Methods for building belief functions

Thierry Denœux[1]

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

Spring School BFTA 2011
Autrans, April 4-8, 2011

Building belief functions

- The basic theory tells us how to reason and compute with belief functions, but it does not tell us where belief functions come from.
- We need formalized methods for modeling expert opinions and statistical information using belief functions.
- Three general approaches:
  - Least Commitment Principle;
  - Using meta-knowledge about information sources (discounting);
  - Predictive belief functions.

Outline

Least Commitment Principle

General approach

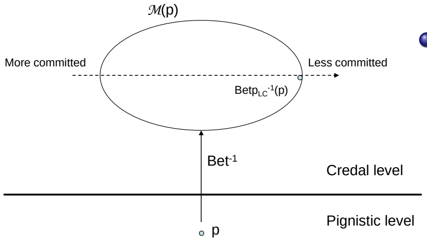
- Least commitment principle: *"When several belief functions are compatible with a set of constraints, the least informative according to some informational ordering (if it exists) should be selected"*.
- General approach:
  1. Express the available information as a set of constraints on an unknown mass function;
  2. Find the least-committed mass function (according to some ordering), compatible with the constraints.
- Three applications:
  - Inverse pignistic transformation;
  - Credal ordering constraints;
  - Deconditioning, Generalized Bayes Theorem (GBT).

Outline

Inverse pignistic transformation
Problem statement

- Assume we want to elicit a mass function m on $\Omega = \{\omega_1, \ldots, \omega_K\}$ from an expert.
- It is easier to elicit the corresponding pignistic probability:
  - For each $\omega_k \in \Omega$ ask for the fair price p_k the expert is willing to pay for a ticket that will allow him to receive 1 euro if $X = \omega_k$, and to receive nothing otherwise.
  - The pignistic probability mass function is $p(\omega_k) = p_k$, $k = 1, \ldots, K$.
- How to compute a mass function m on Ω consistent with p, i.e., such that $p = Bet(m)$?

Inverse pignistic transformation
Discrete case



- There are infinitely many mass functions m such that $Bet(m) = p$.
- The q-least committed solution is a consonant mass function defined by the following possibility distribution:

$$\pi(\omega_k) = \sum_{\ell=1}^{K} \min(p_k, p_\ell).$$

Inverse pignistic transformation

Recovering the mass function

- Let $1 = \pi_{(1)} \geq \pi_{(2)} \geq \ldots \geq \pi_{(K)}$ be the ordered possibility degrees, and $\omega_{(1)}, \ldots, \omega_{(K)}$ the elements of Ω in the corresponding order, i.e., $\pi(\omega_{(i)}) = \pi_{(i)}$, $i = 1, \ldots, K$.
- We have

$$
\begin{aligned}
m(\{\omega_{(1)}\}) &= \pi_{(1)} - \pi_{(2)} \\
&\vdots \\
m(\{\omega_{(1)}, \ldots, \omega_{(i)}\}) &= \pi_{(i)} - \pi_{(i+1)} \\
&\vdots \\
m(\Omega) &= \pi_{(K)}.
\end{aligned}
$$

Inverse pignistic transformation

Example

- Let us consider a frame $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and the pignistic probability mass function

$$p(\omega_1) = 0.7, \quad p(\omega_2) = 0.2, \quad p(\omega_3) = 0.1$$

- We have

$$\begin{aligned} \pi(\omega_1) &= 0.7 + 0.2 + 0.1 = 1 \\ \pi(\omega_2) &= 0.2 + 0.2 + 0.1 = 0.5 \\ \pi(\omega_3) &= 0.1 + 0.1 + 0.1 = 0.3. \end{aligned}$$

- The corresponding mass function is

$$m(\{\omega_1\}) = 0.5, \quad m(\{\omega_1, \omega_2\}) = 0.2, \quad m(\Omega) = 0.3.$$

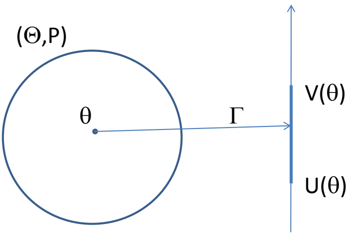
Inverse pignistic transformation

Continuous case

- Assume that the variable of interest X is a continuous variable taking values in $\mathbb{R}$.
- The expert gives us a probability distribution on $\mathbb{R}$. Can we extend the previous line of reasoning to this situation?
- We need to define belief functions on $\mathbb{R}$ and the associated notions (informational orderings, pignistic transformation, etc.).

Belief functions on $\mathbb{R}$

Random intervals



A random interval is defined by a probability space $(\Theta, \mathcal{A}, P)$ and a mapping Γ from Θ to the set $\mathcal{I}$ of closed real intervals:

$$\Gamma : \theta \to \Gamma(\theta) = [U(\theta), V(\theta)],$$

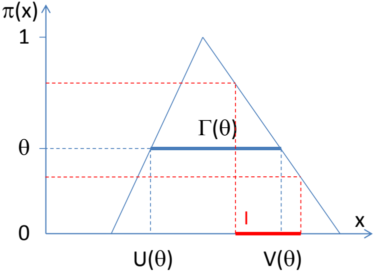
such that (U, V) is a two-dimensional random vector, with $U \leq V$.

We have, for any $I \in \mathcal{I}$:

$$bel(I) = P([U, V] \subseteq I), \quad pl(I) = P([U, V] \cap I \neq \emptyset)$$

$$q(I) = P([U, V] \supseteq I)$$

Random intervals
Example: possibility distribution



- Let π be a possibility distribution on $\mathbb{R}$, $\Theta = [0,1]$, P the Lebesgues measure on $[0,1]$, and $\Gamma(\theta)$ the θ-level cut of π.
- It can be checked that

$$pl(I) = \sup_{x \in I} \pi(x) = \Pi(I)$$

$$bel(I) = 1 - \sup_{x \notin I} \pi(x) = N(I).$$

Pignistic probability density

Discrete case

- Let us assume that $\Gamma(\Theta) = \{I_1, \ldots, I_r\}$. We can define the mass function as

$$m(I_i) = P\left(\{\theta \in \Theta | \Gamma(\theta) = I_i\}\right).$$

- m is a discrete mass function with focal intervals $I_1, \ldots, I_r$.
- Assuming $0 < |I_i| < +\infty$ for all i, the pignistic probability density associated to m is:

$$p_m(x) = \sum_{i=1}^{r} m(I_i) \frac{1_{I_i}(x)}{|I_i|}, \quad \forall x \in \mathbb{R}.$$

- It is a finite mixture of continuous uniform distributions.

Pignistic probability density
Continuous case

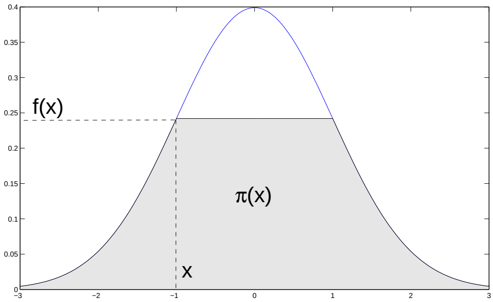
- If (U, V) is a continuous random vector with density f, we can define a "mass density"

$$m([u, v]) = f(u, v), \quad \forall (u, v) \in \mathbb{R}^2, u \leq v.$$

- The pignistic probability density is:

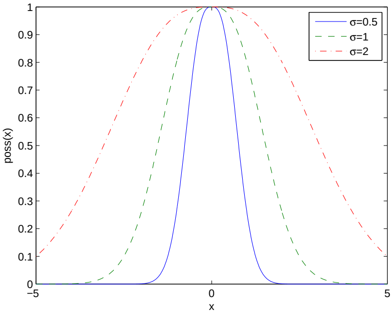
$$p_m(x) = \lim_{\epsilon \to 0} \int_{-\infty}^{x} \int_{x+\epsilon}^{+\infty} \frac{f(u, v)}{v - u} dvdu.$$

Inverse pignistic transformation
General expression



$$\pi(x) = \int_{-\infty}^{+\infty} \min(f(x), f(t))dt.$$

Inverse pignistic transformation
Example: normal distribution



$$\pi(x) = \begin{cases} \frac{2(x-\mu)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + 2\left(1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right) & \text{if } x \geq \mu \\ \frac{2(\mu-x)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + 2\Phi\left(\frac{x-\mu}{\sigma}\right) & \text{otherwise.} \end{cases}$$

Outline

Credal ordering constraint

Problem

- Consider the following problems:
  1. Let X and X' be two variables. Our beliefs on X are represented by m. Additionally, we believe that X' tends to take greater values than X. How to quantify our beliefs on X' using a mass function?
  2. We consider one variable X and two different contexts C and C'. When C holds, our beliefs on X are represented by m. When C' holds, we cannot precisely assess our beliefs on X, but we believe that X tends to take higher values than it does when C holds. How to quantify our beliefs on X in context C'?
- Approach: formalize the notion of “tending to take higher values" as a constraint on a mass function, and find the least-committed solution compatible with that constraint.

Stochastic ordering

- Given two probability distributions P and P' on $\mathbb{R}$, we say that P is stochastically less than or equal to P' ($P \preceq P'$) if

$$P((x,+\infty)) \leq P'((x,+\infty)), \quad \forall x \in \mathbb{R}$$

- Intuitively, this means that distribution P attaches less probability to larger values than P' does.
- Property: the above condition holds holds iff:

$$P \preceq P' \Leftrightarrow \mathbb{E}_P(g) \leq \mathbb{E}_{P'}(g), \quad \forall g \in \mathcal{G}$$

  where $\mathcal{G}$ is the set of measurable and non decreasing real functions.
- How to extend this notion to compare two mass functions m and m' on $\mathbb{R}$?

Credal ordering
Definitions

- Four definitions (credal orderings):
  1. $m \lesssim m'$ iff $bel((x, +\infty)) \leq pl'((x, +\infty)), \quad \forall x \in \mathbb{R}$;
  2. $m \leqslant m'$ iff $bel((x, +\infty)) \leq bel'((x, +\infty)), \quad \forall x \in \mathbb{R}$;
  3. $m \eqslantless m'$ iff $pl((x, +\infty)) \leq pl'((x, +\infty)), \quad \forall x \in \mathbb{R}$;
  4. $m \ll m'$ iff $pl((x, +\infty)) \leq bel'((x, +\infty)), \quad \forall x \in \mathbb{R}$.
- Let $\mathcal{G}_b$ denote the set of bounded, measurable and non decreasing real functions. Then we have:

$$
\begin{aligned}
m \lesssim m' &\Leftrightarrow \underline{\mathbb{E}}_m(g) \leq \overline{\mathbb{E}}_{m'}(g), \quad \forall g \in \mathcal{G}_b\\
m \leqslant m' &\Leftrightarrow \underline{\mathbb{E}}_m(g) \leq \underline{\mathbb{E}}_{m'}(g), \quad \forall g \in \mathcal{G}_b\\
m \eqslantless m' &\Leftrightarrow \overline{\mathbb{E}}_m(g) \leq \overline{\mathbb{E}}_{m'}(g), \quad \forall g \in \mathcal{G}_b\\
m \ll m' &\Leftrightarrow \overline{\mathbb{E}}_m(g) \leq \underline{\mathbb{E}}_{m'}(g), \quad \forall g \in \mathcal{G}_b.
\end{aligned}
$$

Credal ordering constraint

Example of result

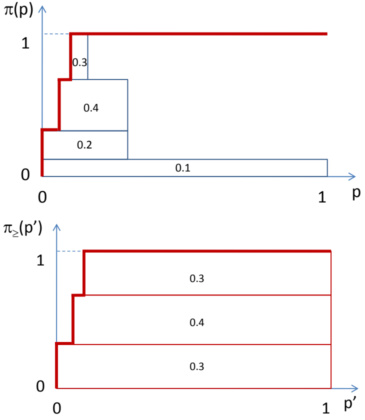
Theorem

The pl-least committed element mass function m' such that $m' \geqslant m$ exists and is unique. It is the consonant mass function $m_{\geqslant}$ with possibility distribution $\pi_{\geqslant}$ given by

$$\pi_{\geqslant}(x) = pl((-\infty, x])$$

where pl is the plausibility function associated to m.

Credal ordering constraint
Example

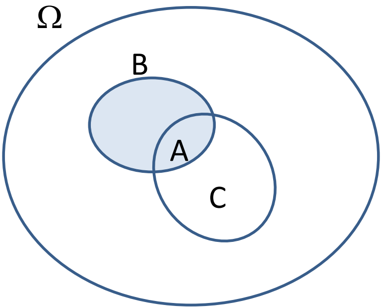


- Assume that m represents an expert's opinion regarding the failure probability p of a component in standard operating condition.
- We want to assess our beliefs regarding the failure probability p' of the same component in a more stringent environment.
- We only know that p' tends to be greater than p: $m_{p'} \geqslant m_p$.

Outline

1. Least Commitment Principle
   - Inverse pignistic transformation
   - Credal ordering constraints
   - Deconditioning
2. Using metaknowledge
   - Discounting
   - Contextual discounting
3. Predictive belief functions
   - Definition
   - Discrete case
   - Continuous case

Deconditioning

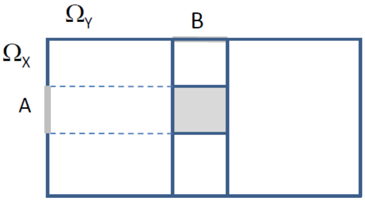


- Let m_0 be a mass function on Ω expressing our beliefs about X in a context where we know that $X \in B$.
- We want to build a mass function m verifying the constraint $m(\cdot|B) = m_0$.
- Any m built from m_0 by transferring each mass $m_0(A)$ to $A \cup C$ for some $C \subseteq \overline{B}$ satisfies the constraint.
- s-least committed solution: transfer $m_0(A)$ to the largest such set $A \cup \overline{B}$:

$$m(D) = \begin{cases} m_0(A) & \text{if } D = A \cup \overline{B} \text{ for some } A \subseteq B, \\ 0 & \text{otherwise} \end{cases}$$

Deconditioning

Ballooning extension



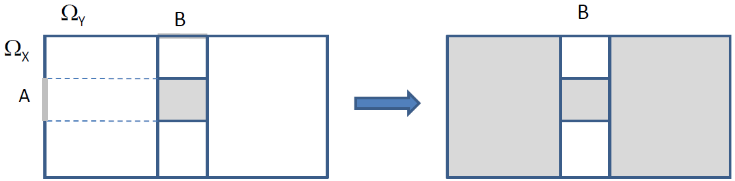
- More complex situation: two frames Ω_X and Ω_Y.
- Let $m_0^{\Omega_X}$ be a mass function on Ω_X expressing our beliefs about X in a context where we know that $Y \in B$ for some $B \subseteq \Omega_Y$.
- We want to find $m^{\Omega_{XY}}$ such that

$$\left(m^{\Omega_{XY}} \textcircled{\cap} (m_B^{\Omega_Y})^{\uparrow \Omega_{XY}}\right)^{\downarrow \Omega_X} = m_0^{\Omega_X}$$

Deconditioning

Ballooning extension (continued)

- s-least committed solution: each mass $m_0^{\Omega_X}(A)$ transferred to $(A \times B) \cup (\Omega_X \times \overline{B})$.



- Notation $m^{\Omega_{XY}} = (m_0^{\Omega_X})^{\Uparrow \Omega_{XY}}$ (ballooning extension).

Application: Generalized Bayes Theorem

Problem statement

- Two variables $X \in \Omega$ et $\theta \in \Theta = \{\theta_1, \ldots, \theta_K\}$.
- Typically:
  - X is observed (sensor measurement),
  - θ is not observed (class, unknown parameter).
- Partial knowledge of X given $\theta = \theta_k$ for each k: $m^\Omega(\cdot|\theta_k)$.
- Prior knowledge about θ: m_0^Θ (may be vacuous).
- We observe $X \in A$.
- Belief function on Θ?

Generalized Bayes Theorem

Solution

- Solution:

$$m^{\Theta}(\cdot|A) = \left(\textcircled{\cap}_{k=1}^{K} m^{\Omega}(\cdot|\theta_k)^{\Uparrow \Omega\times\Theta} \textcircled{\cap} m_A^{\Omega \uparrow \Omega\times\Theta} \textcircled{\cap} m_0^{\Theta \uparrow \Omega\times\Theta} \right)^{\downarrow \Theta}$$

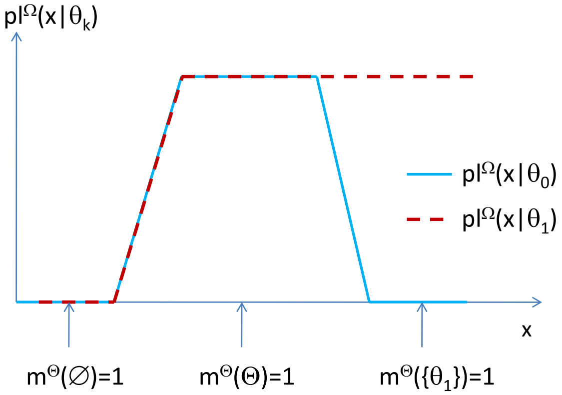
- Expression:

$$m^{\Theta}(\cdot|A) = \textcircled{\cap}_{k=1}^{K} m_k^{\Theta} \textcircled{\cap} m_0^{\Theta},$$

where

$$m_k^{\Theta}(\overline{\{\theta_k\}}) = 1 - pl^{\Omega}(A|\theta_k)$$
$$m_k^{\Theta}(\Theta) = pl^{\Omega}(A|\theta_k)$$

Generalized Bayes Theorem
Example



Generalized Bayes Theorem

Properties

- Property 1: Bayes' theorem is recovered as a special case when the conditional mass functions $m^{\Omega}(\cdot|\theta_k)$ and m_0^{Θ} are Bayesian mass functions.
- Property 2: If X and Y are cognitively independent conditionally on θ, i.e.:

$$pl^{\Omega_X \times \Omega_Y}(A \times B|\theta_k) = pl^{\Omega_X}(A|\theta_k) \cdot pl^{\Omega_Y}(B|\theta_k),$$

for all k, $A \subseteq \Omega_X$ and $B \subseteq \Omega_Y$, then

$$m^{\Theta}(\cdot|X \in A, Y \in B) = m^{\Theta}(\cdot|X \in A) \cap\!\!\!\!\bigcirc m^{\Theta}(\cdot|Y \in B).$$

Outline

Discounting
Problem statement

- A source of information provides:
  - a value;
  - a set of values;
  - a probability distribution, etc..
- The information is:
  - not fully reliable or
  - not fully relevant.
- Examples:
  - Possibly faulty sensor;
  - Measurement performed in unfavorable experimental conditions;
  - Information is related to a situation or an object that only has some similarity with the situation or the object considered (case-based reasoning).

Discounting
Formalization

- A source S provides a mass function m_S^Ω.
- S may be reliable or not. Let $\mathcal{R} = \{R, NR\}$.
- Assumptions:
  - If S is reliable, we accept m_S^Ω as a representation of our beliefs:
  $$m^\Omega(\cdot|R) = m_S^\Omega$$
  - If S is not reliable, we know nothing:
  $$m^\Omega(\cdot|NR) = m_\Omega^\Omega$$
  - The source has a probability $1 - \alpha$ of being reliable:
  $$m^\mathcal{R}(\{NR\}) = \alpha, \quad m^\mathcal{R}(\{R\}) = 1 - \alpha$$

Discounting
Solution

- Solution:

$$^{\alpha}m^{\Omega} = \left(m^{\mathcal{R}\uparrow\Omega\times\mathcal{R}} \textcircled{\cap} m^{\Omega}(\cdot|R)^{\Uparrow\Omega\times\mathcal{R}} \right)^{\downarrow\Omega}.$$

- Simple expressions:

$$\begin{aligned} ^{\alpha}m^{\Omega} &= (1-\alpha)m^{\Omega}_{S} + \alpha m^{\Omega}_{\Omega} \\ &= m^{\Omega}_{S} \textcircled{\cup} m^{\Omega}_{0} \end{aligned}$$

with $m^{\Omega}_{0}(\Omega) = \alpha$ and $m^{\Omega}_{0}(\emptyset) = 1-\alpha$.

- $^{\alpha}m^{\Omega}$ is a s-less committed than (a generalization of) m^{Ω}_{S}:

$$^{\alpha}m^{\Omega} \sqsupseteq_{s} m^{\Omega}_{S}.$$

Outline

Generalization: Contextual Discounting
Formalization

- A more general model allowing us to take into account richer meta-information about the source.
- Let $\Theta = \{\theta_1, \ldots, \theta_L\}$ be a partition of Ω, representing different contexts.
- Let $m^{\mathcal{R}}(\cdot|\theta_k)$ denote the mass function on $\mathcal{R}$ quantifying our belief in the reliability of source S, when we know that the actual value of X is in θ_k.
- We assume that:

$$m^{\mathcal{R}}(\{R\}|\theta_k) = 1 - \alpha_k, \quad m^{\mathcal{R}}(\{NR\}|\theta_k) = \alpha_k.$$

  for each $k \in \{1, \ldots, L\}$.
- Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_L)$.

Contextual Discounting
Example

- Let us consider a simplified aerial target recognition problem, in which we have three classes: airplane ($\omega_1 \equiv a$), helicopter ($\omega_2 \equiv h$) and rocket ($\omega_3 \equiv r$).
- Let $\Omega = \{a, h, r\}$.
- The sensor provides the following mass function: $m_S^\Omega(\{a\}) = 0.5$, $m_S^\Omega(\{r\}) = 0.5$.
- We assume that
  - The probability that the source is reliable when the target is an airplane is equal to $1 - \alpha_1 = 0.4$;
  - The probability that the source is reliable when the target is either a helicopter, or a rocket is equal to $1 - \alpha_2 = 0.9$.
- We have $\Theta = \{\theta_1, \theta_2\}$, with $\theta_1 = \{a\}$, $\theta_2 = \{h, r\}$, and $\boldsymbol{\alpha} = (0.6, 0.1)$.

Contextual Discounting
Solution

- Solution:

$$ {}^{\alpha}m^{\Omega} = \left(\textcircled{\cap}_{k=1}^{L} m^{\mathcal{R}}(\cdot|\theta_k)^{\Uparrow \Omega \times \mathcal{R}} \textcircled{\cap} m^{\Omega}(\cdot|R)^{\Uparrow \Omega \times \mathcal{R}} \right)^{\downarrow \Omega}. $$

- Result:

$$ {}^{\alpha}m^{\Omega} = m_S^{\Omega} \textcircled{\cup} m_1^{\Omega} \textcircled{\cup} \ldots \textcircled{\cup} m_L^{\Omega} $$

with $m_k^{\Omega}(\theta_k) = \alpha_k$ and $m_k^{\Omega}(\emptyset) = 1 - \alpha_k$.

- Standard discounting is recovered as a special case when $\Theta = \{\Omega\}$.

Contextual Discounting

Example (continued)

- The discounted mass function can be obtained by combining disjunctively 3 mass functions:
  - $m_S^\Omega(\{a\}) = 0.5$, $m_S^\Omega(\{r\}) = 0.5$;
  - $m_1^\Omega(\{a\}) = 0.6$, $m_1^\Omega(\emptyset) = 0.4$;
  - $m_1^\Omega(\{h, r\}) = 0.1$, $m_1^\Omega(\emptyset) = 0.9$.
- Result:

A	h	a	r	h, a	h, r	a, r	Ω
$m_S^\Omega(A)$	0	0.5	0.5	0	0	0	0
${}^\alpha m^\Omega(A)$	0	0.45	0.18	0	0.02	0.27	0.08

Outline

Predictive belief functions
Motivation

- Let X be random variable (defined from a repeatable random experiment), with unknown probability $\mathbb{P}_X$.
- We have observed n independent replicates of X:

$$\mathbf{X} = (X_1, \ldots, X_n).$$

- Problem: quantify our beliefs regarding a future realization of X using a belief function $bel(\cdot; \mathbf{X})$: predictive belief function.

Predictive belief functions
Examples

1. Example 1:
   - We have drawn r black balls in n drawings from an urn with replacement:
   - What is our belief that the next ball to be drawn from the urn will be black?
2. Example 2:
   - The lifetimes of 20 bearings have been observed:

     2398, 2812, 3113, 3212, 3523, 5236, 6215,
     6278, 7725, 8604, 9003, 9350, 9460, 11584,
     11825, 12628, 12888, 13431, 14266, 17809.
   - Let X be the lifetime of a bearing taken at random from the same population. Belief function on X?

Predictive belief functions
Requirements

- Requirement 1 (Hacking's frequency principle):
  - If $\mathbb{P}_X$ were known, we would equate our beliefs with probabilities: $bel(\cdot;\mathbb{P}_X) = \mathbb{P}_X$.
  - Weaker version when $\mathbb{P}_X$ is unknown:

$$\forall A \subset \Omega, \quad bel(A;\mathbf{X}) \xrightarrow{P} \mathbb{P}_X(A), \text{ as } n \to \infty,$$

- Requirement 2 (LCP):
  - As n is finite, $bel(\cdot;\mathbf{X})$ should be less committed than $\mathbb{P}_X$. However, the condition $bel(\cdot;\mathbf{X}) \leq \mathbb{P}_X$ is too strong.
  - Weaker requirement:

$$\mathbb{P}\left(bel(A;\mathbf{X}) \leq \mathbb{P}_X(A), \forall A \subset \Omega\right) \geq 1-\alpha.$$

"$bel(\cdot;\mathbf{X})$ is less committed than $\mathbb{P}_X$ most of the time"

Predictive belief functions

Meaning of Requirement 2

$$\mathbf{x} = (x_1, \ldots, x_n) \rightarrow bel(\cdot, \mathbf{x})$$
$$\mathbf{x}' = (x'_1, \ldots, x'_n) \rightarrow bel(\cdot; \mathbf{x}')$$
$$\mathbf{x}'' = (x''_1, \ldots, x''_n) \rightarrow bel(\cdot; \mathbf{x}'')$$
$$\vdots$$

- As the number of realizations of the random sample tends to ∞, the proportion of belief functions less committed than $\mathbb{P}_X$ should tend to a limit at least equal to $1 - \alpha$.

Outline

Using simultaneous confidence intervals

IJAR 42(3):228-252, 2006

- If X is discrete, $\Omega = \{\omega_1, \ldots, \omega_K\}$: a solution can be obtained using a simultaneous confidence intervals on probabilities $p_k = \mathbb{P}(X = \omega_k)$.
- Random intervals $[P_k^-, P_k^+]$, $k = 1, \ldots, K$ are simultaneous confidence intervals at level $1 - \alpha$ if

$$\mathbb{P}\left(P_k^- \leq p_k \leq P_k^+, k = 1, \ldots, K\right) \geq 1 - \alpha$$

- They are asymptotic simultaneous confidence intervals if the above inequality holds in the limit as $n \to \infty$.

Goodman's simultaneous confidence intervals

Asymptotic simultaneous confidence intervals were proposed by Goodman (1965):

$$P_k^- = \frac{b + 2N_k - \sqrt{\Delta_k}}{2(n+b)},$$

$$P_k^+ = \frac{b + 2N_k + \sqrt{\Delta_k}}{2(n+b)},$$

with $N_k = \#\{i | X_i = \omega_k\}$, $b = \chi^2_{1;1-\alpha/K}$ and $\Delta_k = b\left(b + \frac{4N_k(n-N_k)}{n}\right)$.

Goodman's simultaneous confidence intervals
Example

- 220 psychiatric patients categorized as either neurotic, depressed, schizophrenic or having a personality disorder.
- Observed counts: $\mathbf{n} = (91, 49, 37, 43)$.
- Goodman' confidence intervals at confidence level $1 - \alpha = 0.95$:

Diagnosis	N_k/n	P_k^-	P_k^+
Neurotic	0.41	0.33	0.50
Depressed	0.22	0.16	0.30
Schizophrenic	0.17	0.11	0.24
Personality disorder	0.20	0.14	0.27

From confidence intervals to lower probabilities

- To each $\mathbf{p} = (p_1, \ldots, p_K)$ corresponds a probability measure $\mathbb{P}_X$ s.t. $\mathbb{P}_X(\{\omega_k\}) = p_k$ for each k.
- Consequently, simultaneous confidence intervals define a family of probability measures described by the following lower probability measure:

$$P^-(A) = \max\left(\sum_{\omega_k \in A} P_k^-, 1 - \sum_{\omega_k \notin A} P_k^+\right)$$

- P^- satisfies requirements R_1 and R_2:
  - $P^-(A) \xrightarrow{P} \mathbb{P}_X(A)$ as $n \to \infty$, for all $A \subseteq \Omega$,
  - $\mathbb{P}(P^- \leq \mathbb{P}_X) \geq 1 - \alpha$.
- Is it a belief function?

From lower probabilities to belief functions
$K = 2$ or $K = 3$

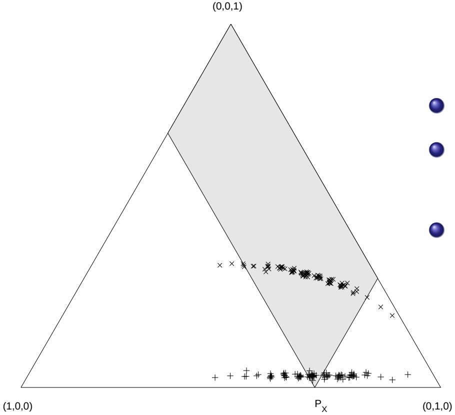
- If $K = 2$ or $K = 3$, P^- is a belief function.
- Case $K = 2$:

$$
\begin{aligned}
m(\{\omega_1\}) &= P_1^- \approx \widehat{p} - u_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \\
m(\{\omega_2\}) &= P_2^- \approx 1 - \widehat{p} - u_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \\
m(\Omega) &= 1 - P_1^- - P_2^- \approx 2u_{1-\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}},
\end{aligned}
$$

with $\widehat{p} = N_1/n$.

The case $K = 2$

Example



- $K = 2$, $p_1 = \mathbb{P}_X(\{\omega_1\}) = 0.3$.
- 100 realizations of a random sample of size $n = 30$.
- 100 predictive belief functions at level $1 - \alpha = 0.95$.

From lower probabilities to belief functions

$K = 2$ or $K = 3$

- If $K > 3$, P^- is not a belief function in general. We can find the most committed belief function satisfying $bel \leq P^-$ by solving the following linear optimization problem:

$$\max_m J(m) = \sum_{A \subseteq \Omega} bel(A) = \sum_{A \subseteq \Omega} \sum_{B \subseteq A} m(B)$$

under the constraints:

$$\sum_{B \subseteq A} m(B) \leq P^-(A), \quad \forall A \subset \Omega,$$

$$\sum_{A \subseteq \Omega} m(A) = 1, \quad m(A) \geq 0, \quad \forall A \subseteq \Omega \, .$$

- The solution satisfies requirements R_1 and R_2: it is a predictive belief function (at confidence level $1 - \alpha$).

The case $K > 3$

Psychiatric Data

A	$P^-(A)$	$bel(A)$	$m(A)$
$\{\omega_1\}$	0.33	0.33	0.33
$\{\omega_2\}$	0.16	0.14	0.14
$\{\omega_1, \omega_2\}$	0.50	0.50	0.021
$\{\omega_3\}$	0.11	0.097	0.097
$\{\omega_1, \omega_3\}$	0.45	0.45	0.020
$\{\omega_2, \omega_3\}$	0.28	0.28	0.036
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\{\omega_1, \omega_3, \omega_4\}$	0.70	0.66	0.038
$\{\omega_2, \omega_3, \omega_4\}$	0.50	0.48	0.019
Ω	1	1	0

Case of ordered data

- Assume Ω is ordered: $\omega_1 < \ldots < \omega_K$.
- The focal sets of *bel* can be constrained to be intervals $A_{k,r} = \{\omega_k, \ldots, \omega_r\}$.
- Under this additional constraint, an analytical solution to the previous optimization problem can be found:

$$m(A_{k,k}) = P_k^-,$$

$$m(A_{k,k+1}) = P^-(A_{k,k+1}) - P^-(A_{k+1,k+1}) - P^-(A_{k,k}),$$

$$m(A_{k,r}) = P^-(A_{k,r}) - P^-(A_{k+1,r}) - P^-(A_{k,r-1}) + P^-(A_{k+1,r-1})$$

for $r > k+1$, and $m(B) = 0$, for all $B \notin \mathcal{I}$.

Example: rain data

- January precipitation in Arizona (in inches), recorded during the period 1895-2004.

class ω_k	n_k	n_k/n	p_k^-	p_k^+
< 0.75	48	0.44	0.32	0.56
$[0.75, 1.25)$	17	0.15	0.085	0.27
$[1.25, 1.75)$	19	0.17	0.098	0.29
$[1.75, 2.25)$	11	0.10	0.047	0.20
$[2.25, 2.75)$	6	0.055	0.020	0.14
≥ 2.75	9	0.082	0.035	0.18

- Degree of belief that the precipitation in Arizona next January will exceed, say, 2.25 inches?

Rain data
Result

$m(A_{k,r})$	1	2	3	4	5	6
1	0.32	0	0	0.13	0.11	0
2	-	0.085	0	0	0.012	0.14
3	-	-	0.098	0	0	0
4	-	-	-	0.047	0	0
5	-	-	-	-	0.020	0
6	-	-	-	-	-	0.035

- We get $bel(X \geq 2.25) = bel(\{\omega_5, \omega_6\}) = 0.055$ and $pl(X \geq 2.25) = 0.317$.
- In 95 % of cases, the interval $[bel(A), pl(A)]$ computed using this method contains $\mathbb{P}_X(A)$.

Outline

Predictive belief functions

Proceedings ISIPTA '07, 11-20, 2007

- If X is absolutely continuous, $\Omega = \mathbb{R}$: a solution can be obtained using a confidence band on the cumulative distribution function F_X of X.
- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an iid sample from X with cdf F_X.
- A pair of functions $(\underline{F}(\cdot; \mathbf{X}), \overline{F}(\cdot; \mathbf{X}))$ computed from $\mathbf{X}$ and such that $\underline{F}(\cdot; \mathbf{X}) \leq \overline{F}(\cdot; \mathbf{X})$ is a confidence band at level $\alpha \in (0, 1)$ if

$$P\left\{\underline{F}(x; \mathbf{X}) \leq F_X(x) \leq \overline{F}(x; \mathbf{X}),\ \forall x \in \mathbb{R}\right\} = 1 - \alpha,$$

Predictive belief functions
Kolmogorov Confidence band

- A non parametric confidence band can be computed using the Kolmogorov statistic:

$$D_n = \sup_x |S_n(x; \mathbf{X}) - F_X(x)|,$$

where $S_n(\cdot; \mathbf{X})$ is the sample cdf.

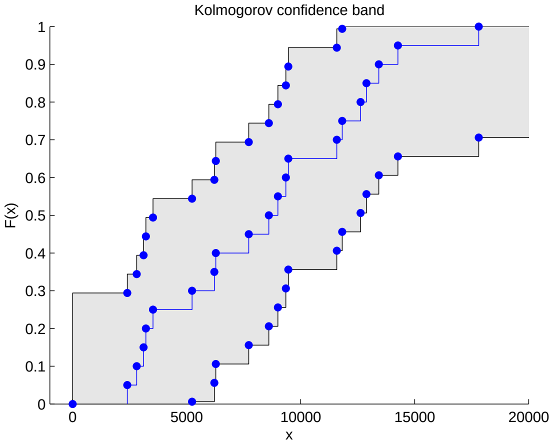
- The probability distribution of D_n can be computed exactly. Let $d_{n,\alpha}$ by the α-critical value of D_n, i.e., $\mathbb{P}(D_n \geq d_{n,\alpha}) = \alpha$.
- The two step functions

$$\begin{aligned} \underline{F}(x; \mathbf{X}) &= \max(0, S_n(x; \mathbf{X}) - d_{n,\alpha}), \\ \overline{F}(x; \mathbf{X}) &= \min(1, S_n(x; \mathbf{X}) + d_{n,\alpha}) \end{aligned}$$

form a confidence band at level $1 - \alpha$.

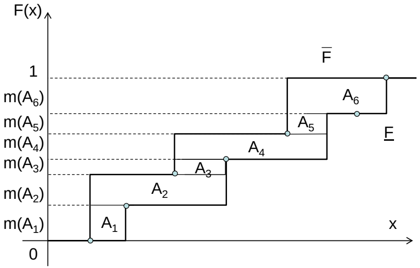
Kolmogorov Confidence band

Bearing data ($1 - \alpha = 0.95$)



Predictive belief functions

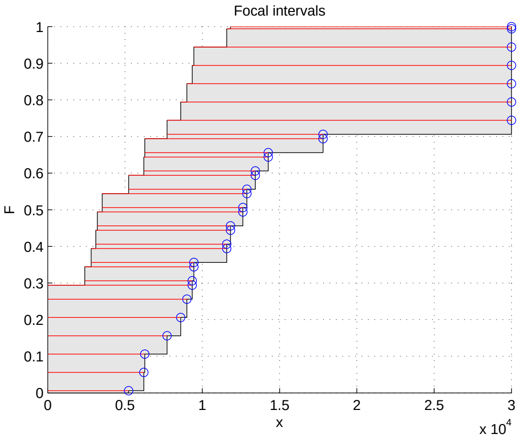
p-boxes and belief functions



- A Kolmogorov confidence band defines a p-box (a set of probability measures with cdf constrained by 2 step functions).
- A p-box defines a discrete random interval.
- The belief function constructed from a Kolmogorov confidence band at level $1-\alpha$ is a predictive belief function at level $1-\alpha$.

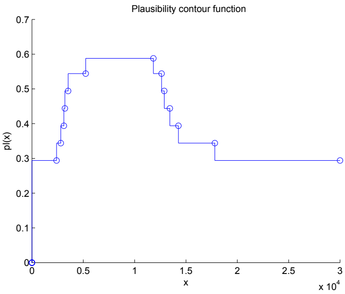
Construction of a mass function from a p-box

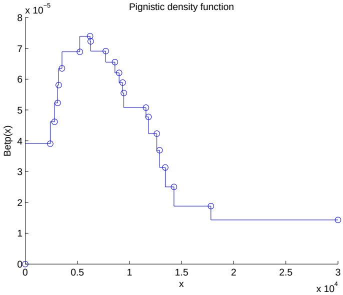
Bearing data



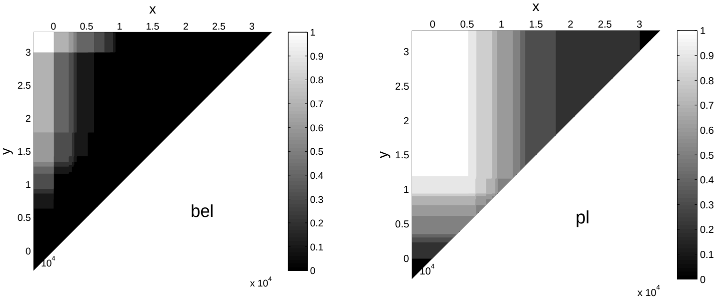
Contour and pignistic density functions

Bearing data





Belief and plausibility functions
Bearing data



Summary

- Developing engineering applications using the Dempster-Shafer framework requires modeling expert knowledge and statistical information using belief functions.
- Systematic and principled methods now exist:
  - Least-commitment principle
  - Discounting
  - GBT
  - Predictive belief functions
  - etc.
- Specific methods will be studied in following lectures (parametric statistical inference, classification, etc.).
- More research on expert knowledge elicitation and statistical inference is needed.

References I

cf. `http://www.hds.utc.fr/˜tdenoeux`

Ph. Smets.
Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem.
International Journal of Approximate Reasoning, 9:1-35, 1993.

T. Denœux.
Extending stochastic ordering to belief functions on the real line.
Information Sciences, Vol. 179, pages 1362-1376, 2009.

D. Mercier, B. Quost and T. Denœux.
Refined modeling of sensor reliability in the belief function framework using contextual discounting.
Information Fusion, Vol. 9, Issue 2, pages 246-258, 2008.

References II

cf. http://www.hds.utc.fr/~tdenoeux

T. Denœux.

Constructing Belief Functions from Sample Data Using Multinomial Confidence Regions.

International Journal of Approximate Reasoning, Vol. 42, Issue 3, Pages 228-252, 2006.

A. Aregui and T. Denœux.

Constructing Predictive Belief Functions from Continuous Sample Data Using Confidence Bands.

In G. De Cooman and J. Vejnarova and M. Zaffalon (Eds), Proceedings of ISIPTA '07, pages 11-20, Prague, Czech Republic, July 2007.