Frequency-calibrated belief functions

Thierry Denœux

Université de Technologie de Compiègne
HEUDIASYC (UMR CNRS 6599)
http://www.hds.utc.fr/~tdenoeux

WUML/WPMSIIP workshop
Compiègne, September 11, 2017

Classical approaches to inference using BFs

- Dempster's approach: uses of a structural equation $X = \varphi(\theta, U)$, where X is the random observed data, $\theta \in \Theta$ is the unknown parameter and U is an auxiliary variable 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.
- Likelihood-based approach: constructs a consonant belief function directly from the likelihood function.
- These methods can thus be seen as implementing a form of prior-free generalization of Bayesian inference (combining the data-conditional belief function with a prior probability distribution using Dempster's rule yields the Bayesian posterior distribution).
- They lack, however, the frequency calibration properties expected by many statisticians.

Blending BF inference with frequentist ideas

- In recent years, several attempts have been made to blend belief function inference with frequentist ideas:
  1. Frequentist predictive belief functions (Denoeux, 2006; Aregui and Denoeux, 2007)
  2. Confidence structures (Balch, 2012)
  3. Inferential models (Martin and Liu, 2013, 2016)
- Are there any links between these approaches? What interpretations underly them?

Notations and terminology

- Let $\boldsymbol{x}$ be the observed data, assumed to be a realization of a random vector $\boldsymbol{X}$ with sample space $\Omega_{\boldsymbol{X}}$. We will consider 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 $\boldsymbol{x}$. 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 $\boldsymbol{X} = \boldsymbol{x}$.
- Let $(\boldsymbol{X}, \boldsymbol{Y})$ be a pair of random variables with joint sample space $\Omega_{\boldsymbol{X}} \times \Omega_{\boldsymbol{Y}}$, where $\boldsymbol{X}$ is the (past) observed data and $\boldsymbol{Y}$ is the (future) not-yet observed data. The joint distribution $\mathbb{P}_{\boldsymbol{X},\boldsymbol{Y}|\theta}$ depends on parameter θ. A predictive belief function $Bel_{\boldsymbol{Y}|\boldsymbol{x}}$ is a data-conditional belief function on $\Omega_{\boldsymbol{Y}}$ quantifying the uncertainty on $\boldsymbol{Y}$ after observing the evidence $\boldsymbol{x}$.

Outline

Intuitive idea

- If we knew the conditional distribution $\mathbb{P}_{Y|\boldsymbol{x},\theta}$ of Y given $\boldsymbol{X} = \boldsymbol{x}$, we would 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, 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 committed than $\mathbb{P}_{Y|\boldsymbol{x},\theta}$:

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

- However, this condition is generally too strict to be of any practical value, as they can be guaranteed only for the vacuous belief function.
- Solution (Denoeux, 2006): 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.

Definition

Definition

A belief function verifying

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left(Bel_{Y|\boldsymbol{X}} \leq \mathbb{P}_{Y|\boldsymbol{X},\theta}\right) \geq 1-\alpha, \tag{2}$$

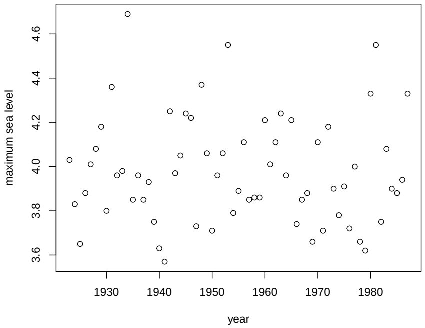
for all $\theta \in \Theta$ is called a predictive belief function at confidence level $1-\alpha$. It is an approximate $1-\alpha$-level predictive belief function if Property (2) holds only in the limit as the sample size tends to infinity.

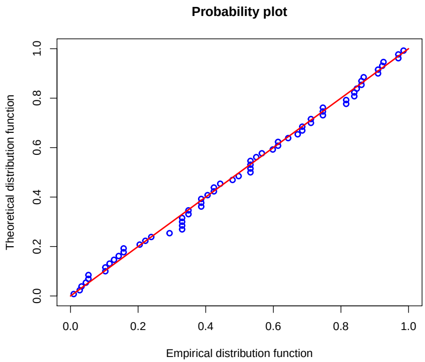
Property: $[Bel_{Y|\boldsymbol{X}}(A), Pl_{Y|\boldsymbol{X}}(A)]$ is a CI for $\mathbb{P}_{Y|\boldsymbol{X},\theta}(A)$ at level $1-\alpha$

Construction

- $X_1, \ldots, X_n, Y$ iid from a discrete distribution: multinomial simultaneous confidence intervals (Denoeux, 2006)
- $X_1, \ldots, X_n, Y$ iid from a continuous distribution: confidence band (Aregui and Denoeux, 2007). In that case, the PBF is a p-box.
- General case: use a structural equation $Y = \varphi(\theta, V)$ and a confidence region $C_\alpha(\boldsymbol{X})$ be a for θ at level $1-\alpha$. The random set $\varphi(C_\alpha(\boldsymbol{x}), V)$ is a predictive belief function at level $1-\alpha$. (Denoeux, 2017)

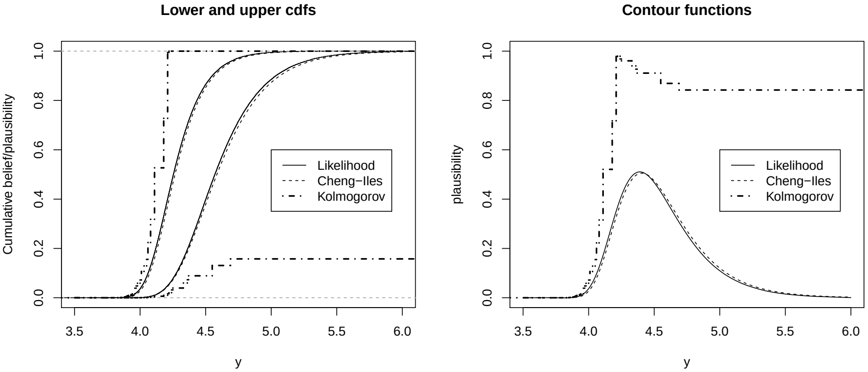
Example





Annual maximum sea-levels recorded at Port Pirie over the period 1923-1987 (left) and probability plot for the Gumbel fit to the data (right)

Example



Port Pirie data: lower and upper cdfs (left) and contour functions (right) of the predictive belief functions about the maximum sea level in the next 10 years at confidence level 95%

Outline

Definition

- A confidence structure as defined by Balch (2012) is an observation-conditional random set that encodes confidence regions at all levels.
- Let $(\Omega_U, \mathcal{B}_U, \mathbb{P}_U)$ be a probability space, $\mathcal{B}_\Theta$ an algebra of subsets of Θ, and

$$\Gamma : \Omega_U \times \Omega_{\boldsymbol{X}} \to \mathcal{B}_\Theta,$$

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 $B \in \mathcal{B}_\Theta$.

Definition

Definition

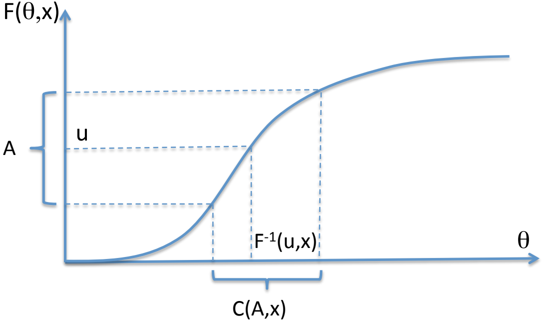
Mapping Γ defines a confidence structure if the following inequality holds for all $\theta \in \Theta$ and all $A \in \mathcal{B}_U$,

$$\mathbb{P}_{\boldsymbol{X}|\theta}\left\{\theta \in \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})\right\} \geq \mathbb{P}_U(A). \tag{3}$$

Condition (3) 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 $\mathbb{P}_U(A)$.

Confidence distribution as a special case

- $F(\theta, \boldsymbol{x})$ is a confidence distribution iff for any $\alpha \in (0, 1)$, $(-\infty, F^{-1}(\alpha, \boldsymbol{X})]$ is a lower-side confidence interval for θ at level α
- Let $\Gamma(u, \boldsymbol{x}) = \{F^{-1}(u, \boldsymbol{x})\}$ and $U \sim \mathcal{U}[0, 1]$



Construction from nested confidence regions

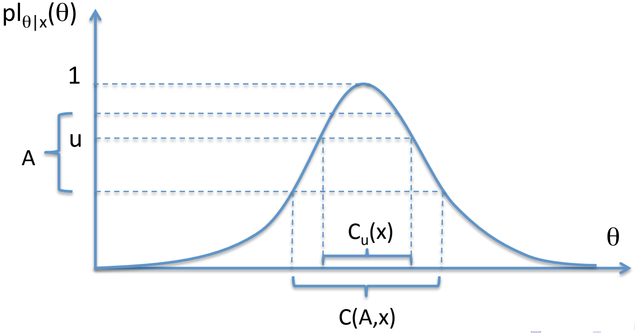
- Let $C_\alpha(\boldsymbol{X})$, $\alpha \in [0,1]$ be a nested family of confidence regions, such that

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

and, for any (α, α'), $\alpha < \alpha' \Rightarrow C_\alpha(\boldsymbol{X}) \supseteq C_{\alpha'}(\boldsymbol{X})$.

- Consider the confidence structure with multivalued mapping

$$\Gamma(u, \boldsymbol{x}) = C_u(\boldsymbol{x}) \text{ and } U \sim \mathcal{U}[0,1]$$



Interpretation

- 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 $\overline{H}$ contains a realization of a confidence region at level at least equal to $1 - \alpha$.

Predictive confidence structures

- An observation-dependent random set $\Gamma(U, \boldsymbol{X})$ is called a predictive confidence structure (Denoeux, 2017) if the following inequalities hold

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left\{Y \in \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})\right\} \geq \mathbb{P}_U(A). \quad (4)$$

  for all $\theta \in \Theta$ and all measurable subset $A \subseteq \Omega_U$.
- Having observed $\boldsymbol{X} = \boldsymbol{x}$, the induced PBF is

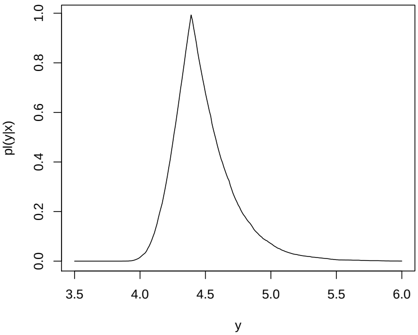
$$Bel_{Y|\boldsymbol{x}}(B) = \mathbb{P}_U\left(\Gamma(U, \boldsymbol{x}) \subseteq B\right) \text{ for all } B \subseteq \Omega_Y$$

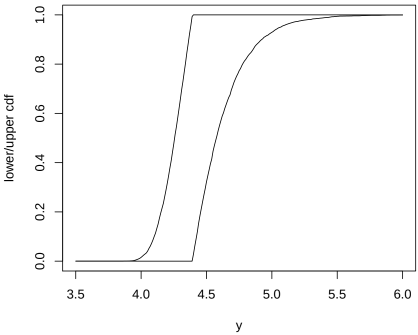
- Meaning: for any measurable subset A of Ω_U, the set $C(A, \boldsymbol{X}) = \bigcup_{u \in A} \Gamma(u, \boldsymbol{X})$ is a prediction region for Y at confidence level $\mathbb{P}_U(A)$. Any subset B of Ω_Y thus contains a realization of a confidence region for Y with confidence level at least equal 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.

Construction

- Frequentist predictive distribution
- Prediction region
- Structural equation (Denoeux, 2017)

Example





Consonant predictive belief function for the Port Pirie data, derived from a frequentist predictive distribution. (left) contour function; (right) lower and upper cdfs.

Outline

Example

- The notion of a valid belief function (Martin and Liu, 2013) 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 $\alpha \in (0,1)$,
$$\sup_{\theta \notin H} \mathbb{P}_{\boldsymbol{X}|\theta}\left\{Bel_{\theta|\boldsymbol{X}}(H) \geq 1-\alpha\right\} \leq \alpha. \tag{5}$$
The belief function $Bel_{\theta|\boldsymbol{x}}$ is valid if it is valid for any H.

Proposition

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

1. $\forall \alpha \in (0,1), \forall \theta \in \Theta, \forall H \subset \Theta,$
$$\sup_{\theta \notin H} \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ Bel_{\theta|\boldsymbol{X}}(H) \geq 1-\alpha \right\} \leq \alpha. \tag{6a}$$
2. $\forall \alpha \in (0,1), \forall \theta \in \Theta, \forall H \subset \Theta,$
$$\sup_{\theta \in H} \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ Pl_{\theta|\boldsymbol{X}}(H) \leq \alpha \right\} \leq \alpha. \tag{6b}$$
3. $\forall \alpha \in (0,1),$
$$\forall \theta \in \Theta, \quad \mathbb{P}_{\boldsymbol{X}|\theta} \left\{ pl_{\theta|\boldsymbol{X}}(\theta) \leq \alpha \right\} \leq \alpha. \tag{6c}$$
4. For all $\alpha \in (0,1)$, let $C_\alpha(\boldsymbol{X}) = \{\theta \in \Theta \mid pl_{\theta|\boldsymbol{X}}(\theta) > \alpha\}$. Then, $\forall \theta \in \Theta$,
$$\mathbb{P}_{\boldsymbol{X}|\theta}\{C_\alpha(\boldsymbol{X}) \ni \theta\} \geq 1-\alpha. \tag{6d}$$

Construction

- The previous proposition suggests a simple way to build valid belief functions from a nested set of confidence regions $C_\alpha(\boldsymbol{X})$. 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 are not valid.
- A method to generate valid belief functions using Inferential Models (IMs) has been introduced by Martin and Liu (2013).

Valid predictive belief functions

- The notion of validity can straightforwardly be extended to predictive belief functions (Martin and Lingham, 2016). 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

$$\mathbb{P}_{\boldsymbol{X},Y|\theta}\left\{pl_{Y|\boldsymbol{X}}(Y) \leq \alpha\right\} \leq \alpha, \tag{7}$$

  for any $\theta \in \Theta$ and any $\alpha \in (0,1)$.
- Equivalent condition: 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 for all $\alpha \in (0,1)$

Construction

- 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 $U \sim \mathcal{U}[0, 1]$.
- Other method for constructing valid predictive belief functions was recently proposed by Martin and Lingham (2016)

Summary I

- 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 $B \subseteq \Omega_Y$. For some predictive belief function $Bel_{Y|\boldsymbol{x}}$, let $Bel_{Y|\boldsymbol{x}}(B) = \beta_B$ and $Pl_{Y|\boldsymbol{x}}(B) = 1 - Bel_{Y|\boldsymbol{x}}(\overline{B}) = \pi_B$.

1. 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 $\mathbb{P}_{Y|\boldsymbol{X}}(B)$. 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.
2. 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.

Summary II

3. Assume $Bel_{Y|\boldsymbol{x}}$ is a valid predictive belief function. For any $y \in B$, $pl(y) \leq \pi_B$. The event $pl(Y) \leq \pi_B$ 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, if $y \notin B$, then $pl(y) \leq Pl(\overline{B}) = 1 - \beta_B$, and this event 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.

Conclusions

- All three definitions of calibration make sense and are consistent with the usual semantics of belief functions. For instance, in prediction problems
  - A high value of $Bel_{Y|\boldsymbol{x}}(B)$ is evidence that the event $Y \in B$ will happen
  - A low value of $Pl_{Y|\boldsymbol{x}}(B)$ is evidence that the event $Y \in B$ will not happen
- The notions of confidence structure and valid BF are equivalent in the consonant case
- Links with several classical statistical notions (confidence bands/distributions/curves)
- Are there any other relations between the three notions?

References I

T. Denœux.
Constructing belief functions from sample data using multinomial confidence regions
International Journal of Approximate Reasoning 42 (3) (2006) 228–252.

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.

M. S. Balch
Mathematical foundations for a theory of confidence structures,
International Journal of Approximate Reasoning 53 (7) (2012) 1003 – 1019.

References II

R. Martin, C. Liu
Inferential models: A framework for prior-free posterior probabilistic inference
Journal of the American Statistical Association 108 (2013) 301–313.

R. Martin, R. T. Lingham
Prior-free probabilistic prediction of future observations
Technometrics 58 (2) (2016) 225–235.

T. Denœux, S. Li.
Frequency-calibrated belief functions: A review
Submitted to *International Journal of Approximate Reasoning*, 2017
Available at `https://hal.archives-ouvertes.fr/hal-01576356`