geq 1 - \alpha \tag{27}$$

and, for any (α, α'),

$$\alpha < \alpha' \Rightarrow C_\alpha(\boldsymbol{X}) \supseteq C_{\alpha'}(\boldsymbol{X}). \tag{28}$$

Consider the confidence structure with multivalued mapping

$$\Gamma(u, \boldsymbol{x}) = C_{1-u}(\boldsymbol{x})$$

and $U \sim \mathcal{U}[0, 1]$. For any measurable $A \subseteq [0, 1]$, we have

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left(\theta \in \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})\right) = \mathbb{P}_{\boldsymbol{X}|\theta}\left(\theta \in C_{\inf(A)}(\boldsymbol{X})\right) = 1 - \inf(A) \geq \mathbb{P}_U(A).$$

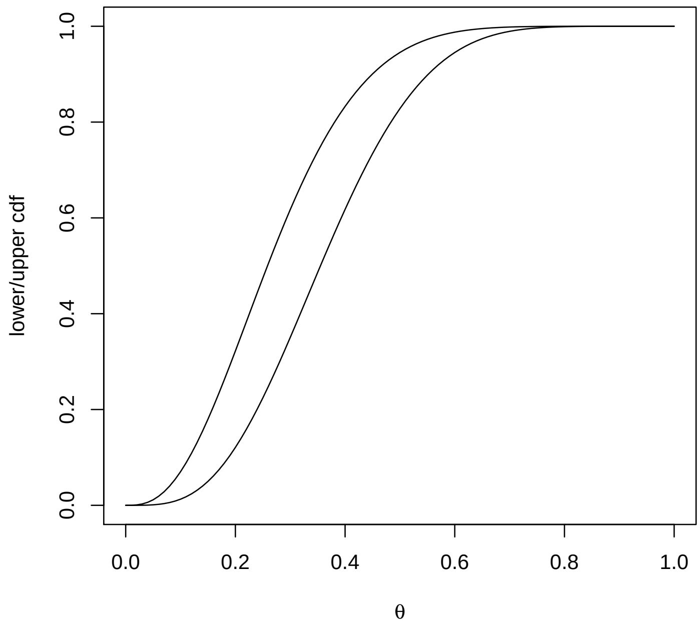


Figure 5: C-box for a binomial probability with $x = 3$ successes in $n = 10$ trials.

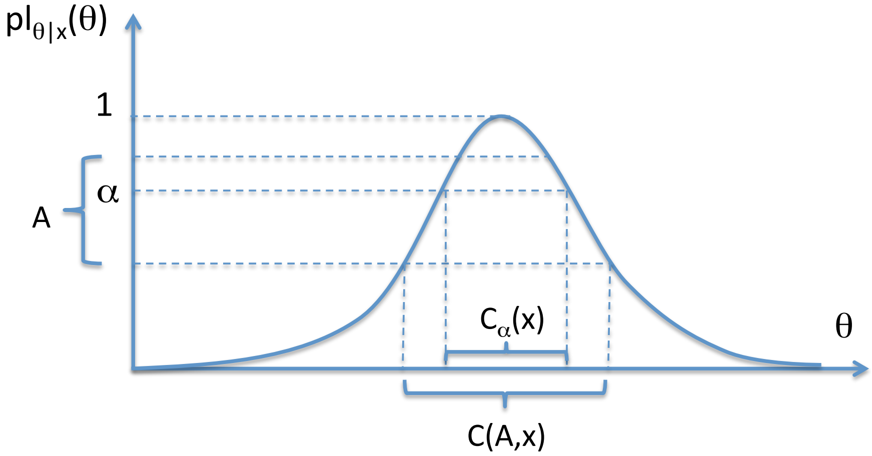


Figure 6: Constructing a confidence structure from nested confidence intervals.

Consequently, the random set $\Gamma(U, \boldsymbol{x}) = C_{1-U}(\boldsymbol{x})$ defines a confidence structure. If (27) holds only asymptotically as the amount of data tends to infinity, then $C_\alpha(\boldsymbol{X})$ is an approximate $100(1-\alpha)\%$ confidence region, and the random set $C_{1-U}(\boldsymbol{x})$ can be called an *approximate confidence structure*.

We can remark that, due to (28), the confidence structure constructed by this method induces a consonant belief function. The associated plausibility function is then known to be a possibility measure [50], and the following equation holds

$$Pl_{\theta|\boldsymbol{x}}(A) = \sup_{\theta \in A} pl_{\theta|\boldsymbol{x}}(\theta) \tag{29}$$

for all $A \subseteq \Theta$, where $pl_{\theta|\boldsymbol{x}}$ is the contour function associated to $Pl_{\theta|\boldsymbol{x}}$. This method of constructing a possibility distribution from nested confidence intervals was already suggested in [51]. When θ is scalar, the contour function $pl_{\theta|\boldsymbol{x}}(\theta)$ is a *confidence curve*, a notion introduced by Birnbaum [52]. See, e.g., [53, 54] for more recent references on this notion, and [55] for an extension to the case of a multidimensional parameter. A confidence curve can easily be constructed from a confidence distribution [48]. If $F(\cdot, \boldsymbol{x})$ is a confidence distribution for θ, the corresponding confidence curve is

$$pl_{\theta|\boldsymbol{x}}(\theta) = 2 \min \left\{ F(\theta, \boldsymbol{x}), 1 - F(\theta, \boldsymbol{x}) \right\}. \tag{30}$$

A confidence curve can also be obtained from a C-box $(\underline{F}_{\theta|\boldsymbol{x}}(\theta), \overline{F}_{\theta|\boldsymbol{x}}(\theta))$ (constructed, e.g., by pivoting the cdf as explained above) as

$$pl_{\theta|\boldsymbol{x}}(\theta) = \begin{cases} 2\overline{F}_{\theta|\boldsymbol{x}}(\theta) & \text{if } \theta \leq \overline{F}^{-1}_{\theta|\boldsymbol{x}}(0.5), \\ 1 & \text{if } \overline{F}^{-1}_{\theta|\boldsymbol{x}}(0.5) < \theta \leq \underline{F}^{-1}_{\theta|\boldsymbol{x}}(0.5), \\ 2(1 - \underline{F}_{\theta|\boldsymbol{x}}(\theta)) & \text{if } \theta > \underline{F}^{-1}_{\theta|\boldsymbol{x}}(0.5). \end{cases} \tag{31}$$

The relative likelihood (14) often provides a convenient way to obtain a nested family of confidence regions [56, Chapter 5], from which a confidence structure can be constructed. In particular, when the conditions of Wilks' theorem hold, the confidence regions $C_\alpha(\boldsymbol{x})$ have simple analytical expressions (15). The contour function $pl'_{\theta|\boldsymbol{x}}(\theta)$ corresponding to these confidence regions is then related to the relative likelihood $pl_{\theta|\boldsymbol{x}}(\theta)$ defined in (14) by the following equation,

$$c_{pl'_{\theta|\boldsymbol{x}}(\theta)} = pl_{\theta|\boldsymbol{x}}(\theta),$$

with $c_\alpha = \exp(-0.5\chi^2_{p;1-\alpha})$. The solution is

$$pl'_{\theta|\boldsymbol{x}}(\theta) = 1 - F_{\chi^2_p} \left\{ -2 \log pl_{\theta|\boldsymbol{x}}(\theta) \right\}, \tag{32}$$

where $F_{\chi^2_p}$ is the cdf of the chi square distribution with p degrees of freedom. We then have, by construction,

$$\mathbb{P}_{\boldsymbol{X}|\theta} \left\{ pl'_{\theta|\boldsymbol{X}}(\theta) \geq \alpha \right\} \approx 1 - \alpha.$$

Transformation (32) can be seen as a calibration of the likelihood-based belief function introduced in [2] and studied in [14, 15]. Figure 7 shows $pl'(\theta)$ as a function of $pl(\theta)$ for different values of p. For $p = 1$, the calibrated belief function is more specific than the likelihood-based belief function. For $p = 2$, they are identical, and for $p > 2$ calibration results in a loss of specificity. The likelihood-based belief function thus corresponds to an approximate confidence structure for

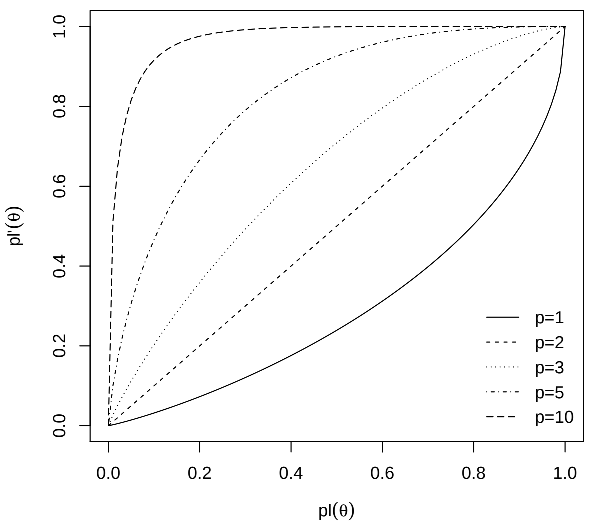


Figure 7: Transformation of the relative likelihood to calibrate a likelihood-based belief function, for different values of the number p of parameters.

$p = 1$ and $p = 2$ (and it is conservative for $p = 1$). It can be calibrated to become an approximate confidence structure for $p > 2$ (and also for $p = 1$ to make it less conservative). A refinement of this method, based on the estimation of the exact distribution of $pl_{\theta|\boldsymbol{X}}(\theta)$ by simulation, has been recently proposed in [57].

Example 4. *Consider again the binomial case with* $X \sim \mathcal{B}(n, \theta)$. *Figure 8 shows the relative likelihood function* $pl_{\theta|\boldsymbol{X}}(\theta)$, *the calibrated contour function* $pl'_{\theta|\boldsymbol{X}}(\theta)$ *obtained from (32) with* $p = 1$, *and the contour function* $pl''_{\theta|\boldsymbol{X}}(\theta)$ *computed from the C-box (26) using (31), for* $n = 100$ *and* $x = 30$. *We note that the* α*-level cuts*

$$C_{1-\alpha}(X) = \{\theta \in \Theta | pl''_{\theta|\boldsymbol{X}}(\theta) \geq \alpha\}$$

are the Clopper-Pearson confidence intervals (4) at level $1 - \alpha$. *As these confidence intervals are conservative, the confidence curve* $pl''_{\theta|\boldsymbol{X}}(\theta)$ *defines an exact confidence structure. In contrast, the* α*-level cuts of* $pl'_{\theta|\boldsymbol{X}}(\theta)$ *are only approximate confidence intervals.* □

In the case of a consonant confidence structure, there is an interesting connection between plausibilities and p-values. Specifically, consider a null hypothesis $H_0 = \{\theta_0\}$. Each confidence region $C_\alpha(\boldsymbol{X})$ corresponds to a test of H_0 at level α, which rejects H_0 when $C_\alpha(\boldsymbol{X}) \not\ni \theta_0$. For a given realization $\boldsymbol{x}$ of $\boldsymbol{X}$, the p-value can be defined as the smallest level α for which H_0 is rejected. From Figure 6, it is clear that the p-value equals $pl_{\theta|\boldsymbol{x}}(\theta_0)$. Thanks to Eq. (29), the same property holds in the case of a composite null hypothesis, the p-value being equal to the plausibility $Pl_{\theta|\boldsymbol{x}}(H_0)$. Consequently, in the case of a consonant confidence structure, there is a close connection between notions of plausibility and p-value. This connection was already noticed by Martin and Liu [58], who showed that, for most practical hypothesis testing problems, there

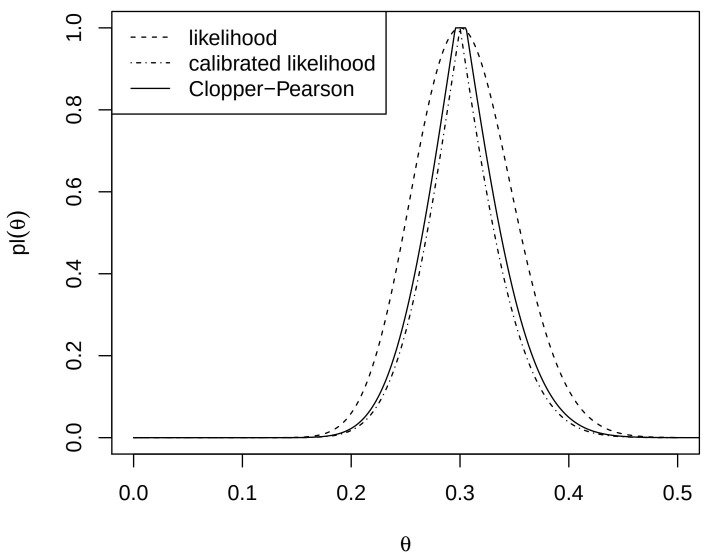


Figure 8: Relative likelihood $pl_{\theta|\boldsymbol{X}}(\theta)$, calibrated relative likelihood $pl'_{\theta|\boldsymbol{X}}(\theta)$ and Clopper-Pearson confidence curve for the binomial example with $n = 100$ and $x = 30$.

exists a data-conditional plausibility function (constructed from an inferential model, see Section 4.1), which, evaluated at the null hypothesis, is the p-value.

Confidence nets. Confidence nets, as defined in [59], also have a close connection with confidence structures[2]. For a parametric statistical model with random variable $\boldsymbol{X}$, a confidence net is a collection of data dependent sets $S_j(\boldsymbol{X})$, $j = 1, \ldots, N$, whose union is the whole parameter space, and such that

1. Each set $S_j(\boldsymbol{X})$ is a confidence region for θ with known coverage probability p_j, and
2. Any intersection of the S_j covers θ with probability zero.

For instance, let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an iid sample from a univariate continuous distribution symmetrically distributed about scalar parameter θ, and let $Z_1, \ldots, Z_{N-1}$, $N = 2^n$, denote the subsample means (i.e., the means of all nonempty subsets of the n random observations). Then, the collection of intervals between the ordered means $S_1(\boldsymbol{X}) = (-\infty, Z_{(1)})$, $S_2(\boldsymbol{X}) = (Z_{(1)}, Z_{(2)})$, $\ldots$, $S_N(\boldsymbol{X}) = (Z_{(N-1)}, +\infty)$ forms a confidence net with $p_j = 1/N$, $j = 1, \ldots, N$ [59]. It is clear that a confidence net corresponds to a special kind of C-box. Specifically, let U be a random variable with a standard normal distribution, and let Γ be the multivalued mapping defined by $\Gamma(u, \boldsymbol{x}) = S_j(\boldsymbol{x})$ for all $u \in [(j-1)/N, j/N)$, $j = 1, \ldots, N$. Then, the random set $\Gamma(U, \boldsymbol{X})$ defines a C-box, with disjoint focal sets $S_j(\boldsymbol{X})$. This principle was used by Balch and Smarslok [60] to construct a confidence structure on the offset between two otherwise similar distributions.

[2]We thank an anonymous referee for drawing our attention on confidence nets and their relation with confidence structures.

Propagation of confidence structures. As noted by Balch [20] and Ferson et al. [21], an interesting property of confidence structures is that they can be propagated through numerical equations. More precisely, let $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$ be two parameters, with confidence structures $\Gamma_1(U_1, \boldsymbol{X}_1)$ and $\Gamma_2(U_2, \boldsymbol{X}_2)$ induced by independent observations $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. Let $\eta = g(\theta_1, \theta_2)$ be a parameter defined as a function of θ_1 and θ_2, and let $V = (V_1, V_2)$ be a pair of independent random variables such that V_i has the same marginal distribution as U_i, $i = 1, 2$. Then, the random set

$$\Gamma(V, \boldsymbol{X}) = g\left\{\Gamma_1(V_1, \boldsymbol{X}_1), \Gamma_2(V_2, \boldsymbol{X}_2)\right\},$$

where $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ is a confidence structure for η. This strong result, proved in [20], shows that we can propagate confidence structures in equations using the usual Dempster-Shafer calculus (see, e.g, [61, 62, 63]), and get a confidence structure as an output. Confidence structure are thus, in this respect, compatible with the Demspter-Shafer semantics of belief functions.

3.2. Confidence structures for prediction

The notion of confidence structure can easily be extended to prediction problems, resulting in a new notion of *predictive confidence structure.* Using the notations introduced in Section 1, consider a pair of r.v. $(\boldsymbol{X}, Y)$ whose joint distribution $\mathbb{P}_{\boldsymbol{X},Y|\theta}$ depends on a parameter $\theta \in \Theta$. A data-conditional random set $\Gamma(U, \boldsymbol{X})$ will be called a predictive confidence structure if the following inequalities hold:

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left\{Y \in \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})\right\} \geq \mathbb{P}_U(A). \tag{33}$$

for all $\theta \in \Theta$ and all $A \in \mathcal{B}_U$. Having observed $\boldsymbol{X} = \boldsymbol{x}$, the induced predictive belief function is

$$Bel_{Y|\boldsymbol{x}}(B) = \mathbb{P}_U\left(\Gamma(U, \boldsymbol{x}) \subseteq B\right).$$

for all $B \in \mathcal{B}_Y$. The meaning of a predictive confidence structure is similar as that of a confidence structure in estimation problems: for any measurable subset A of Ω_U, the set $C(A, \boldsymbol{X}) = \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})$ is a prediction region for Y at confidence level $\mathbb{P}_U(A)$. Any subset B of Ω_Y thus contains a realization of a prediction region for Y with confidence level at least equal to $Bel_{Y|\boldsymbol{x}}(B)$. Most of the time (i.e., for most of the observed data $\boldsymbol{X}$ and the future data Y), regions B with a high degree of belief thus contain the future data Y.

Link with frequentist predictive distributions. Just as confidence distributions are particular confidence structures, frequentist predictive distributions [64] are a special kind of predictive confidence structures. Any method for constructing predictive confidence distributions thus provides us with predictive confidence structures. For example, the so-called pivotal method [65] starts with a pivotal quantity $W = q(Y, \boldsymbol{X})$ whose distribution function $G(w)$ does not depend on θ. If q is nondecreasing in Y, α prediction limits $L_\alpha(\boldsymbol{X})$ on Y by can be obtained by solving the inequality

$$\mathbb{P}(W \leq w_\alpha) = \alpha$$

for Y, where w_α is the α-quantile of W, leading to $\mathbb{P}(Y \leq L_\alpha(\boldsymbol{X})) = \alpha$. A frequentist predictive distribution $\widetilde{F}_{Y|\boldsymbol{x}}(y)$ can then be defined by treating the confidence limits $L_\alpha(\boldsymbol{X})$ as α-quantiles, i.e., $\widetilde{F}_{Y|\boldsymbol{x}}(L_\alpha(\boldsymbol{x})) = \alpha$. The predictive distribution can then be obtained from G as

$$\widetilde{F}_{Y|\boldsymbol{x}}(y) = G\left\{q(y, \boldsymbol{x})\right\}.$$

When W is only asymptotically pivotal, i.e., when its distribution function $G(w;\theta)$ depends on θ but approaches a fixed distribution asymptotically, then we can approximate the distribution $G(w;\theta)$ in the case of finite sample size by $\widetilde{G}(w) = G(w;\widehat{\theta})$, where $\widehat{\theta}$ is an estimate of θ. We can then proceed as if W was pivotal, which gives us an approximate predictive distribution $\widetilde{F}_{Y|\boldsymbol{x}}(y) = \widetilde{G}\left\{q(y,\boldsymbol{x})\right\}$. Given a predictive distribution $\widetilde{F}_{Y|\boldsymbol{x}}(y)$, the interval $[\widetilde{F}^{-1}_{Y|\boldsymbol{x}}(\alpha/2), \widetilde{F}^{-1}_{Y|\boldsymbol{x}}(1-\alpha/2)]$ is an exact or approximate predictive interval for Y. A general approach, [65] is to consider the following pivotal or asymptotically pivotal quantity

$$W = F_{Y|\boldsymbol{X},\widehat{\theta}(\boldsymbol{X})}(Y).$$

We assume that $\widehat{\theta}$ is a consistent estimator of θ as the information about θ increases, and W is asymptotically distributed as $\mathcal{U}[0,1]$ [65]. The predictive distribution is then

$$\widetilde{F}_{Y|\boldsymbol{x}}(y) = G\left\{F_{Y|\boldsymbol{x};\widehat{\theta}(\boldsymbol{x})}(y)\right\}, \tag{34}$$

and G can be replaced by $\widetilde{G}$ if W is asymptotically pivotal. When an analytical expression of $\widetilde{G}$ is not available, it can be estimated by a parametric bootstrap approach [65]. Specifically, let $\boldsymbol{x}_1^*,\ldots,\boldsymbol{x}_B^*$ and $y_1^*,\ldots,y_B^*$ be B bootstrap replicates of $\boldsymbol{x}$ and y, respectively. We can compute the corresponding values $w_b^* = F_{Y|x_b^*,\widehat{\theta}(\boldsymbol{x}_b^*)}(y_i^*)$, $b=1,\ldots,B$, and the distribution of W can be approximated by the empirical cdf

$$\widetilde{G}(w) = \frac{1}{B}\sum_{b=1}^{B} I(w_b^* \leq w).$$

Example 5. *Let us consider again the sea-level data of Example 2, assuming $\boldsymbol{X}=(X_1,\ldots,X_n)$ to be iid from the Gumbel distribution (16) and Y to be independently distributed according to (17). Here, the exact distribution of the quantity $W = F_{Y|\boldsymbol{X},\widehat{\theta}(\boldsymbol{X})}(Y)$ is intractable, but it can be estimated by the parametric bootstrap technique described above. Figure 9(a) shows the empirical cdf $\widetilde{G}(v)$ estimated with $B=10,000$ bootstrap samples. As we can see, the distribution of W is very close to uniform. Consequently, the predictive distribution $\widetilde{F}_{Y|\boldsymbol{x}}$ is very close to the plug-in distribution $F_{Y|\boldsymbol{x},\widehat{\theta}(\boldsymbol{x})}$, as shown in Figure 9(b).*

Prediction regions. Just as confidence structures can be derived from a nested family of confidence regions, as shown in Section 3.1, predictive confidence structures can be derived from a nested family of prediction regions. We recall that a prediction region at level $1-\alpha$ is a random set $R_\alpha(\boldsymbol{X})$ such that

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left(Y \in R_\alpha(\boldsymbol{X})\right) \geq 1-\alpha. \tag{35}$$

Assume that the family $\left(R_\alpha(\boldsymbol{X})\right)_{\alpha\in[0,1]}$ is nested, i.e., a condition similar to (28) holds for any (α,α'). Then, the multivalued mapping

$$\Gamma(u,\boldsymbol{x}) = R_{1-u}(\boldsymbol{x})$$

and $U \sim \mathcal{U}[0,1]$ induces a predictive confidence structure. (The proof is similar to the one given in Section 3 for the confidence structure case.) Again, the predictive belief function $Bel_{Y|\boldsymbol{x}}$ induced by this predictive confidence structure is consonant.

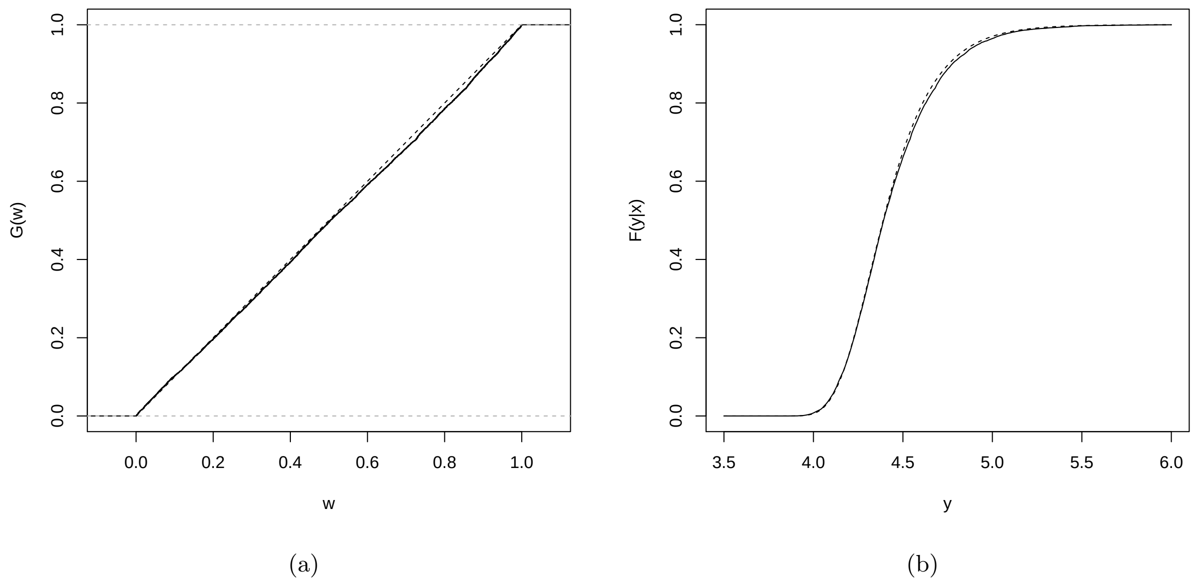


Figure 9: Frequentist predictive distribution for the Port Pirie data (Example 5). (a) bootstrap distribution $\widetilde{G}(w)$ of pivotal variable $W = F_{Y|\boldsymbol{X},\widehat{\theta}(\boldsymbol{X})}(Y)$; (b) predictive distribution $\widetilde{F}_{Y|\boldsymbol{x}}$ (solid line) and plug-in distribution $F_{Y|\boldsymbol{x},\widehat{\theta}(\boldsymbol{x})}$ (dashed line).

This method is more general than the previous one based on predictive distributions, because it can also be applied when Y is multidimensional. If, however, a predictive distribution $\widetilde{F}(y|\boldsymbol{x})$ is available, then a "prediction curve", the equivalent of a confidence curve, can be obtained as

$$pl_{Y|\boldsymbol{x}}(y) = 2\min\left\{\widetilde{F}_{Y|\boldsymbol{x}}(y), 1 - \widetilde{F}_{Y|\boldsymbol{x}}(y)\right\}, \tag{36}$$

which parallels (30). Each α-cut of this contour function is a $1-\alpha$ prediction interval. The predictive plausibility function is then

$$Pl_{Y|\boldsymbol{x}}(A) = \sup_{y\in A} pl_{Y|\boldsymbol{x}}(y)$$

for any $A \subseteq \Omega_Y$.

Example 6. *Figures 10(a) and 10(b) show, respectively, the contour function $pl_{Y|\boldsymbol{x}}(y)$ and the lower and upper cdfs for the consonant predictive belief function $Bel_{Y|\boldsymbol{x}}$ constructed from the predictive distribution $\widetilde{F}_{Y|\boldsymbol{x}}$ of Example 5.*

Use of a structural equation. A predictive confidence structure can also be built from a confidence structure via a structural equation such as (9). More precisely, if Y can be written as a function of the unknown parameter θ and some pivotal variable V as $Y = \varphi(\theta, V)$, then we obtain a predictive confidence structure by plugging a confidence structure on θ in this structural equation. This result is expressed by the following theorem.

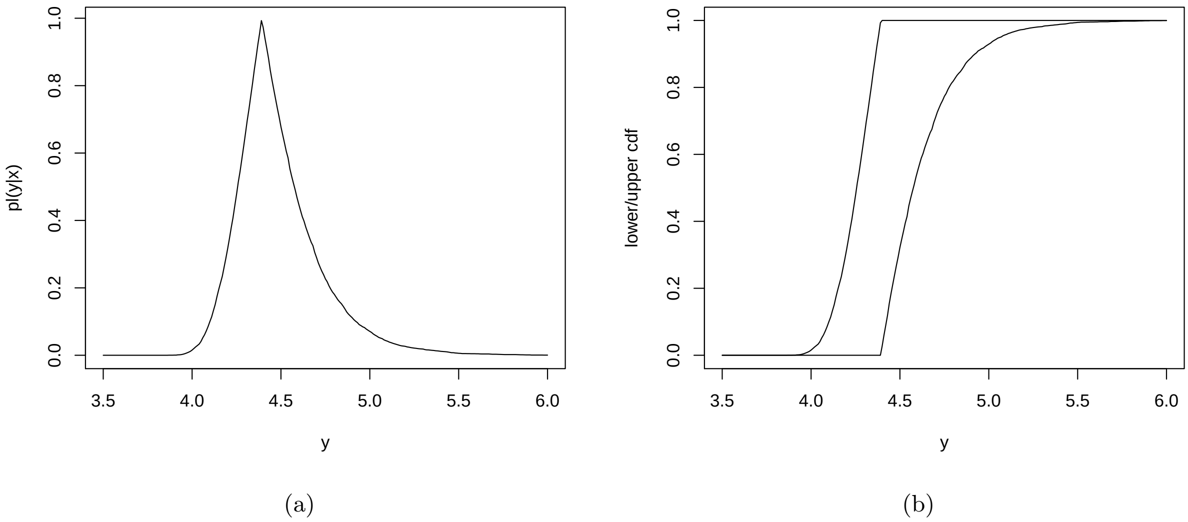


Figure 10: Consonant predictive belief function for the Port Pirie data, derived from the predictive distribution of Example 5. (a) contour function; (b) lower and upper cdfs.

Theorem 2. *Let $Y = \varphi(\theta, V)$ be a random variable, and $\Gamma(U, \boldsymbol{X})$ a confidence structure for θ. Then, the random set $\Pi(U, V, \boldsymbol{X}) = \varphi\left(\Gamma(U, \boldsymbol{X}), V\right)$ is a predictive confidence structure for Y.*

Proof. We need to prove that, for any measurable subset $A \subseteq \Omega_U \times \Omega_V$ and any $\theta \in \Theta$,

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left\{ Y \in \bigcup_{(u,v)\in A} \varphi\left(\Gamma(u, \boldsymbol{X}), v\right) \right\} \geq \mathbb{P}_{U,V}(A).$$

For any $v_0 \in \Omega_V$, let $A(v_0) = \{u \in \Omega_U \mid (u, v_0) \in A\}$. As $\Gamma(U, \boldsymbol{X})$ is a confidence structure, we have, for any $\theta \in \Theta$,

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left\{ \theta \in \bigcup_{u\in A(v_0)} \Gamma(u, \boldsymbol{X}) \right\} \geq \mathbb{P}_U(A(v_0)).$$

Now,

$$\left(\theta \in \bigcup_{u\in A(v_0)} \Gamma(u, \boldsymbol{X}) \right) \Rightarrow \left(\varphi(\theta, v) \in \bigcup_{u\in A(v_0)} \varphi(\Gamma(u, \boldsymbol{X}), v_0) \right).$$

Hence,

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left\{\varphi(\theta,v_0)\in\bigcup_{(u,v)\in A}\varphi(\Gamma(u,\boldsymbol{X}),v)\right\}\geq$$

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left\{\varphi(\theta,v_0)\in\bigcup_{u\in A(v_0)}\varphi(\Gamma(u,\boldsymbol{X}),v_0)\right\}\geq$$

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left\{\theta\in\bigcup_{u\in A(v_0)}\Gamma(u,\boldsymbol{X})\right\}\geq\mathbb{P}_U(A(v_0)).$$

Consequently,

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left\{Y\in\bigcup_{(u,v)\in A}\varphi\left(\Gamma(u,\boldsymbol{X}),v\right)\right\}=$$

$$\int\mathbb{P}_{\boldsymbol{X}|\theta}\left\{\varphi(\theta,v_0)\in\bigcup_{(u,v)\in A}\varphi(\Gamma(u,\boldsymbol{X}),v)\right\}f_V(v_0)dv_0\geq$$

$$\int\mathbb{P}_U(A(v_0))f_V(v_0)dv_0=\mathbb{P}_{U,V}(A).$$

□

In the special case where θ is a scalar parameter and the confidence structure on θ is a confidence distribution, it follows from Theorem 2 that the predictive confidence structure on $Y=\varphi(\theta,V)$ is a frequentist predictive distribution. Let us illustrate this point by the following example.

Example 7. *Let $X_1,\ldots,X_n,Y$ be iid from $\mathcal{N}(\theta,\sigma^2)$ with known σ^2. From the upper α confidence limit $\overline{X}+\frac{\sigma}{\sqrt{n}}\Phi^{-1}(\alpha)$ we get the confidence distribution of θ*

$$F(\theta,\boldsymbol{x})=\Phi\left(\frac{\sqrt{n}(\theta-\overline{x})}{\sigma}\right),$$

which is the normal cdf with mean $\overline{x}$ and standard deviation $\sigma/\sqrt{n}$. The confidence structure on θ is thus $F^{-1}(U,\boldsymbol{X})$ with $U\sim\mathcal{U}[0,1]$. Now, we have the structural equation

$$Y=\theta+V$$

with $V\sim\mathcal{N}(0,\sigma^2)$. We thus get the following predictive confidence structure predictive distribution for Y:

$$\Pi(U,V,\boldsymbol{X})=F^{-1}(U,\boldsymbol{X})+V,$$

which corresponds to a normal distribution with mean $\overline{x}$ and standard deviation $\sigma\sqrt{1+1/n}$,

$$\widetilde{F}_{Y|\boldsymbol{x}}(y)=\Phi\left(\frac{y-\overline{x}}{\sigma\sqrt{1+1/n}}\right)\tag{37}$$

We can easily check that $\widetilde{F}_{Y|\boldsymbol{x}}$ defined by (37) is a frequentist predictive distribution: its α-quantile is

$$\widetilde{F}^{-1}_{Y|\boldsymbol{x}}(\alpha) = \overline{x} + \Phi^{-1}(\alpha)\sigma\sqrt{1+\frac{1}{n}},$$

and

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left\{Y \leq \overline{X} + \Phi^{-1}(\alpha)\sigma\sqrt{1+\frac{1}{n}}\right\} = \alpha.$$

□

In the previous example, both the confidence structure on θ and the predictive confidence structure on Y induce probability distributions. When parameter θ is multidimensional, there is no confidence distribution: the belief functions on θ and Y will usually be non-additive. This more general case is illustrated by the following example.

Example 8. *Consider again the sea-level example (Example 2). The structural equation is given by (18). As mentioned in Section 3.1, the likelihood-based belief function, defined as the consonant belief function with contour function equal to the relative likelihood (14), is an approximate confidence structure when $p=2$, as the sets*

$$\Gamma(\alpha, \boldsymbol{X}) = \{\theta \in \Theta \mid pl_{\theta|\boldsymbol{X}}(\theta) \geq \alpha\}$$

are approximate $1-\alpha$ confidence regions. The predictive confidence structure $\varphi(\Gamma(U,\boldsymbol{X}),V)$ with (U,V) uniformly distributed in $[0,1]^2$ is thus an approximate predictive confidence structure. We note that the corresponding predictive belief function is identical to the one studied in [41, 25]. As the likelihood function is unimodal and continuous, the sets $\varphi(\Gamma(u,\boldsymbol{X}),v)$ are closed intervals. Their lower and upper bounds are, respectively, the minimum and the maximum of

$$\varphi(\theta, v) = \mu - \sigma \log\log v^{-1/m}$$

subject to $pl_{\theta|\boldsymbol{x}}(\theta) \geq u$. As $\varphi(\Gamma(U,\boldsymbol{X}),V)$ is a predictive confidence structure, the sets

$$C(A,\boldsymbol{X}) = \bigcup_{(u,v)\in A} \varphi\left(\Gamma(u,\boldsymbol{X}),v\right)$$

are confidence regions at level $\mathbb{P}_{U,V}(A)$, for any measurable subset A of $[0,1]$. For instance, consider the following family of sets,

$$A_\alpha = \left[1-\sqrt{1-\alpha}, 1\right] \times \left[\frac{1-\sqrt{1-\alpha}}{2}, \frac{1+\sqrt{1-\alpha}}{2}\right]$$

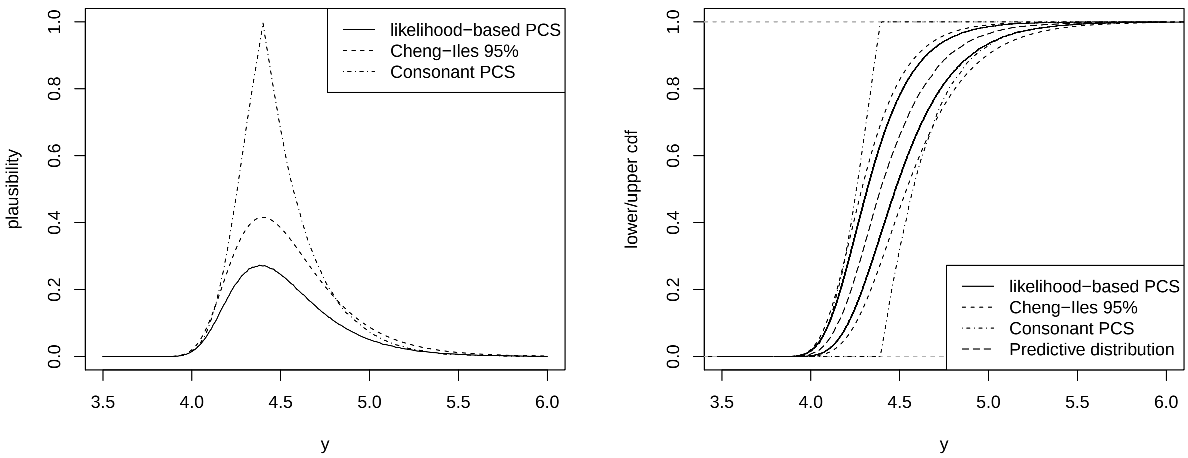
The corresponding $1-\alpha$ confidence regions are closed intervals

$$\left[\min_{\mu,\sigma}\left\{\mu - \sigma\log\log\left(\frac{1-\sqrt{1-\alpha}}{2}\right)^{-1/m}\right\}, \max_{\mu,\sigma}\left\{\mu - \sigma\log\log\left(\frac{1+\sqrt{1-\alpha}}{2}\right)^{-1/m}\right\}\right],$$

where the minimum and maximum are computed subject to the constraint $pl_{\theta|\boldsymbol{x}}(\theta) \geq \sqrt{1-\alpha}$. We estimated the coverage probabilities by simulation, for $\alpha \in \{0.5, 0.3, 0.1, 0.05\}$, by drawing $B = 1000$

Table 3: Estimated coverage probabilities of prediction regions in Example 8.

$1-\alpha$	0.5	0.7	0.9	0.95
p	0.881	0.938	0.989	0.998



(a) (b)

Figure 11: Port Pirie data: contour functions (a) and lower/upper cdfs (b) of three different predictive belief functions: predictive confidence structure derived from the likelihood and the structural equation method ("likelihood-based PCS"), predictive belief function obtained from the Cheng-Iles confidence band ("Cheng-Iles 95%"), and consonant predictive confidence structure derived from the predictive distribution computed by Lawless' method ("Consonant PCS").

observed datasets $\boldsymbol{x}_b^*$, $b = 1, \ldots, B$ *and* B *realizations* y_b^*, $b = 1, \ldots, B$ *of* Y, *from their respective distributions with* $\theta = \widehat{\theta}$. *The results are shown in Table 3. We can see that the prediction regions are very conservative. Other choices of sets* A *may lead to less conservative solutions.*

Figures 11(a) and 11(b) show, respectively, the contour functions and the lower/upper cdfs for the likelihood-based predictive belief function induced by the predictive confidence structure $\varphi(\Gamma(U, \boldsymbol{X}), V)$, *together with those obtained from the Cheng-Iles 95% confidence band (Section 2.2), and consonant predictive confidence structure already shown in Figure 10. Figure 11(b) also shows the predictive distribution* $\widetilde{F}(y|\boldsymbol{x})$.

Nonparametric predictive confidence structures. In the case where $X_1, \ldots, X_n, Y$ is iid from a univariate continuous distribution, a nonparametric confidence structure can be designed using a construction procedure similar to that used to build confidence nets (Section 3.1). Specifically, as all permutations of the $n+1$ values are equally likely, the events $Y \leq X_{(1)}$, $X_{(i)} < Y \leq X_{(i+1)}$ for $i = 1, \ldots, n-1$, and $X_{(n)} < Y$ have equal probability $1/(n+1)$. Consequently, the multivalued mapping

$$\Gamma(U, \boldsymbol{X}) = \begin{cases} (-\infty, X_{(1)}] & \text{if } U \leq \frac{1}{n+1}, \\ (X_{(i)}, X_{(i+1)}] & \text{if } \frac{i}{n+1} < U \leq \frac{i+1}{n+1}, \; i = 1, \ldots, n-1 \\ (X_{(n)}, +\infty) & \text{if } U > \frac{n}{n+1}, \end{cases}$$

with $U \sim \mathcal{U}[0,1]$, is a predictive confidence structure for Y.

4. Valid belief functions

As we have seen in Section 3, the notion of confidence structure is intimately related to those of confidence or prediction regions. For estimation problem, a degree of belief $Bel_{\theta|\boldsymbol{x}}(H) = \alpha$ in a proposition "$\theta \in H$" means that H contains a realization $C(\boldsymbol{x})$ of a confidence region for θ, at a level at least equal to α. Because an α-level confidence region contains θ for at least $100\alpha\%$ of the training samples, knowing that H contains such a confidence region with a high value of α is strong evidence for H and should logically be reflected by a high degree of belief.

However, the condition defining a confidence structure is not very restrictive, as it can be met by very different belief functions. For instance, in the case of a scalar parameter, a nested set of confidence intervals may induce both a Bayesian belief function (the confidence distribution) and a consonant one. The latter seems to be more in line with the usual interpretation of belief functions, as it assigns plausibilities between 0 and 1 to single hypotheses $\{\theta\}$. In some cases, such plausibilities are equal to relative likelihoods, or a transformation thereof. In contrast, a continuous confidence distribution assigns zero plausibility to any single hypothesis, which means that no single parameter value can be considered as plausible after observing the data, a rather counterintuitive statement.

In this section, we examine a different notion of calibration for belief functions, introduced in [22] and [23]. This notion will be reviewed in Section 4.1 in the case of estimation problems. A corresponding notion for prediction problems will be examined in Section 4.2.

4.1. Valid estimative belief functions

The notion of a *credible* [22, 23] or *valid* [18, 19] belief function captures the idea that a false hypothesis (i.e., a subset $H \subset \Theta$ that does not contain the true value of the parameter) should rarely receive a high degree of belief, or, conversely, a true hypothesis should rarely have a low plausibility. Formally, a belief function $Bel_{\theta|\boldsymbol{x}}$ is said to be valid for hypothesis H if, for any $\alpha \in (0,1)$,

$$\sup_{\theta \notin H} \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ Bel_{\theta|\boldsymbol{X}}(H) \geq 1 - \alpha \right\} \leq \alpha. \tag{38}$$

The belief function $Bel_{\theta|\boldsymbol{x}}$ is valid if it is valid for any H. This condition can be equivalently expressed in different forms, as shown by the following proposition.

Proposition 1. *Let $Bel_{\theta|\boldsymbol{X}}$ be an estimative belief function. The following conditions are equivalent:*

1. $\forall \alpha \in (0,1)$, $\forall \theta \in \Theta$, $\forall H \subset \Theta$,

$$\sup_{\theta \notin H} \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ Bel_{\theta|\boldsymbol{X}}(H) \geq 1 - \alpha \right\} \leq \alpha. \tag{39a}$$

2. $\forall \alpha \in (0,1)$, $\forall \theta \in \Theta$, $\forall H \subset \Theta$,

$$\sup_{\theta \in H} \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ Pl_{\theta|\boldsymbol{X}}(H) \le \alpha \right\} \le \alpha. \tag{39b}$$

3. $\forall \alpha \in (0,1)$,

$$\forall \theta \in \Theta, \quad \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ pl_{\theta|\boldsymbol{X}}(\theta) \le \alpha \right\} \le \alpha. \tag{39c}$$

4. *For all* $\alpha \in (0,1)$, *let* $C_\alpha(\boldsymbol{X}) = \{\theta \in \Theta \mid pl_{\theta|\boldsymbol{X}}(\theta) > \alpha\}$. *Then,* $\forall \theta \in \Theta$,

$$\mathbb{P}_{\boldsymbol{X}|\theta}\{C_\alpha(\boldsymbol{X}) \ni \theta\} \ge 1 - \alpha. \tag{39d}$$

Proof. We summarize the proof given in [18] for completeness. The equivalence between (39a) and (39b) results from the equality

$$Pl_{\theta|\boldsymbol{X}}(H) = 1 - Bel_{\theta|\boldsymbol{X}}(\overline{H}).$$

Condition (39c) is a special case of (39b) with $H = \{\theta\}$, and (39c) implies (39b) because $Pl_{\theta|\boldsymbol{X}}(H) \ge pl_{\theta|\boldsymbol{X}}(\theta)$ whenever $\theta \in H$, by the monotonicity of $Pl_{\theta|\boldsymbol{X}}$. The equivalence between (39c) and (39d) results from the equivalence

$$\mathbb{P}_{\boldsymbol{X}|\theta} \left\{ pl_{\theta|\boldsymbol{X}}(\theta) > \alpha \right\} \ge 1 - \alpha \Leftrightarrow \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ pl_{\theta|\boldsymbol{X}}(\theta) \le \alpha \right\} \le \alpha.$$

□

Each of the equivalent definitions of validity in Proposition 1 allows us to grasp the meaning of this notion. Condition (39b) means that, for any $\theta \in \Theta$ and any H such that $\theta \in H$, the random variable $Pl_{\theta|\boldsymbol{X}}(H)$ is stochastically greater than a random variable U with the standard uniform distribution. This condition formalizes the requirement that true hypotheses should often be assigned relatively high plausibilities. An immediate consequence is that, for a valid belief function, the testing rule that rejects H whenever $Pl_{\theta|\boldsymbol{X}}(H) \le \alpha$ has type-I error at level α [18]. Condition (39d) means that the $100(1-\alpha)\%$ *plausibility regions* $C_\alpha(\boldsymbol{X})$ are $100(1-\alpha)\%$ confidence regions for θ. Informally, this means that true hypotheses H rarely have a small plausibility.

We can remark that the validity condition does not imply that the belief function $Bel_{\theta|\boldsymbol{X}}$ conveys any information about θ. It is satisfied, in particular, by the vacuous belief function for which $pl_{\theta|\boldsymbol{X}}(\theta) = 1$ for all $\theta \in \Theta$: we then have $\mathbb{P}_{\boldsymbol{X}|\theta} \left\{ pl_{\theta|\boldsymbol{X}}(\theta) \le \alpha \right\} = 0 \le \alpha$ for any $\alpha \in (0,1)$. A more stringent condition can be imposed by replacing the rightmost inequality in (39c) and the inequality (39d) by equalities, i.e.,

$$\forall \theta \in \Theta, \quad \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ pl_{\theta|\boldsymbol{X}}(\theta) \le \alpha \right\} = \alpha \tag{40a}$$

and

$$\mathbb{P}_{\boldsymbol{X}|\theta}\{C_\alpha(\boldsymbol{X}) \ni \theta\} = 1 - \alpha. \tag{40b}$$

A belief function verifying (40a) or, equivalently (40b) for any θ and any $\alpha \in (0,1)$ is said to be *efficient* [18]. It is *asymptotically efficient* if these conditions hold in the limit as the sample size tends to infinity.

Proposition 1 suggests a simple way to build valid belief functions from a nested set of confidence regions $C_\alpha(\boldsymbol{X})$ verifying (27)-(28). The consonant belief function induced by the confidence structure

$$\Gamma(u, \boldsymbol{x}) = C_{1-u}(\boldsymbol{x})$$

with $U \sim \mathcal{U}[0,1]$ verifies $\{\theta \in \Theta \mid pl_{\theta|\boldsymbol{X}}(\theta) > \alpha\} = C_\alpha(\boldsymbol{X})$. Consequently, it is valid. This remark shows that the notion of confidence structure and that of valid belief function coincide in the case of consonant belief functions. However, other types of confidence structures such as confidence distributions or C-boxes (such as constructed in Example 3) are not valid. A method to generate valid belief functions using *Inferential Models* (IMs) has been introduced by Martin and Liu [18, 19]. This method is summarized below.

Inferential models. The IM approach is an adaptation of Dempster's method of inference [37, 66, 7] that guarantees that the computed belief function is valid. As Dempster, Martin and Liu [18, 19] start with a sampling model taking the form of a structural equation

$$\boldsymbol{X} = \varphi(\theta, U), \tag{41}$$

where U is an auxiliary random variable with known distribution. Having observed $\boldsymbol{X} = \boldsymbol{x}$, we must have

$$\boldsymbol{x} = \varphi(\theta, u^*) \tag{42}$$

for some unknown u^*. If u^* was known, then the set of possible values of θ could be found by solving Equation 42 for θ. The set of solutions is

$$\Gamma(u^*; \boldsymbol{x}) = \{\theta \in \Theta \mid \boldsymbol{x} = \varphi(\theta, u^*)\}.$$

At this point, Dempster's method and the IM approach diverge. Dempster [7] postulates that our beliefs about u^* are represented by the probability measure $\mathbb{P}_U$. The resulting belief function $Bel_{\theta|\boldsymbol{x}}$ is then induced by the random set $\Gamma(U; \boldsymbol{x})$, i.e., for any hypothesis H,we have

$$Bel_{\theta|\boldsymbol{x}}(H) = \mathbb{P}_U\left\{\Gamma(U; \boldsymbol{x}) \subseteq H \mid \Gamma(U; \boldsymbol{x}) \neq \emptyset\right\} \tag{43a}$$

$$Pl_{\theta|\boldsymbol{x}}(H) = \mathbb{P}_U\left\{\Gamma(U; \boldsymbol{x}) \cap H \neq \emptyset\right\}. \tag{43b}$$

In general, belief function $Bel_{\theta|\boldsymbol{x}}$ defined by (43) is not valid. To ensure this property, the IM approach consists in "predicting" u^* by a random set $S(U)$, thus "weakening" the belief function obtained in Dempster's model. Let $pl(u^*) = \mathbb{P}_U(S(U) \ni u^*)$ denote the contour function giving, for each u^*, the probability that the random set $S(U)$ hits u^*. The random set $S(U)$ is said to be valid if

$$\mathbb{P}_U(pl(U) \leq \alpha) \leq \alpha, \tag{44}$$

for all $\alpha \in (0,1)$, i.e., if the random variable $pl(U)$ stochastically dominates the standard uniform distribution. A typical choice for the mapping $u \to S(u)$ in the frequent case where $U \sim \mathcal{U}[0,1]$ is

$$S(u) = [u/2, 1 - u/2].$$

The random set $S(U)$ then induces a consonant belief function with contour function

$$pl(u) = 1 - |1 - 2u|,$$

and

$$\mathbb{P}_U(pl(U) \leq \alpha) = \mathbb{P}_U(1 - |1 - 2U| \leq \alpha) = \alpha.$$

The resulting belief function on θ is then the belief function induced by the random set $\Gamma_{\boldsymbol{x}}(S(U))$, i.e., we have

$$Bel_{\theta|\boldsymbol{x}}(H) = \mathbb{P}_U \left\{ \Gamma(S(U); \boldsymbol{x}) \subseteq H \mid \Gamma(S(U); \boldsymbol{x}) \neq \emptyset \right\} \tag{45a}$$

$$Pl_{\theta|\boldsymbol{x}}(H) = \mathbb{P}_U \left\{ \Gamma(S(U); \boldsymbol{x}) \cap H \neq \emptyset \right\}. \tag{45b}$$

It can easily be shown that the validity of $S(U)$ implies the validity of the belief function defined by (45) ([18, Theorem 2]).

Example 9. *As in Example 3, consider again the case where X has a binomial distribution $\mathcal{B}(n, \theta)$ and we wish to estimate θ. Here, a possible choice of a structural equation is*

$$F_{X|\theta}(X-1) \leq 1 - U < F_{X|\theta}(X), \tag{46}$$

with $U \sim \mathcal{U}[0,1]$ (see, e.g. [23]). Using again formula (25) relating the binomial and beta cdfs, (46) can be written as

$$1 - \mathrm{pBeta}_{\mathrm{X,n-X+1}}(\theta) \leq 1 - \mathrm{U} < 1 - \mathrm{pBeta}_{\mathrm{X+1,n-X}}(\theta). \tag{47}$$

Solving Equation (47) for θ, we get the set of solutions

$$\Gamma_X(U) = \left[\mathrm{qBeta}_{\mathrm{X,n-X+1}}(\mathrm{U}), \mathrm{qBeta}_{\mathrm{X+1,n-X}}(\mathrm{U}) \right]. \tag{48}$$

The belief function induced by the random set (48) corresponds to Dempster's solution in [37]. It is also identical to the C-box (26) obtained in Example 3. It is thus a confidence structure, but the induced belief function is not valid. To obtain a valid belief function, we need to replace U in (48) by a valid random set, such as $S(U) = [U/2, 1 - U/2]$. We then get the IM

$$\Gamma_X(S(U)) = \left[\mathrm{qBeta}_{\mathrm{X,n-X+1}}\left(\frac{\mathrm{U}}{2}\right), \mathrm{qBeta}_{\mathrm{X+1,n-X}}\left(1 - \frac{\mathrm{U}}{2}\right) \right]. \tag{49}$$

We can see that the focal intervals (49) are nested, and they are the Clopper-Pearson confidence intervals (4). Consequently, the IM approach yields the consonant belief function already found in Example 4, by applying formula 31 to the C-box (48). Its contour function $pl_{\theta|\boldsymbol{x}}(\theta)$ is the Clopper-Pearson confidence curve shown in Figure 8 for the case $n = 100$ and $x = 30$. We can notice that another choice of valid predictive random set yields another belief function, not necessarily consonant. For instance, the random set $S'(U) = [U/2, (1 + U)/2]$ is also valid, and it yields the IM

$$\Gamma_X(S'(U)) = \left[\mathrm{qBeta}_{\mathrm{X,n-X+1}}\left(\frac{\mathrm{U}}{2}\right), \mathrm{qBeta}_{\mathrm{X+1,n-X}}\left(\frac{1+\mathrm{U}}{2}\right) \right]. \tag{50}$$

As the bounds of $\Gamma_X(S'(U))$ are co-monotonic, this random set defines a p-box. However, its contour function is identical to that induced by (49).

4.2. Valid predictive belief functions

The notion of validity can straightforwardly be extended to predictive belief functions [24]. Using the notations introduced in Section 1, a predictive belief function with contour function $pl_{Y|\boldsymbol{X}}$ is said to be valid if the random variable $pl_{Y|\boldsymbol{X}}(Y)$ stochastically dominates the uniform distribution, i.e., if

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left\{pl_{Y|\boldsymbol{X}}(Y) \leq \alpha\right\} \leq \alpha, \tag{51}$$

for any $\theta \in \Theta$ and any $\alpha \in (0,1)$. Using the same line of reasoning as in Section 4.1, it is easy to show that a predictive belief function with contour function $pl_{Y|\boldsymbol{X}}$ is valid if and only if the $100(1-\alpha)\%$ plausibility sets

$$R_\alpha(\boldsymbol{X}) = \{y \in \Omega_Y \mid pl_{Y|\boldsymbol{X}}(y) > \alpha\}.$$

are $100(1-\alpha)\%$ prediction regions, i.e., if

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\{R_\alpha(\boldsymbol{X}) \ni Y\} \geq 1-\alpha, \tag{52}$$

for any $\theta \in \Theta$ and any $\alpha \in (0,1)$. Consequently, a valid predictive belief function can be constructed from a nested family of confidence regions $\{R_\alpha(\boldsymbol{X})\}$ for $\alpha \in (0,1)$, through the multi-valued mapping

$$\Gamma(u, \boldsymbol{x}) = R_{1-u}(\boldsymbol{x})$$

with $U \sim \mathcal{U}[0,1]$. Such a family of confidence regions can be obtained, for instance, by first constructing a predictive distribution such as (34) using Lawless' method [65], and then using transformation (36) to obtain a contour function $pl_{Y|\boldsymbol{x}}$. This contour function defines a valid and consonant predictive belief function. In other words, a consonant predictive confidence structure induces a valid predictive belief functions. This construction, which, to the best of our knowledge, has not been considered before as a means to obtain valid predictive belief functions, was illustrated in Example 5 (see Figure 10).

Another method for constructing valid predictive belief functions was recently proposed by Martin and Lingham in [24]. Basically, the method uses two structural equations for $\boldsymbol{X}$ and Y. Solving the first equation for θ and plugging in to the second equation, they get a new equation relating Y to $\boldsymbol{X}$ and an auxiliary variable U. Predicting U using a valid random set $S(U)$ verifying (44) then yields a valid predictive belief function for Y. The following example is taken from [24].

Example 10. *Assume that $X \sim \mathcal{B}(n,\theta)$ and $Y \sim \mathcal{B}(m,\theta)$ are two binomial variables with the same parameter θ and known numbers of trials n and m. Using again the structural equation (46), we get*

$$F_{X|\theta}(X-1) \leq 1-U < F_{X|\theta}(X) \tag{53}$$

and

$$F_{Y|\theta}(Y-1) \leq 1-V < F_{Y|\theta}(Y), \tag{54}$$

where (U,V) has a uniform distribution in $[0,1]^2$. Writing (53) in the form (47) using (25), and solving for θ, we get

$$\mathrm{qBeta}_{\mathrm{X,n-X+1}}(\mathrm{U}) \leq \theta < \mathrm{qBeta}_{\mathrm{X+1,n-X}}(\mathrm{U}). \tag{55}$$

Solving (54) for Y, we obtain

$$F^{-1}_{Y|\theta}(1-V) < Y \leq 1 + F^{-1}_{Y|\theta}(1-V). \tag{56}$$

Using the fact that $F^{-1}_{Y|\theta}(v)$ is an increasing function of θ for all v, we can plug in the interval (55) into (56) to get

$$F^{-1}_{Y|\theta_1(X,U)}(1-V) < Y \leq 1 + F^{-1}_{Y|\theta_1(X,U)}(1-V), \tag{57}$$

where $\theta_1(X,U)$ and $\theta_2(X,U)$, respectively, the lower and upper endpoints of interval (55). Predicting U and V using valid random sets, we get a valid predictive belief function of Y.

According to Martin and Lingham in [24], the prediction intervals $R_\alpha(X)$ obtained using the method detailed in Example 10 are close to those constructed using Wang's method described in [67]. For this problem, we could thus equivalently start from these prediction intervals and build a valid consonant belief function as explained above. In general, the available techniques for constructing confidence regions, both for continuous and for discrete distributions, provide easy ways to obtain valid predictive belief functions for a wide range of problems. A more extensive comparison between this approach and the recent method proposed in [24] remains to be performed.

5. Summary and Conclusions

In this paper, we have tried to put in perspective recent streams of research on statistical inference in the belief function framework, which propose different ways to compute degrees of belief with frequency-related interpretation from statistical evidence. Three different definitions of a calibrated belief function, recapitulated in Table 1, have been put forward. Each of these definitions relates degrees of belief with long-run frequencies in repeated experiments, ensuring that true statements are "often" assigned a high degree of belief, while false statements "often" have a low plausibility. Consider, for instance, the problem of predicting some future observation Y based on past observation $\boldsymbol{X} = \boldsymbol{x}$, and consider some statement "$Y \in B$" for some $B \in \mathcal{B}_Y$. For some predictive belief function $Bel_{Y|\boldsymbol{x}}$, let $Bel_{Y|\boldsymbol{x}}(B) = \beta_B$ and $Pl_{Y|\boldsymbol{x}}(B) = 1 - Bel_{Y|\boldsymbol{x}}(\overline{B}) = \pi_B$.

- If $Bel_{Y|\boldsymbol{x}}$ is a predictive belief function at confidence level $1-\alpha$, then $[\beta_B, \pi_B]$ is a realization of a $100(1-\alpha)\%$ prediction interval for $\mathbb{P}_{Y|\boldsymbol{X}}(B)$. In other words, this interval was computed by a method which, most of the time, provides an interval that encloses the conditional probability of the event $Y \in B$. A high value of β_B (respectively, a low value of π_B) is thus logically associated with a high degree of belief that the event $Y \in B$ will (respectively, will not) happen.
- If $Bel_{Y|\boldsymbol{x}}$ is induced by a predictive confidence structure, then we know that B and $\overline{B}$ contain realizations of prediction regions for Y at levels, respectively, β_B and $1 - \pi_B$. Again, a high value of β_B (respectively, a low value of π_B) corresponds to strong evidence in favor of (respectively, against) B.
- Assume $Bel_{Y|\boldsymbol{x}}$ is a valid predictive belief function. For any $y \in B$, $pl_{Y|\boldsymbol{x}}(y) \leq \pi_B$. The event $pl_{Y|\boldsymbol{X}}(Y) \leq \pi_B$ has a probability less than π_B. If π_B is small, we are thus inclined to believe that the event $Y \in B$ will not occur. Conversely, for any $y \notin B$, $pl_{Y|\boldsymbol{x}}(y) \leq Pl(\overline{B}) = 1 - \beta_B$, and the event $pl_{Y|\boldsymbol{X}}(Y) \leq 1 - \beta_B$ has a probability less than $1 - \beta_B$: a high value of β_B thus corresponds to a good reason to believe that the event $Y \in B$ will happen.

It results from this discussion that all three definitions of calibration make sense and are consistent with the usual semantics of belief functions. From a practical point of view, the necessity to fix a confidence level $1 - \alpha$ in the first approach can be seen as a drawback. The other two approaches do not have this limitation, and they are equivalent in the case of consonant belief functions. The availability of simple procedures for constructing such belief functions, based, e.g., on nested families of confidence or prediction intervals, is another argument in favor of these notions.

We believe that frequentist interpretations of degrees of belief such as discussed in this paper may facilitate the acceptance of belief function analyses by scientist and engineers (see, e.g., [68] for an engineering application of belief functions and statistical inference). On the other hand, it must be stressed that, by making belief functions compatible with frequentist concepts such as confidence regions and confidence distributions, we generally loose compatibility with Bayesian inference. In particular, combining a data-conditional belief function, constructed using any of the methods reviewed in this paper, with a probabilistic prior using Dempster's rule will not yield the Bayesian posterior. This is in contrast with Dempster's method of inference [7], and with the likelihood-based methods described in [14, 25]. It thus seems that "frequentist" and "generalized Bayesian" views of belief functions cannot be easily reconciled and have to coexist, just as frequentist and Bayesian procedures in mainstream statistics [69]. It remains to be seen if a useful compromise between these two approaches can be found, perhaps drawing ideas from the calibrated Bayes paradigm [70].

Acknowledgement

This research was supported by a grant from the Beijing Government as part of the "Overseas Talents" program.

References

[1] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Annals of Mathematical Statistics 38 (1967) 325–339.

[2] G. Shafer, A mathematical theory of evidence, Princeton University Press, Princeton, N.J., 1976.

[3] G. Shafer, Perspectives in the theory and practice of belief functions, International Journal of Approximate Reasoning 4 (1990) 323–362.

[4] T. Denoeux, 40 years of Dempster-Shafer theory, International Journal of Approximate Reasoning 79 (2016) 1–6.

[5] G. Shafer, A mathematical theory of evidence turns 40, International Journal of Approximate Reasoning 79 (2016) 7–25.

[6] A. P. Dempster, A generalization of Bayesian inference (with discussion), J. R. Statistical Society B 30 (1968) 205–247.

[7] A. P. Dempster, The Dempster-Shafer calculus for statisticians, International Journal of Approximate Reasoning 48 (2) (2008) 365–377.

[8] A. W. F. Edwards, Likelihood (expanded edition), The John Hopkins University Press, Baltimore, USA, 1992.

[9] G. A. Barnard, G. M. Jenkins, C. B. Winsten, Likelihood inference and time series, Journal of the Royal Statistical Society 125 (3) (1962) 321–372.

[10] A. Birnbaum, On the foundations of statistical inference, Journal of the American Statistical Association 57 (298) (1962) 269–306.

[11] L. A. Wasserman, Belief functions and statistical evidence, The Canadian Journal of Statistics 18 (3) (1990) 183–196.

[12] Y. Y. Chen, Statistical inference based on the possibility and belief measures, Transactions of the American Mathematical Society 347 (5) (1995) 1855–1863.

[13] M. Aickin, Connecting Dempster-Shafer belief functions with likelihood-based inference, Synthese 123 (2000) 347–364.

[14] T. Denœux, Likelihood-based belief function: justification and some extensions to low-quality data, International Journal of Approximate Reasoning 55 (7) (2014) 1535–1547.

[15] T. Denœux, Rejoinder on "likelihood-based belief function: Justification and some extensions to low-quality data", International Journal of Approximate Reasoning 55 (7) (2014) 1614–1617.
[16] T. Denœux, Constructing belief functions from sample data using multinomial confidence regions, International Journal of Approximate Reasoning 42 (3) (2006) 228–252.
[17] A. Aregui, T. Denœux, Constructing predictive belief functions from continuous sample data using confidence bands, in: G. De Cooman, J. Vejnarová, M. Zaffalon (Eds.), Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '07), Prague, Czech Republic, 2007, pp. 11–20.
[18] R. Martin, C. Liu, Inferential models: A framework for prior-free posterior probabilistic inference, Journal of the American Statistical Association 108 (2013) 301–313.
[19] R. Martin, C. Liu, Inferential Models: Reasoning with Uncertainty, CRC Press, Boca Raton, 2016.
[20] M. S. Balch, Mathematical foundations for a theory of confidence structures, International Journal of Approximate Reasoning 53 (7) (2012) 1003 – 1019.
[21] S. Ferson, M. Balch, K. Sentz, J. Siegrist, Computing with confidence, in: F. Cozman, T. Denœux, S. Destercke, T. Seidenfeld (Eds.), ISIPTA'13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, Compiègne, 2013, pp. 129–138.
[22] R. Martin, J. Zhang, C. Liu, Dempster-Shafer theory and statistical inference with weak beliefs, Statistical Science 25 (2010) 72–87.
[23] J. Zhang, C. Liu, Dempster-Shafer inference with weak beliefs, Statistica Sinica 21 (2011) 475–494.
[24] R. Martin, R. T. Lingham, Prior-free probabilistic prediction of future observations, Technometrics 58 (2) (2016) 225–235.
[25] O. Kanjanatarakul, S. Sriboonchitta, T. Denœux, Statistical estimation and prediction using belief functions: principles and application to some econometric models, International Journal of Approximate Reasoning 72 (2016) 71–94.
[26] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, International Journal of General Systems 12 (3) (1986) 193–226.
[27] C. J. Clopper, E. S. Pearson, The use of confidence or fiducial limits illustrated in the case of the binomial, Biometrika 26 (1934) 404–413.
[28] C. P. Quesenberry, D. C. Hurst, Large sample simultaneous confidence intervals for multinomial proportions, Technometrics 6 (2) (1964) 191–195.
[29] L. A. Goodman, On simultaneous confidence intervals for multinomial proportions, Technometrics 7 (2) (1965) 247–254.
[30] M.-H. Masson, T. Denœux, Inferring a possibility distribution from empirical data, Fuzzy Sets and Systems 157 (3) (2006) 319–340.
[31] E. L. Lehman, Testing statistical hypotheses, 2nd Edition, Springer-Verlag, New-York, 1986.
[32] S. Ferson, V. Kreinovitch, L. Ginzburg, D. S. Myers, K. Sentz, Constructing probability boxes and Dempster-Shafer structures, Tech. Rep. SAND2002-4015, Sandia National Laboratories, Albuquerque, NM (2003).
[33] E. Kriegler, H. Held, Utilizing belief functions for the estimation of future climate change, International Journal of Approximate Reasoning 39 (2–3) (2005) 185–209.
[34] A. N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione, Istituto Italiano degli Attuari 4 (1933) 83–91.
[35] M. Kendall, A. Stuart, The advanced theory of statistics, 4th Edition, Vol. 2, Charles Griffin and Co Ltd, London, 1979.
[36] R. C. H. Cheng, T. C. Iles, Confidence bands for cumulative distribution functions of continuous random variables, Technometrics 25 (1) (1983) 77–86.
[37] A. P. Dempster, New methods for reasoning towards posterior distributions based on sample data, Annals of Mathematical Statistics 37 (1966) 355–374.
[38] D. J. Hudson, Interval estimation from the likelihood function, J. R. Statistical Society B 33 (2) (1973) 256–262.
[39] S. S. Wilks, The large-sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Math. Statist. 9 (1) (1938) 60–62.
[40] H. W. Peers, Likelihood ratio and associated test criteria, Biometrika 58 (3) (1971) 577–587.
[41] O. Kanjanatarakul, S. Sriboonchitta, T. Denœux, Forecasting using belief functions: an application to marketing econometrics, International Journal of Approximate Reasoning 55 (5) (2014) 1113–1128.
[42] S. G. Coles, An Introduction to Statistical Modelling of Extreme Values, Springer, London, 2001.
[43] A. Aregui, T. Denœux, Constructing consonant belief functions from sample data using confidence sets of pignistic probabilities, International Journal of Approximate Reasoning 49 (3) (2008) 575–594.
[44] P. Smets, R. Kennes, The Transferable Belief Model, Artificial Intelligence 66 (1994) 191–243.

[45] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, International Journal of Approximate Reasoning 38 (2005) 133–147.
[46] D. Dubois, H. Prade, P. Smets, A definition of subjective possibility, International Journal of Approximate Reasoning 48 (2) (2008) 352–364.
[47] T. Schweder, N. L. Hjort, Confidence and likelihood, Scandinavian Journal of Statistics 29 (2002) 309–332.
[48] M.-G. Xie, K. Singh, Confidence distribution, the frequentist distribution estimator of a parameter: A review, International Statistical Review 81 (1) (2013) 3–39.
[49] G. Casella, R. L. Berger, Statistical Inference, 2nd Edition, Thomson Learning, Duxbury Thomson Learning, 2002.
[50] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.
[51] D. Dubois, Possibility theory and statistical reasoning, Computational Statistics and Data Analysis 51 (1) (2006) 47–69.
[52] A. Birnbaum, Confidence curves: An omnibus technique for estimation and testing statistical hypotheses, Journal of the American Statistical Association 56 (294) (1961) 246–249.
[53] H. Blaker, Confidence curves and improved exact confidence intervals for discrete distributions, Canadian Journal of Statistics 28 (4) (2000) 783–798.
[54] R. Bender, G. Berg, H. Zeeb, Tutorial: Using confidence curves in medical research, Biometrical Journal 47 (2) (2005) 237–247.
[55] T. Schweder, Confidence nets for curves, in: V. Nair (Ed.), Advances in Statistical Modeling and Inference. Essays in honor of Kjell A. Doksum, World Scientific, 2017, pp. 593–609.
[56] Y. Pawitan, In all likelihood: Statistical Modelling and Inference Using Likelihood, Clarendon Press, Oxford, 2001.
[57] R. Martin, Plausibility functions and exact frequentist inference, Journal of the American Statistical Association 110 (512) (2015) 1552–1561.
[58] R. Martin, C. Liu, A note on p-values interpreted as plausibilities, Statistica Sinica 24 (4) (2014) 1703–1716.
[59] A. R. Francis, M. Stehlk, H. P. Wynn, "Building" exact confidence nets, Bernoulli 23 (4B) (2017) 3145–3165.
[60] M. S. Balch, B. P. Smarslok, A pre-validation study on supersonic wind tunnel data collected from legacy aerothermal experiments, in: 16th AIAA Non-Deterministic Approaches Conference, American Institute of Aeronautics and Astronautics, 2014, pp. 1–11.
[61] R. R. Yager, Arithmetic and other operations on Dempster-Shafer structures, Int. J. Man-Machines Studies 25 (1986) 357–366.
[62] G. Shafer, P. P. Shenoy, K. Mellouli, Propagating belief functions in qualitative Markov trees, International Journal of Approximate Reasoning 1 (1987) 349–400.
[63] J. Kohlas, Modeling uncertainty with belief functions in numerical models, European Journal of Operational Research 40 (3) (1989) 377–388.
[64] O. E. Barndorff-Nielsen, D. R. Cox, Prediction and asymptotics, Bernoulli 2 (4) (1996) 319–340.
[65] J. F. Lawless, M. Fredette, Frequentist prediction intervals and predictive distribution, Biometrika 92 (3) (2005) 529–542.
[66] A. P. Dempster, Upper and lower probability inferences based on a sample from a finite univariate population, Biometrika 57 (1967) 515–528.
[67] H. Wang, Closed form prediction intervals applied for disease counts, The American Statistician 64 (3) (2010) 250–256.
[68] N. Ben Abdallah, N. Mouhous Voyneau, T. Denœux, Combining statistical and expert evidence using belief functions: Application to centennial sea level estimation taking into account climate change, International Journal of Approximate Reasoning 55 (2014) 341–354.
[69] M. J. Bayarri, J. O. Berger, The interplay of Bayesian and frequentist analysis, Statist. Sci. 19 (1) (2004) 58–80. doi:10.1214/088342304000000116.
[70] R. J. Little, Calibrated Bayes: A Bayes/frequentist roadmap, The American Statistician 60 (3) (2006) 213–223.