Forecasting using belief functions

Application to innovation diffusion

Thierry Denœux[1], Orakanya Kanjanatarakul[2], Songsak Sriboonchitta[3]

[1]Heudiasyc, CNRS, Université de Technologie de Compiègne, France
[2]Faculty of Management Sciences, Chiang Mai Rajabhat University, Thailand
[3]Faculty of Economics, Chiang Mai University, Thailand

TES 2014,
Chiang Mai, Thailand,
9 January, 2014

Motivation

- Forecasting quantities of interest based on past observations is an important issue in econometrics.
- The uncertainty of the forecast is usually considered very important information to be provided to the decision-maker.
- Usual formalisms for describing forecast uncertainty:
  1. Prediction intervals (how to combine with utilities for rational decision-making?);
  2. Bayesian predictive probability distributions (rely on prior probability distribution).
- In this talk,
  - We argue that the theory of belief functions is a valuable alternative model to describe forecast uncertainty;
  - We apply this approach to the prediction of innovation diffusion.

Outline

1. Theory of belief functions
   - Representation of evidence
   - Combination of evidence
2. Statistical Inference and forecasting
   - Statistical inference
   - Forecasting
3. Application to innovation diffusion
   - Bass model
   - Sales forecasting
   - Example

Outline

Historical perspective

- Also known as Dempster-Shafer theory or Evidence theory.
- Initially introduced by Dempster (1966, 1968) with the objective to reconcile Bayesian and fiducial inference.
- Shafer (1976) later formalized this approach as a general framework for reasoning and decision-making under uncertainty.
- Many applications in statistics, artificial intelligence, risk analysis, etc.

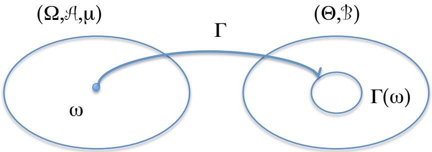
Main features

- The theory of belief function subsumes both the logical and probabilistic approaches to uncertainty: a belief function may be seen as
  - a non-additive measure or as
  - a generalized set.
- The belief function approach coincides with the Bayesian approach when all variables are described by probability distributions.
- However, due to its greater expressive power, the theory of belief functions allows us to handle more general forms of information.

Outline

Belief and plausibility functions

Mathematical definitions



- Let $(\Omega, \mathcal{A}, \mu)$ be a probability space, $(\Theta, \mathcal{B})$ a measurable space, and Γ be a multivalued mapping from Ω to $\mathcal{B}$ defining a random set.
- Belief and plausibility functions on Θ are defined as follows (assuming certain measurability requirements): for all $B \in \mathcal{B}$,

$$Bel(B) = \mu(\{\omega \in \Omega | \Gamma(\omega) \subseteq B\})$$

$$Pl(B) = \mu(\{\omega \in \Omega | \Gamma(\omega) \cap B \neq \emptyset\}) = 1 - Bel(\overline{B}).$$

Belief and plausibility functions

Interpretation

- Typically, Θ is the domain of an unknown quantity θ, and Ω is a set of interpretations of a given piece of evidence about θ, only one of which is true.
- If $\omega \in \Omega$ holds, then the evidence tells us that $\Gamma(\omega) \ni \theta$, and nothing more.
- Then
  - $Bel(B)$ is the probability that the evidence implies B ;
  - $Pl(B)$ is the probability that the evidence is consistent with B.
- Obviously, $Bel(A) \leq Pl(A)$ for all A.
- *Bel* and *Pl* are non additive in general:

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$$

$$Pl(A \cup B) \leq Pl(A) + Pl(B) - Pl(A \cap B)$$

Special cases

1. The sets $\Gamma(\omega) \subseteq \Theta$ are called the focal sets of *Bel*.
2. If there is only one focal set A, then the evidence tells us that $\theta \in A$ for sure, and nothing more. The corresponding belief function $Bel_{\{A\}}$ is said to be logical. In particular, the vacuous belief function $Bel_{\{\Theta\}}$ encodes complete ignorance.
3. If all focal sets are singletons, then $Bel = Pl$ is a probability measure.
4. If the focal sets are nested, *Bel* is said to be consonant. We then have:

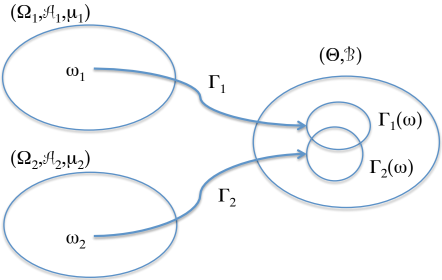
$$Pl(A) = \sup_{\theta \in A} pl(\theta)$$

where $pl : \theta \to Pl(\{\theta\})$ is the contour function of *Bel*.

Outline

Dempster's rule
Definition

Let us assume that we have two pieces of evidence that induce belief functions Bel_1 and Bel_2 on Θ.



- If interpretations ω_1 and ω_2 both hold, we know that $\theta \in \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$.
- If the two pieces of evidence are independent, the probability that ω_1 and ω_2 both hold is $\mu_1(\omega_1)\mu_2(\omega_2)$.
- If $\Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) = \emptyset$, we know that ω_1 and ω_2 cannot hold simultaneously. The joint probability distribution on $\Omega_1 \times \Omega_2$ must be conditioned to eliminate such pairs.
- This random set induces a new combined belief function $Bel_{12} = Bel_1 \oplus Bel_2$.

Dempster's rule

Properties

- Commutativity, associativity. Neutral element: $Bel_{\{\Theta\}}$.
- Generalization of intersection: if $Bel_{\{A\}}$ and $Bel_{\{B\}}$ are logical belief functions and $A \cap B \neq \emptyset$, then

$$Bel_{\{A\}} \oplus Bel_{\{B\}} = Bel_{\{A \cap B\}}$$

- Generalization of probabilistic conditioning: if P is a probability measure and $Bel_{\{A\}}$ is a logical function, then $P \oplus Bel_{\{A\}}$ is the conditional probability measure $P(\cdot|A)$.

Outline

1. Theory of belief functions
   - Representation of evidence
   - Combination of evidence

2. Statistical Inference and forecasting
   - Statistical inference
   - Forecasting

3. Application to innovation diffusion
   - Bass model
   - Sales forecasting
   - Example

Outline

The problem

- We consider a statistical model $\{f_\theta(x), x \in \mathbb{X}, \theta \in \Theta\}$, where $\mathbb{X}$ is the sample space and Θ the parameter space.
- Having observed x, how to quantify the uncertainty about Θ, without specifying a prior probability distribution?
- Two solutions using belief functions:
    1. Dempster's solution based an auxiliary variable with a pivotal probability distribution (Dempster, 1967);
    2. Likelihood-based approach (Shafer, 1976; Wasserman, 1990; Denœux, 2013).

Likelihood-based belief function

Requirements

- Let Bel_x^Θ be a belief function representing our knowledge about θ after observing x. We impose the following requirements:
    1. Likelihood principle: Bel_x^Θ should be based only on the likelihood function $\theta \to L_x(\theta) = f_\theta(x)$.
    2. Compatibility with Bayesian inference: when a Bayesian prior P_0 is available, combining it with Bel_x^Θ using Dempster's rule should yield the Bayesian posterior:

    $$Bel_x^\Theta \oplus P_0 = P(\cdot|x).$$

    3. Principle of minimal commitment: among all the belief functions satisfying the previous two requirements, Bel_x^Θ should be the least committed (least informative).

Likelihood-based belief function

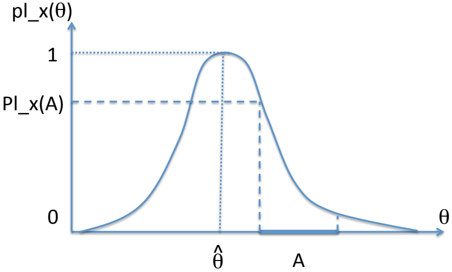
Solution (Denœux, 2013)

- Bel_x^Θ is the consonant belief function with contour function equal to the normalized likelihood:

$$pl_x(\theta) = \frac{L_x(\theta)}{L_x(\hat{\theta})},$$

where $\hat{\theta}$ is the MLE of θ.

- Corresponding plausibility function:

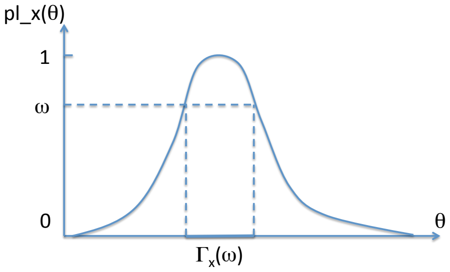


$$Pl_x^\Theta(A) = \sup_{\theta \in A} pl_x(\theta), \quad \forall A \subseteq \Theta$$

Equivalent random set

- Corresponding random set: $(\Omega, \mathcal{B}(\Omega), \lambda, \Gamma_x)$ with $\Omega = [0, 1]$, λ is the Lebesgue measure and

$$\Gamma_x(\omega) = \{\theta \in \Theta | pl_x(\theta) \geq \omega\}.$$



- $\lambda(\{\omega \in \Omega | \Gamma_x(\omega) \cap A \neq \emptyset\}) = \sup_{\theta \in A} pl_x(\theta)$.

Discussion

- The likelihood-based method is easy to implement, even for complex models.
- By construction, it boils down to Bayesian inference when a Bayesian prior is available.
- It is compatible with usual likelihood-based inference:
  - Assume that $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$ and θ_2 is a nuisance parameter. The marginal contour function on Θ_1

$$pl_x(\theta_1) = \sup_{\theta_2 \in \Theta_2} pl_x(\theta_1, \theta_2) = \frac{\sup_{\theta_2 \in \Theta_2} L_x(\theta_1, \theta_2)}{\sup_{(\theta_1, \theta_2) \in \Theta} L_x(\theta_1, \theta_2)}$$

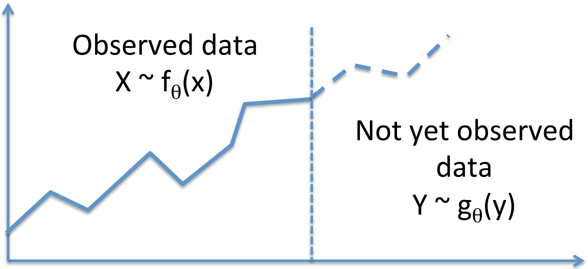
    is the relative profile likelihood function.
  - Let $H_0 \subset \Theta$ be a composite hypothesis. Its plausibility

$$Pl_x^{\Theta}(H_0) = \frac{\sup_{\theta \in H_0} L_x(\theta)}{\sup_{\theta \in \Theta} L_x(\theta)}.$$

    is the usual likelihood ratio statistics $\Lambda(x)$.

Outline

The forecasting problem



Sampling model

- We consider a sampling model of the form

$$Y = \varphi(\theta, Z),$$

  where
  - $Z \in \mathbb{Z}$ is an unobserved auxiliary variable with known probability distribution μ independent of θ;
  - φ is defined in such a way that the distribution of Y for fixed θ is $g_\theta(y)$.
- Example 1: $Y \sim \mathcal{B}(\theta)$ can be written as:

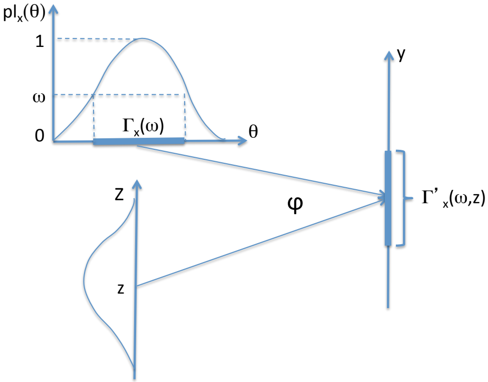
$$Y = \varphi(\theta, Z) = \begin{cases} 1 & \text{if } Z \leq \theta \\ 0 & \text{otherwise,} \end{cases} \quad \text{with } Z \sim \mathcal{U}([0,1]).$$

- Example 2: $Y \sim \mathcal{N}(m, \sigma)$ can be written as:

$$Y = m + \sigma Z \quad \text{with } Z \sim \mathcal{N}(0,1)$$

Predictive belief function

From the equation $Y = \varphi(\theta, Z)$, the random set (belief function) on θ an the probability distribution of Z, we can deduce a random set (belief function) on Y:



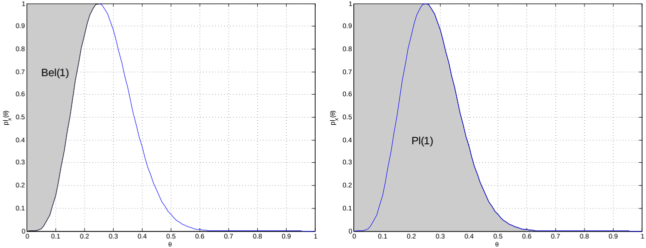
- Let $\Gamma'_x : [0,1] \times \mathbb{Z} \to 2^{\mathbb{Y}}$ be the multi-valued mapping s.t. $\Gamma'_x(\omega, z) = \varphi(\Gamma_x(\omega), z)$.
- The product measure $\lambda \otimes \mu$ on $[0,1] \times \mathbb{Z}$ and Γ'_x induce a predictive belief function on $\mathbb{Y}$.

Example

- Assume that $X \sim \mathcal{B}(n,\theta)$ and $Y \sim \mathcal{B}(1,\theta)$. We have:

$$pl_x(\theta) = \frac{\theta^x(1-\theta)^{n-x}}{\hat\theta^x(1-\hat\theta)^{n-x}} = \left(\frac{\theta}{\hat\theta}\right)^{n\hat\theta}\left(\frac{1-\theta}{1-\hat\theta}\right)^{n(1-\hat\theta)},$$

  for all $\theta \in \Theta = [0,1]$, where $\hat\theta = x/n$.
- Predictive belief function:



Outline

Innovation diffusion

- Forecasting the diffusion of an innovation has been a topic of considerable interest in the last fifty years.
- Typically, when a new product is launched, sale forecasts have to be based on little data and uncertainty has to be quantified to avoid making wrong business decisions based on unreliable forecasts.
- The approach described in this paper uses the Bass model (Bass, 1969) for innovation diffusion together with past sales data to quantify the uncertainty on future sales using the formalism of belief functions.

Outline

Bass model

- Fundamental assumption (Bass, 1969): the probability that an initial purchase of an innovative product will be made at t, given that no purchase has yet been made, is an affine function of the number of previous buyers.
- This implies that the probability $\Phi_\theta(t)$ that an individual taken at random from the population will buy the product before time t is

$$\Phi_\theta(t) = \frac{c(1 - \exp[-(p+q)t])}{1 + (p/q)\exp[-(p+q)t]},$$

where

  - p is the coefficient of innovation;
  - q the coefficient of imitation;
  - c is the probability of eventually adopting the product;
  - $\theta = (p, q, c)$.

Parameter estimation

- Data: $x_1, \ldots, x_{T-1}$, where x_i = observed number of adopters in time interval $[t_{i-1}, t_i)$.
- The number of individuals in the sample of size M who did not adopt the product at time t_{T-1} is $x_T = M - \sum_{i=1}^{T-1} x_i$.
- The probability of adopting the innovation between times t_{i-1} and t_i is $p_i = \Phi_\theta(t_i) - \Phi_\theta(t_{i-1})$ for $1 \leq i \leq T-1$, and the probability of not adopting the innovation before t_{T-1} is $p_T = 1 - \Phi_\theta(t_{T-1})$.
- Consequently, $\mathbf{x} = (x_1, \ldots, x_T)$ is a realization of $\mathbf{X} \sim \mathcal{M}(M, p_1, \ldots, p_T)$ and the likelihood function is

$$L_{\mathbf{x}}(\theta) \propto \prod_{i=1}^{T} p_i^{x_i} = \left(\prod_{i=1}^{T-1} [\Phi_\theta(t_i) - \Phi_\theta(t_{i-1})]^{x_i} \right) [1 - \Phi_\theta(t_{T-1})]^{x_T}.$$

- The belief function on θ is defined by $pl_{\mathbf{x}}(\theta) = L_{\mathbf{x}}(\theta)/L_{\mathbf{x}}(\hat{\theta})$.

Outline

Problem formulation

- Let us assume we are at time at time t_{T-1} and we wish to forecast the number Y of sales between times τ_1 and τ_2, with $t_{T-1} \leq \tau_1 < \tau_2$.
- Y has a binomial distribution $\mathcal{B}(Q, \pi_\theta)$, where
  - Q is the number of potential adopters at time $T-1$;
  - π_θ is the probability of purchase for an individual in that period, given that no purchase has been made before t_{T-1}:

$$\pi_\theta = \frac{\Phi_\theta(\tau_2) - \Phi_\theta(\tau_1)}{1 - \Phi_\theta(t_{T-1})}.$$

- Y can be written as $Y = \varphi(\theta, \mathbf{Z}) = \sum_{i=1}^{Q} \mathbb{1}_{[0,\pi_\theta]}(Z_i)$, where

$$\mathbb{1}_{[0,\pi_\theta]}(Z_i) = \begin{cases} 1 & \text{if } Z_i \leq \pi_\theta \\ 0 & \text{otherwise} \end{cases}$$

and $\mathbf{Z} = (Z_1, \ldots, Z_Q)$ has a uniform distribution in $[0,1]^Q$.

Predictive belief function

Multi-valued mapping

- The predictive belief function on Y is induced by the multi-valued mapping $(\omega, \mathbf{z}) \rightarrow \Gamma'_{\mathbf{x}}(\omega, \mathbf{z}) = \varphi(\Gamma_x(\omega), \mathbf{z})$.
- The range of π_θ when θ varies in $\Gamma_x(\omega)$ is $[\pi_\theta^L(\omega), \pi_\theta^U(\omega)]$, with

$$\pi_\theta^L(\omega) = \min_{\{\theta | pl_x(\theta) \geq \omega\}} \pi_\theta,$$

$$\pi_\theta^U(\omega) = \max_{\{\theta | pl_x(\theta) \geq \omega\}} \pi_\theta.$$

- We have

$$\varphi(\Gamma_x(\omega), \mathbf{z}) = [Y^L(\omega, \mathbf{z}), Y^U(\omega, \mathbf{z})],$$

where $Y^L(\omega, \mathbf{z})$ and $Y^U(\omega, \mathbf{z})$ are, respectively, the number of z_i's that are less than $\pi_\theta^L(\omega)$ and $\pi_\theta^U(\omega)$.

Predictive belief function
Calculation

- The belief and plausibilities that Y will be less than, or equal to y are equal to

$$Bel_x^{\mathbb{Y}}([0,y]) = \int_0^1 F_{Q,\pi_\theta^L(\omega)}(y)d\omega$$
$$Pl_x^{\mathbb{Y}}([0,y]) = \int_0^1 F_{Q,\pi_\theta^U(\omega)}(y)d\omega,$$

where $F_{Q,p}$ denotes the cdf of the binomial distribution $\mathcal{B}(Q,p)$.

- The contour function of Y is

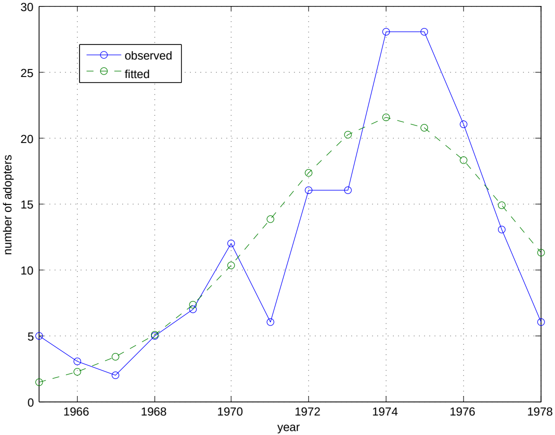
$$pl_x(y) = \int_0^1 \left(F_{Q,\pi_\theta^L(\omega)}(y) - F_{Q,\pi_\theta^U(\omega)}(y-1)\right) d\omega.$$

- Theses integrals can be approximated by Monte-Carlo simulation.

Outline

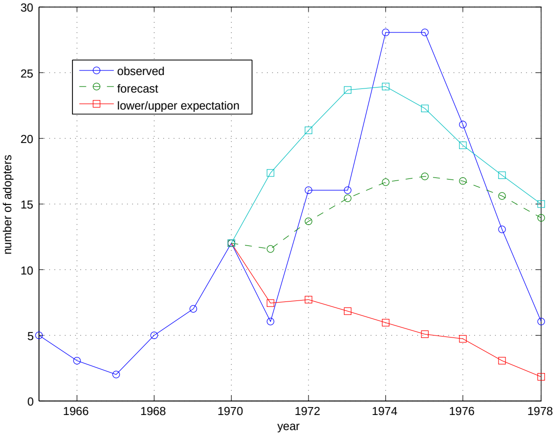
Ultrasound data

Data collected from 209 hospitals through the U.S.A. (Schmittlein and Mahajan, 1982) about adoption of an ultrasound equipment.



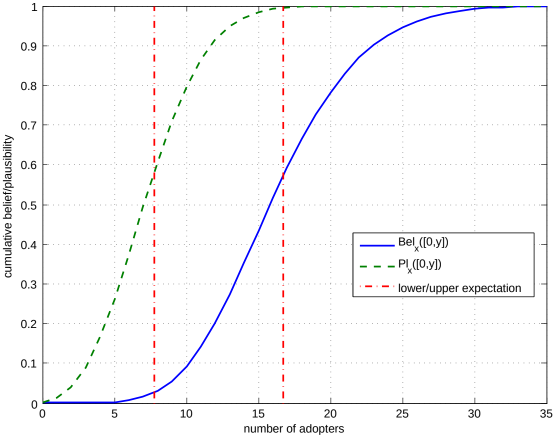
Forecasting

Predictions made in 1970 for the number of adopters in the period 1971-1978, with their lower and upper expectations:



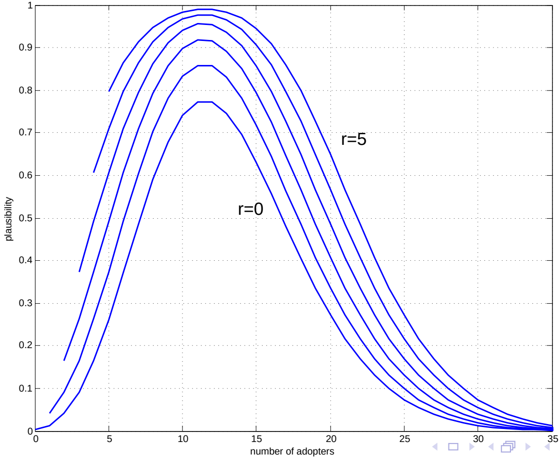
Cumulative belief and plausibility functions

Lower and upper cumulative distribution functions for the number of adopters in 1971, forecasted in 1970:



Plausibilities

Plausibilities $Pl_x^{\mathbb{Y}}([y-r, y+r])$ as functions of y, from $r=0$ (lower curve) to $r=5$ (upper curve), for the number of adopters in 1971, forecasted in 1970:



Conclusions

- Uncertainty quantification is an important component of any forecasting methodology. The approach introduced in this paper allows us to represent forecast uncertainty in the belief function framework, based on past data and a statistical model.
- The method is based on two steps:
  1. Estimation: evidence on the parameter θ is represented by a consonant belief function defined from the normalized likelihood function.
  2. Prediction: the quantity of interest Y is written as $\varphi(\theta, Z)$, where Z is an auxiliary variable; beliefs on θ and Z are then propagated through φ, resulting in a belief function on Y.
- The Bayesian predictive probability distribution is recovered when a prior on θ is available.
- The belief function formalism makes it possible to combine information from several sources (such as expert opinions and statistical data) and fits with a utility-based decision framework.

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

- T. Denœux.
  Likelihood-based belief function: justification and some extensions to low-quality data.
  International Journal of Approximate Reasoning (in press), 2014.

- O. Kanjanatarakul, S. Sriboonchitta and T. Denœux
  Forecasting using belief functions. An application to marketing econometrics.
  International Journal of Approximate Reasoning, To appear, 2014.