Statistics and Information Fusion

Paul-André Monney
Independent Consultant
Ely, Iowa, USA

Spring School
Belief Functions Theory and Applications
Autrans, April 4–8, 2011

Outline

- Statistical Inference and Belief Functions
- Conditional Embedding for Frequency Distributions
- Functional Models

Statistical Specification

For each $\theta \in \Theta$, the function

$$q_\theta : \ \Theta \longrightarrow X$$

represents the probability distribution of the values $x \in X$ when θ is the correct value of the parameter. This is called a *statistical specification.*

Then an observed outcome is evidence as to which element of Θ is the correct value of θ.

The problem is then to determine a belief function S_x on Θ, also called a support function, that represents the available information.

The value $S_x(A)$ is the degree of support for the hypothesis that the true value of θ is within the subset A of Θ.

Requirements

Intuitive constraint on the plausibility function Pl_x associated with S_x: the contour function $Pl_x(\theta)$ is proportional to the likelihood function $q_\theta(x)$:

$$Pl_x(\theta) = c \cdot q_\theta(x)$$

for all $\theta \in \Theta$ and c is a constant that is independent of θ.

It can be proved that there is one and only one *consonant* plausibility function Pl_x (and hence consonant support, belief function S_x) satisfying this constraint.

Results

The proportionality constant is given by

$$c = (\max_{\theta \in \Theta} q_\theta(x))^{-1}$$

and for all non-empty subsets A of Θ,

$$Pl_x(A) = \frac{\max_{\theta \in A} q_\theta(x)}{\max_{\theta' \in \Theta} q_{\theta'}(x)}$$

$$S_x(A) = 1 - Pl_x(\bar{A}) = 1 - \frac{\max_{\theta \in \bar{A}} q_\theta(x)}{\max_{\theta \in \Theta} q_\theta(x)}$$

Several Observations

The requirement of consonance applies only to the case of a single observation x.

If we have independent observations $x_1, \ldots, x_n$ from the same statistical specification $q_\theta(x)$, then the consonant support functions S_{x_i} can be combined by Dempster's rule into the possibly dissonant belief function

$$S = S_{x_1} \oplus \cdots \oplus S_{x_n}$$

Example

A possibly biased coin.

$$X = \{H, T\}, \qquad \Theta = \{0, 1, \ldots, 10\}$$

$$q_\theta(H) = \frac{\theta}{10} \qquad q_\theta(T) = 1 - \frac{\theta}{10}.$$

In this case, observing a single H produces the support function

$$\begin{aligned} S_H(A) &= 1 - \frac{\max_{\theta \in \bar{A}} q_\theta(H)}{\max_{\theta \in \Theta} q_\theta(H)} \\ &= 1 - \max_{\theta \in \bar{A}} \frac{\theta}{10} \\ &= \min_{\theta \in \bar{A}} \left(1 - \frac{\theta}{10}\right) \end{aligned}$$

Example, cont.

Similarly, the observation of a single T results in the consonant support function

$$S_T(A) = \min_{\theta \in \bar{A}} \frac{\theta}{10}.$$

In general, if n physically independent tosses result in k heads and $n-k$ tails, then it will produce a support function equal to the Dempster's rule combination of k copies of S_H and $n-k$ copies of S_T.

Example, cont.

Two coin tosses : 1 heads and 1 tails. Rounded to the nearest 0.01, degree of support for the set

$$\{i, \ldots, j\}$$

i \ j	1	2	3	4	5	6	7	8	9
1	.02	.07	.13	.22	.33	.47	.62	.80	1.00
2		.02	.07	.13	.22	.33	.47	.62	.80
3			.02	.07	.13	.22	.33	.47	.62
4				.02	.07	.13	.22	.33	.47
5					.02	.07	.13	.22	.33
6						.02	.07	.13	.22
7							.02	.07	.13
8								.02	.07
9									.02

Example, cont.

Ten coin tosses : 5 heads and 5 tails. Rounded to the nearest 0.01, degree of support for the set

$$\{i, \ldots, j\}$$

i \ j	1	2	3	4	5	6	7	8	9
1	.00	.02	.11	.30	.57	.81	.96	1.00	1.00
2		.02	.11	.30	.57	.81	.95	1.00	1.00
3			.08	.26	.53	.77	.91	.95	.96
4				.15	.39	.63	.77	.81	.81
5					.18	.39	.53	.57	.57
6						.15	.26	.30	.30
7							.08	.11	.11
8								.02	.02
9									.00

Aleatory Combination

So far, we have fused independent data at the epistemic level.

We can also fuse independent data at the aleatory level: we consider multiple independent observations as one compound observation.

The compound observation $(x_1, \ldots, x_n)$ is the outcome of a random experiment governed by the product density

$$q_\theta^n : \ X^n \longrightarrow [0, 1]$$

Then the usual method described above can be applied to find the corresponding consonant support function on Θ.

Contrast

In general, the two support functions will be different, but their relative plausibility of singletons are the same.

$$S_{x_1} \oplus \cdots \oplus S_{x_n} \neq S_{x_1,\ldots,x_n}$$

The contrast between the two methods is primarily in their treatment of conflict among the observations.

Bayesian Prior

If we have a Bayesian (probability distribution) prior Bel_0 on Θ, we can combine it with S_x by Dempster's rule.

It can be proved that the result of the combination is a Bayesian (probability distribution) belief function Bel_1 on Θ given by

$$Bel_1(\theta) = \frac{Bel_0(\theta)q_\theta(x)}{\sum_{\theta\in\Theta} Bel_0(\theta)q_\theta(x)}$$

This is the classical Bayes's Theorem !

$$P(\theta|x) = \frac{P(\theta)P(x|\theta)}{\sum_{\theta\in\Theta} P(\theta)P(x|\theta)}$$

Empirical Distributions

Knowledge: empirical distributions $P_\theta(x)$ on the sample space X.

We want a belief function Bel on $\Theta \times X$ that has $P_\theta(x)$ as conditional distribution and is vacuous (non-informative) on Θ.

Once Bel is determined, we can combine it by using Dempster's rule with an actual observation x after vacuous extension on the Cartesian product $\Theta \times X$.

The result is then marginalized to Θ to determine degrees of support and plausibility of hypotheses $A \subseteq \Theta$.

Conditional Embedding

The first step is to represent P_θ by its conditional embedding Bel_θ.

Then all the Bel_θ are combined by Dempster's rule,

$$Bel = \oplus_{\theta \in \Theta} Bel_\theta$$

The focal elements of Bel_θ are in one-to-one correspondence with the elements of X.

Corresponding to $x \in X$ is the focal element

$$\{(\theta, x)\} \cup ((\Theta - \{\theta\}) \times X)$$

with m-value $P_\theta(x)$.

Example

$$\Theta = \{\theta_1, \theta_2\} \qquad X = \{0, 1\}$$

$$P_{\theta_1}(1) = p_1 \quad P_{\theta_2}(1) = p_2$$

Then the focal sets of Bel are

$$\{(\theta_1, 1), (\theta_2, 1)\}, \quad p_1 p_2$$

$$\{(\theta_1, 1), (\theta_2, 0)\}, \quad p_1(1 - p_2)$$

$$\{(\theta_1, 0), (\theta_2, 1)\}, \quad (1 - p_1)p_2$$

$$\{(\theta_1, 0), (\theta_2, 0)\}, \quad (1 - p_1)(1 - p_2).$$

Example, cont.

Conditioning Bel on the observation $x = 1$ yields the degrees of belief

$$Bel(\theta_1|x = 1) = \frac{p_1(1 - p_2)}{1 - (1 - p_1)(1 - p_2)}$$

$$Bel(\theta_2|x = 1) = \frac{(1 - p_1)p_2}{1 - (1 - p_1)(1 - p_2)}$$

If $p_1 = 0$, then

$$Bel(\theta_1|x = 1) = 0$$

$$Bel(\theta_2|x = 1) = 1.$$

If $p_1 = 1$, then

$$Bel(\theta_1|x_1 = 1) = 1 - p_2$$

$$Bel(\theta_2|x_1 = 1) = 0$$

Several Observations

Suppose multiple independent observations $x_1, \ldots, x_n$ obtained from P_θ are available.

Then consider the product distributions P_θ^n on X^n, conditionally embed those to obtain the belief function Bel_θ^n on $\Theta \times X^n$, combine them by Dempster's rule to obtain a belief function Bel^n on $\Theta \times X^n$.

Then vacuously extend the observations $x_1, \ldots, x_n$, combine by Dempster's rule, and then marginalize to Θ.

This gives a belief function $Bel_n(\cdot|x_1, \ldots, x_n)$ on Θ.

Epistemic Combination

Or we could also determine the n belief functions

$$Bel(\cdot|x = x_i)$$

and then combine the results by Dempster's rule to obtain the belief function

$$Bel(\cdot|x_1) \oplus \cdots \oplus Bel(\cdot|x_n)$$

Interestingly, it can be proved that the two methods give the same results,

$$Bel_n(\cdot|x_1, \ldots, x_n) = Bel(\cdot|x_1) \oplus \cdots \oplus Bel(\cdot|x_n)$$

This can be applied to the case of n observations in the previous example.

Functional Models

The statistical experiment $\mathcal{E}$ is described by a functional model.

The functional models defines how data x is generated from the unknown parameter and some stochastic element,

$$x = f(\theta, \omega).$$

The probability distribution of the random element ω is assumed to be known. There is a known probability measure P on the domain Ω of ω.

Example

Fair coin, faces 1 and 2.

Two possible cases:

a) θ_0: face 1 carries heads and face 2 carries tails.

b) θ_1: both faces carry heads.

The experiment consist in throwing the coin and observing the outcome (heads or tails).

The chance element is whether face 1 or face 2 turns up, with probability 0.5 each.

$\Theta = \{\theta_0, \theta_1\},\ \Omega = \{1, 2\},$
$X = \{heads,\ tails\},\ P(1) = P(2) = 0.5.$

$$x = \begin{cases} heads & \text{if } \theta = \theta_0 \text{ and } \omega = 1, \\ heads & \text{if } \theta = \theta_1, \\ tails & \text{if } \theta = \theta_0 \text{ and } \omega = 2. \end{cases}$$

Assumption-Based Reasoning

- Consider a functional model f describing an experiment $\mathcal{E}$. Let P denote the probability measure on Ω.
- The outcome of the experiment is observed to be x.
- Given data x and the experiment $\mathcal{E}$, what inference can be made about the value of the unknown parameter θ ?
- The basic idea of assumption-based reasoning is to assume a random element ω and determine what can be deduced about the unknown parameter under this assumption.
- Afterwards, the deductions are weighed by the probabilities of the assumptions.

Analysis

- Given data x, the set of possible assumptions is

$$v_x = \{\omega \in \Omega : \exists\ \theta \in \Theta\ s.t.\ x = f(\theta, \omega)\}.$$

- Conditioning of P on v_x,

$$p'(\omega) = \frac{p(\omega)}{P(v_x)}.$$

- If $\omega \in v_x$ is the chance element that caused the observation x, then the possible parameter values can be restricted to the set

$$T_x(\omega) = \{\theta \in \Theta : x = f(\theta, \omega\}.$$

- Summary represented by a hint

$$\mathcal{H}_x = (v_x, p', T_x, \Theta)$$

Mathematical Theory of Hints (Kohlas & Monney, 1995).

Example, cont.

- We observe $x = heads$.
- The set of compatible assumptions is

$$v_{heads} = \{1, 2\} = \Omega.$$

  So no conditioning is necessary.

- We have

$$T_{heads}(1) = \{\theta_0, \theta_1\}, \quad T_{heads}(2) = \{\theta_1\}.$$

- The degree of support of the hypothesis that the coin carries both heads and tails is

$$sp_{heads}(\theta_0) = 0$$

- The degree of support of the hypothesis that the coin carries two heads is

$$sp_{heads}(\theta_1) = 0.5$$

Example, cont.

- We observe $x = tails$.
- The set of compatible assumptions is

$$v_{tails} = \{2\}$$

- Conditioning: $p'(2) = \frac{p(2)}{p(2)} = 1$.
- We have

$$T_{tails}(2) = \{\theta_0\}$$

- The degree of support of the hypothesis that the coin carries both heads and tails is

$$sp_{tails}(\theta_0) = 1$$

- The degree of support of the hypothesis that the coin carries two heads is

$$sp_{tails}(\theta_1) = 0$$

Repeated Observations

- For $i = 1, \ldots, n$, the individual functional model is

$$x_i = f_i(\theta, \omega_i) \quad x_i \in X_i,\ \omega_i \in \Omega_i,\ \theta \in \Theta$$

- Compound functional model $x = f(\theta, \omega)$ where $x = (x_1, \ldots, x_n) \quad \omega = (\omega_1, \ldots, \omega_n)$ and

$$f(\theta, \omega) = (f_1(\theta, \omega_1), \ldots, f_n(\theta, \omega_n)).$$

- **Theorem 1.** If the experiments are independent, then

$$\mathcal{H}_x = \mathcal{H}_{x_1} \oplus \mathcal{H}_{x_2} \oplus \cdots \oplus \mathcal{H}_{x_n}.$$

- Simple coin example: suppose that heads is observed n times in n throws. We get

$$sp_n(\theta_1) = 1 - 0.5^n \quad pl_n(\theta_1) = 1$$

$$sp_n(\theta_0) = 0 \quad pl_n(\theta_0) = 0.5^n$$

Statistical Model

- A functional model induces a statistical model given by

$$p_\theta(x) = \sum_{\omega : x = f(\theta, \omega)} p(\omega).$$

- Such a statistical model is usually a primary object in Bayesian methods.
- In general, several different functional models induce the same statistical model.
- This means that a functional model generally contains more information than a Bayesian model.
- For a given statistical model, the likelihood function associated with observation x is

$$l_x : \Theta \longrightarrow [0, 1]$$

$$l_x(\theta) = p_\theta(x).$$

Likelihood and Plausibility

- Given a functional model and an observation x, the function

$$pl_x : \Theta \longrightarrow [0, 1]$$

  associated with the inferred hint $\mathcal{H}_x$ is called the contour function.

- **Theorem 2.** Given a functional model and an observation x, the contour function is proportional to the likelihood function. In other words, there is a positive constant c such that

$$pl_x(\theta) = c \cdot l_x(\theta).$$

- This provides a clear interpretation for the likelihood function: up to a multiplicative constant, the likelihood of θ measures the level of compatibility between the data and the hypothesis that θ is the correct value.

Fiducial Probability

- In the statistical literature, the normalized likelihood function is called the fiducial distribution of the parameter (Fisher, 1930).
- The fiducial probability of the parameter θ is

$$\varphi(\theta) = \frac{l_x(\theta)}{\sum_{\theta \in \Theta} l_x(\theta)}$$

- There are some discussions about the meaning of fiducial probability.
- In the context of functional models, the meaning is clear because, for invertible models, we have

$$\varphi(\theta) = pl_x(\theta) = sp_x(\theta).$$

  The fiducial probability of θ is the degree of support for the hypothesis θ provided by the data.

Including Additional Information

- First piece of information: the experiment $\mathcal{E}$ and the data $x \longrightarrow$ hint $\mathcal{H}_x$ on Θ.
- Second piece of information: knowledge about the value of the parameter given by a hint $\mathcal{H}_0$ on Θ.
- The complete information on θ is obtained by combining both hints by Dempster's rule,

$$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_x.$$

Probabilistic Information

- Special case: prior probability distribution $p_0(\theta)$ on the parameter $\longrightarrow$ hint $\mathcal{H}_0 = (\Theta, p_0, id, \Theta)$.
- The combined hint $\mathcal{H}$ is a precise hint whose plausibility function is a probability measure given by

$$pl(\theta) = \frac{p_0(\theta)p_\theta(x)}{\sum_{\theta\in\Theta} p_0(\theta)p_\theta(x)}.$$

- This is the Bayesian theorem, but with a different meaning to the posterior distribution.
- Only the statistical model associated with the functional model is needed.
- Since $pl(\theta) = sp(\theta)$, this is also the degree of support for the hypothesis.