Statistical estimation and prediction using belief functions

Thierry Denœux[1]

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

CIMI Imprecise Probability Workshop
Toulouse, May 28, 2015

Motivations

Uncertainty in estimation and prediction

- Estimation vs. prediction:
  - Estimation problem: given some randomly generated observations, make statements about the generating process (population)
  - Prediction problem: given some randomly generated observations, make statements about future observations to be drawn from the same (or a related) process (population)
- In each case, the statements we can make are based on partial knowledge and are thus subject to uncertainty
- Describing this uncertainty in a formal way is an important issue in statistics

Motivations

Limitations of classical approaches

- Frequentist and Bayesian methods are, by far, the most popular approaches
- They provide reasonable conclusions most of the time, but they have some conceptual and practical shortcomings:
  - Frequentist methods provide pre-experimental measures of the accuracy of statistical evidence, which are not conditioned on specific data (a 95% confidence interval contains the parameter of interest for 95% of the samples, but the 95% value is just an average, and the interval may certainly –or certainly not– contain the parameter for some specific samples)
  - Bayesian methods require the statistician to provide a prior probability distribution, which is problematic when no prior knowledge, or only weak information, is available

Motivations
Approach proposed in this talk

- In this talk, I advocate an approach to statistical estimation and prediction, based on the theory of belief functions
  - First proposed for estimation by Shafer (1976) and studied by Wasserman (1990), among others
  - In line with likelihood-based inference as advocated by Fisher in his later work (Fisher, 1922) and, later, by Birnbaum (1962), Barnard (1962) and Edwards (1992), etc.
  - Retains the idea that "all we need to know about the result of a random experiment is contained in the likelihood function", but reinterprets it as defining a consonant belief function
- The method was recently extended from estimation to prediction (Kanjanataraul et al., IJAR, 2014)
- It boils down to bayesian inference when probabilistic prior information is available

Outline

Outline

Dempster-Shafer theory

- The Dempster-Shafer (DS) theory of belief functions (Dempster, 1966; Shafer, 1976) is now a well established formal framework for reasoning with uncertainty
- It has been successfully applied to many problems, including sensor fusion, classification and clustering , image segmentation, state estimation, scene perception, etc.
- In spite of the initial focus on statistical inference, the application of DS theory in this area has remained limited, partly because of the complexity of Dempster's initial method of inference
- Here, I present a tractable approach to statistical inference and prediction using belief functions

Belief function

Definition

Let $(\Omega, \mathcal{B})$ *be a measurable space. A* *belief function (BF)* *on* $\mathcal{B}$ *is a mapping* $Bel : \mathcal{B} \rightarrow [0, 1]$ *verifying the following three conditions:*

1. $Bel(\emptyset) = 0$;
2. $Bel(\Omega) = 1$;
3. *Bel is* *completely monotone*, *i.e., for any* $k \geq 2$ *and any collection* $B_1, \ldots, B_k$ *of elements of* $\mathcal{B}$,

$$Bel\left(\bigcup_{i=1}^{k} B_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} B_i\right).$$

Plausibility function

Definition

A plausibility function on $\mathcal{B}$ is a mapping $Pl : \mathcal{B} \rightarrow [0,1]$ s.t. $Pl(\emptyset) = 0$, $Pl(\Omega) = 1$ and for any $k \geq 2$ and any collection $B_1, \ldots, B_k$ of elements of $\mathcal{B}$,

$$Pl\left(\bigcap_{i=1}^{k} B_i\right) \leq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} Pl\left(\bigcup_{i \in I} B_i\right).$$

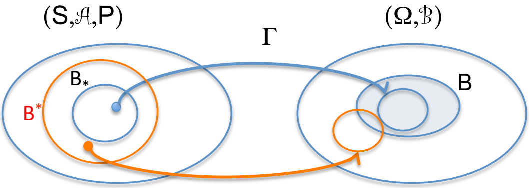
(Pl is completely alternating)

Proposition

Bel is a BF if and only if Pl defined by $Pl(B) = 1 - Bel(\overline{B})$ for all $B \in \mathcal{B}$ is a plausibility function.

Belief function induced by a source

Lower and upper inverses of a multi-valued mapping



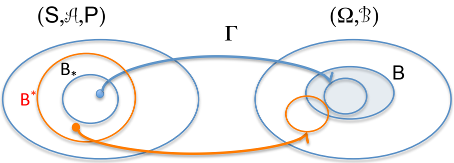
- Let $(S, \mathcal{A}, \mathbb{P})$ be a probability space, $(\Omega, \mathcal{B})$ a measurable space, and Γ a multivalued mapping from S to 2^{Ω}
- Lower and upper inverse: for all $B \in \mathcal{B}$,

$$\Gamma_*(B) = B_* = \{s \in S | \Gamma(s) \neq \emptyset, \Gamma(s) \subseteq B\}$$

$$\Gamma^*(B) = B^* = \{s \in S | \Gamma(s) \cap B \neq \emptyset\}$$

Belief function induced by a source

Lower and upper probabilities

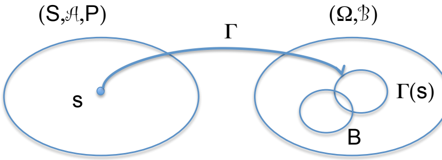


- Γ is strongly measurable wrt $\mathcal{A}$ and $\mathcal{B}$ if, for all $B \in \mathcal{B}$, $B^* \in \mathcal{A}$
- Lower and upper probabilities:

$$\forall B \in \mathcal{B}, \quad \mathbb{P}_*(B) = \frac{\mathbb{P}(B_*)}{\mathbb{P}(\Omega^*)}, \quad \mathbb{P}^*(B) = \frac{\mathbb{P}(B^*)}{\mathbb{P}(\Omega^*)} = 1 - Bel(\overline{B})$$

- $\mathbb{P}_*$ is a BF, and $\mathbb{P}^*$ is the dual plausibility function
- $(S, \mathcal{A}, \mathbb{P}, \Gamma)$ is called a source ($\equiv$ random set) for the BF $Bel = \mathbb{P}_*$

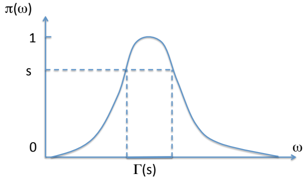
Interpretation



- Typically, Ω is the domain of an unknown quantity ω, and S is a set of interpretations of a given piece of evidence about ω
- If $s \in S$ holds, then the evidence tells us that $\omega \in \Gamma(s)$, and nothing more (imprecision)
- Then
  - $Bel(B)$ is the probability that the evidence implies B
  - $Pl(B)$ is the probability that the evidence is consistent with B

Outline

Consonant random set



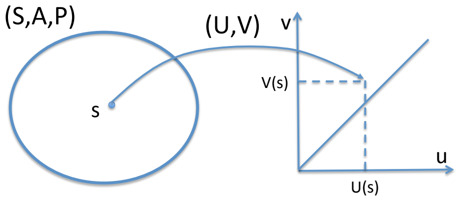
- Let $S = [0,1]$, $\Omega = \mathbb{R}^d$, let π be a mapping from Ω to $S = [0,1]$ s.t. $\sup \pi = 1$, and Γ the mapping from S to 2^Ω defined by

$$\forall s \in [0,1], \quad \Gamma(s) = \{\omega \in \Omega | \pi(\omega) \geq s\}$$

- The source $([0,1], \mathcal{B}([0,1]), \lambda, \Gamma)$ defines a consonant random set, which induces a consonant BF on Ω, with contour function $pl(\omega) = \pi(\omega)$
- The corresponding plausibility function is a possibility measure

$$\forall B \subseteq \mathbb{R}^d, \quad Pl(B) = \sup_{\omega \in B} pl(\omega)$$

Random closed interval



- Let (U, V) be a bi-dimensional random vector from a probability space $(S, \mathcal{A}, \mathbb{P})$ to $\mathbb{R}^2$ such that $\mathbb{P}(\{s \in S | U(s) \leq V(s)\}) = 1$
- The mapping

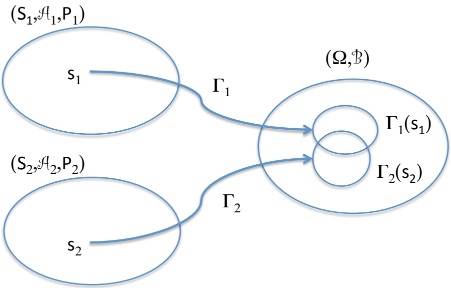
$$\Gamma : s \rightarrow \Gamma(s) = [U(s), V(s)],$$

  is strongly measurable. It defines a random closed interval
- The source $(S, \mathcal{A}, \mathbb{P}, \Gamma)$ defines a BF on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

Outline

Combination of evidence

Dempster's rule of combination



- Let $(S_i, \mathcal{A}_i, \mathbb{P}_i, \Gamma_i)$, $i = 1, 2$ be two sources representing independent items of evidence, inducing BF Bel_1 and Bel_2
- The combined BF $Bel = Bel_1 \oplus Bel_2$ is induced by the source $(S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2, \Gamma_\cap)$ with

$$\Gamma_\cap(s_1, s_2) = \Gamma_1(s_1) \cap \Gamma_2(s_2)$$

Approximate computation

Monte Carlo simulation

Require: Desired number of focal sets N

$i \leftarrow 0$

while $i < N$ **do**

Draw s_1 in S_1 from $\mathbb{P}_1$

Draw s_2 in S_2 from $\mathbb{P}_2$

$\Gamma_\cap(s_1, s_2) \leftarrow \Gamma_1(s_1) \cap \Gamma_2(s_2)$

if $\Gamma_\cap(s_1, s_2) \neq \emptyset$ **then**

$i \leftarrow i + 1$

$B_i \leftarrow \Gamma_\cap(s_1, s_2)$

end if

end while

$\widehat{Bel}(B) \leftarrow \frac{1}{N}\#\{i \in \{1, \ldots, N\} | B_i \subseteq B\}$

$\widehat{Pl}(B) \leftarrow \frac{1}{N}\#\{i \in \{1, \ldots, N\} | B_i \cap B \neq \emptyset\}$

Outline

The estimation problem

- Let $\boldsymbol{y} \in \mathbb{Y}$ denote the observed data and $f_{\boldsymbol{\theta}}(\boldsymbol{y})$ the probability mass or density function describing the data-generating mechanism, where $\boldsymbol{\theta} \in \Theta$ is an unknown parameter
- Having observed $\boldsymbol{y}$, how to quantify the uncertainty about Θ, without specifying a prior probability distribution?
- Likelihood-based solution (Shafer, 1976; Wasserman, 1990; Denœux, 2014)

Likelihood-based belief function

Requirements

Let $Bel_{\boldsymbol{y}}^{\Theta}$ be a belief function representing our knowledge about θ after observing $\boldsymbol{y}$. We impose the following requirements:

1. Likelihood principle: $Bel_{\boldsymbol{y}}^{\Theta}$ should be based only on the likelihood function

$$\theta \rightarrow L_{\boldsymbol{y}}(\theta) = f_{\theta}(\boldsymbol{y})$$

2. Compatibility with Bayesian inference: when a Bayesian prior P_0 is available, combining it with Bel_{z}^{Θ} using Dempster's rule should yield the Bayesian posterior:

$$Bel_{\boldsymbol{y}}^{\Theta} \oplus P_0 = P(\cdot|\boldsymbol{y})$$

3. Principle of minimal commitment: among all the belief functions satisfying the previous two requirements, $Bel_{\boldsymbol{y}}^{\Theta}$ should be the least committed (least informative)

Likelihood-based belief function

Solution (Denœux, 2014)

- $Bel_{\boldsymbol{y}}^{\Theta}$ is the consonant belief function such that

$$pl_{\boldsymbol{y}}(\boldsymbol{\theta}) = \frac{L_{\boldsymbol{y}}(\boldsymbol{\theta})}{L_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}})},$$

  where $\widehat{\boldsymbol{\theta}}$ is a MLE of $\boldsymbol{\theta}$, and it is assumed that $L_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}}) < +\infty$
- Corresponding plausibility function

$$Pl_{\boldsymbol{y}}^{\Theta}(A) = \sup_{\theta \in A} pl_{\boldsymbol{y}}(\boldsymbol{\theta}), \quad \forall A \subseteq \Theta$$

- Source: $([0,1], \mathcal{B}([0,1]), \lambda, \Gamma_{\boldsymbol{y}})$, with

$$\Gamma_{\boldsymbol{y}}(s) = \left\{ \boldsymbol{\theta} \in \Theta | \frac{L_{\boldsymbol{y}}(\boldsymbol{\theta})}{L_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}})} \geq s \right\}$$

Profile likelihood

- Assume that $\theta = (\xi, \nu)$, where ξ is a parameter of interest and ν is a nuisance parameter
- Then, the marginal contour function for ξ is

$$pl_{\mathbf{y}}(\xi) = \sup_{\nu} pl_{\mathbf{y}}(\xi, \nu),$$

  which is the profile relative likelihood function
- The profiling method for eliminating nuisance parameter thus has a natural justification in our approach
- When the quantities $pl_{\mathbf{y}}(\xi)$ cannot be derived analytically, they have to be computed numerically using an iterative optimization algorithm

Relation with likelihood-based inference

- The approach to statistical inference outlined in the previous section is very close to the "likelihoodist" approach advocated by Birnbaum (1962), Barnard (1962), and Edwards (1992), among others
- The main difference resides in the interpretation of the likelihood function as defining a belief function
- This interpretation allows us to quantify the uncertainty in statements of the form $\theta \in H$, where H may contain multiple values. This is in contrast with the classical likelihood approach, in which only the likelihood of single hypotheses is defined
- The belief function interpretation provides an easy and natural way to combine statistical information with other information, such as expert judgements

Outline

Adaptation of flood defense structures

- Commonly, flood defenses in coastal areas are designed to withstand at least 100 years return period events.
- However, due to climate change, they will be subject during their life time to higher loads than the design estimations.
- The main impact is related to the increase of the mean sea level, which affects the frequency and intensity of surges.
- For adaptation purposes, we need to combine
  - statistics of extreme sea levels derived from historical data
  - expert judgement about the future sea level rise (SLR)

Model

- The annual maximum sea level Z at a given location is often assumed to have a Gumbel distribution

$$P(Z \leq z) = \exp\left[-\exp\left(-\frac{z-\mu}{\sigma}\right)\right]$$

  with mode μ and scale parameter σ
- Current design procedures are based on the return level z_T associated to a return period T, defined as the quantile at level $1-1/T$. Here,

$$z_T = \mu - \sigma \log\left[-\log\left(1-\frac{1}{T}\right)\right]$$

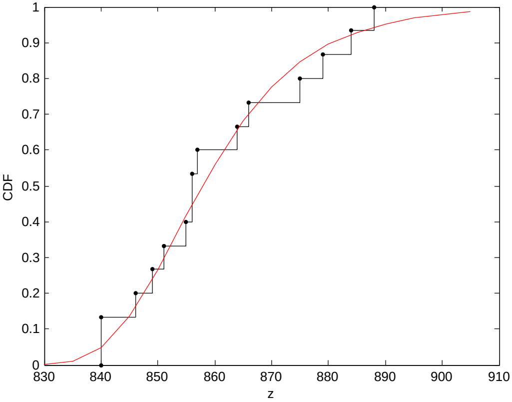
- Because of climate change, it is assumed that the distribution of annual maximum sea level at the end of the century will be shifted to the right, with shift equal to the SLR:

$$z'_T = z_T + SLR$$

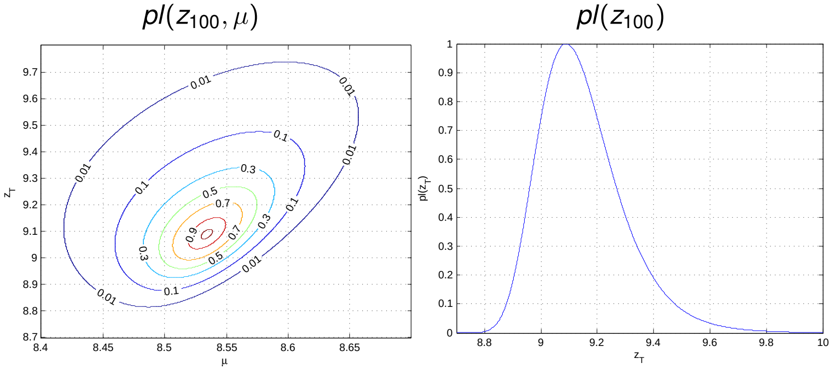
Approach

1. Represent the evidence on z_T by a likelihood-based belief function using past sea level measurements
2. Represent the evidence on SLR by a belief function describing expert opinions
3. Combine these two items of evidence to get a belief function on $z'_T = z_T + SLR$

Sea level data at Le Havre, France (15 years)



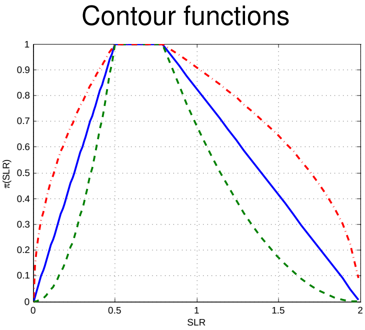
Contour functions

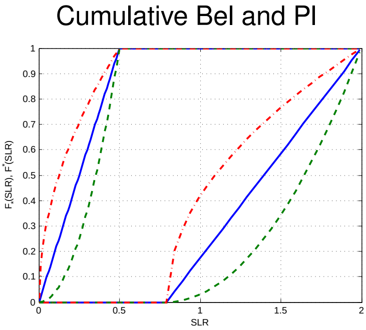


Representation of expert opinions about the SLR

- From a review of the literature (in 2007)
  - The interval $[0.5, 0.79] = [0.18, 0.79] \cap [0.5, 1.4]$ seems to be fully supported by the available evidence
  - Values outside the interval $[0, 2]$ are considered as practically impossible
- Three representations:
  - Consonant random intervals with core $[0.5, 0.79]$, support $[0, 2]$ and different contour functions π;
  - p-boxes with same cumulative belief and plausibility functions as above;
  - Random sets $[U, V]$ with independent U and V and same cumulative belief and plausibility functions as above.

Representation of expert opinions about the SLR





Combination
Principle

- Let $[U_{z_T}, V_{z_T}]$ and $[U_{SLR}, V_{SLR}]$ be the independent random intervals representing evidence on z_T and SLR, respectively.
- The random interval for $z'_T = z_T + SLR$ is

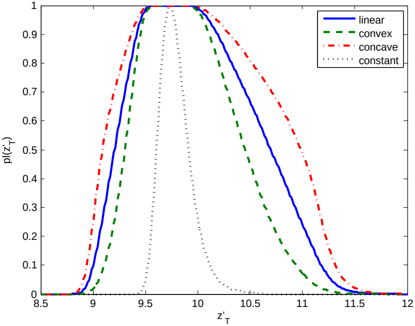
$$[U_{z_T}, V_{z_T}] + [U_{SLR}, V_{SLR}] = [U_{z_T} + U_{SLR}, V_{z_T} + V_{SLR}]$$

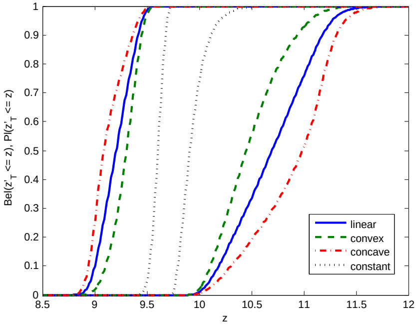
- The corresponding belief and plausibility functions are

$$\begin{aligned} Bel(A) &= P([U_{z_T} + U_{SLR}, V_{z_T} + V_{SLR}] \subseteq A) \\ Pl(A) &= P([U_{z_T} + U_{SLR}, V_{z_T} + V_{SLR}] \cap A \neq \emptyset) \end{aligned}$$

  for all $A \in \mathcal{B}(\mathbb{R})$.
- $Bel(A)$ and $Pl(A)$ can be estimated by Monte Carlo simulation.

Result

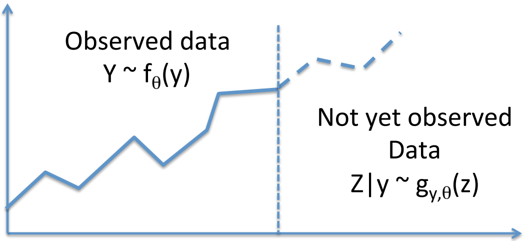

linear
convex
concave
constant
pl(z'_T)
z'_T
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
8.5
9
9.5
10
10.5
11
11.5
12


linear
convex
concave
constant
Bel(z'_T <= z), Pl(z'_T <= z)
z
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
8.5
9
9.5
10
10.5
11
11.5
12

Outline

1. Belief functions
   - Basic definitions
   - Practical models
   - Dempster's rule
2. Estimation
   - Likelihood-based belief function
   - Exemple: sea level rise
3. Prediction
   - Predictive belief function
   - Example: linear regression

The prediction problem



Given some knowledge about θ obtained by observing $\boldsymbol{y}$, what can we say about some not yet observed data $\boldsymbol{z} \in \mathbb{Z}$, whose conditional distribution $g_{\boldsymbol{y},\theta}(\boldsymbol{z})$ given $\boldsymbol{y}$ depends on θ?

Simple example

- $(y_1, \ldots, y_n, z)$ iid from $\mathcal{N}(\theta, 1)$
- Problem: predict z after observing $\boldsymbol{y} = (y_1, \ldots, y_n)$
- We can write

$$z = \theta + \Phi^{-1}(w) = \varphi(\theta, w) \text{ with } w \sim \mathcal{U}([0, 1])$$

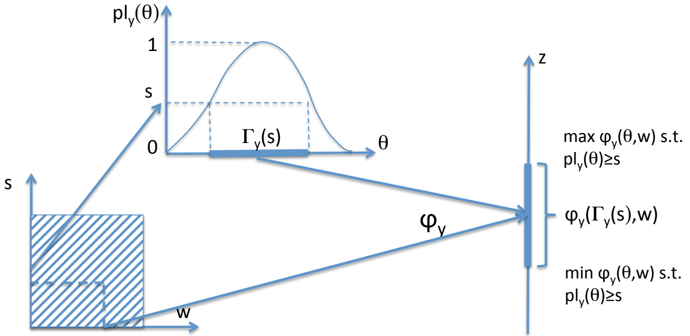
- Corresponding multi-valued mapping

$$w \rightarrow \Gamma(w) = \{(\theta, z) \in \Theta \times \mathbb{Z} | z = \varphi(\theta, w)\}$$

- The source $([0, 1]), \mathcal{B}([0, 1]), \lambda, \Gamma)$ induces a joint belief function $Bel^{\Theta \times \mathbb{Z}}$
- Predictive belief function on $\mathbb{Z}$

$$Bel_y^{\mathbb{Z}} = (Bel^{\Theta \times \mathbb{Z}} \oplus Bel_{\boldsymbol{y}}^{\Theta})^{\downarrow \mathbb{Z}}$$

Predictive belief function



$Bel_{\boldsymbol{y}}^{\mathbb{Z}}$ induced by the source $([0,1]^2, \mathcal{B}([0,1]^2), \lambda^2, \Gamma'_{\boldsymbol{y}})$ where $\Gamma'_{\boldsymbol{y}}$ is the multi-valued mapping $(s, w) \rightarrow \varphi_{\boldsymbol{y}}(\Gamma_{\boldsymbol{y}}(s), w)$

Analytical expression

- In this example, the predictive belief function corresponds to the random closed interval

$$\varphi(\Gamma_y(s), w) = \left[\overline{y} - \sqrt{\frac{-2\ln s}{n}} + \Phi^{-1}(w), \overline{y} + \sqrt{\frac{-2\ln s}{n}} + \Phi^{-1}(w)\right]$$

- As $n \to +\infty$, both bounds converge in distribution to a rv with the same distribution as z

General approach

- Principle
  1. Using the sampling model of $\boldsymbol{z}$ given $\boldsymbol{y}$, construct a joint belief function $Bel_{\boldsymbol{y}}^{\mathbb{Z}\times\Theta}$ on $\mathbb{Z}\times\Theta$
  2. Combine $Bel_{\boldsymbol{y}}^{\mathbb{Z}\times\Theta}$ with the likelihood-based belief function $Bel_{\boldsymbol{y}}^{\Theta}$
  3. Marginalize on $\mathbb{Z}$ to a obtain a predictive belief function $Bel_{\boldsymbol{y}}^{\mathbb{Z}}$
- $Bel_{\boldsymbol{y}}^{\mathbb{Z}}$ can be approximated by a combination of Monte Carlo simulation and constrained optimization

O. Kanjanatarakul, S. Sriboonchitta and T. Denœux
Statistical estimation and prediction using belief functions: principles and application to some econometric models
Submitted, 2015

Remarks

- If θ is fixed to to its true value θ_0, then $Bel_{\boldsymbol{y}}^{\mathbb{Z}}$ equals the true conditional probability distribution of z given $\boldsymbol{y}$
- If the likelihood-based belief function $Bel_{\boldsymbol{y}}^{\Theta}$ is combined by a Bayesian prior P_0, then
  - $Bel_{\boldsymbol{y}}^{\Theta} \oplus P_0$ is the posterior on θ
  - $Bel_{\boldsymbol{y}}^{\mathbb{Z}}$ become the Bayesian posterior predictive distribution of z given $\boldsymbol{y}$

Outline

Model

We consider the following standard regression model

$$\boldsymbol{y} = X\beta + \boldsymbol{\epsilon}$$

where

- $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is the vector of n observations of the dependent variable
- X is the fixed design matrix of size $n \times (p + 1)$
- $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)' \sim \mathcal{N}(\mathbf{0}, I_n)$ is the vector of errors
- The vector of coefficients is $\boldsymbol{\theta} = (\beta', \sigma)'$.

Likelihood-based belief function

- The likelihood function for this model is

$$L_{\boldsymbol{y}}(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\beta)'(\boldsymbol{y} - X\beta)\right].$$

- The contour function can thus be readily calculated as

$$pl_{\boldsymbol{y}}(\boldsymbol{\theta}) = \frac{L_{\boldsymbol{y}}(\boldsymbol{\theta})}{L_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}})}$$

with $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}', \widehat{\sigma})'$, where
  - $\widehat{\beta} = (X'X)^{-1}X'\boldsymbol{y}$ is the ordinary least squares estimate of β
  - $\widehat{\sigma}$ is the standard deviation of residuals

Plausibility of linear hypotheses

- Assertions (hypotheses) H of the form $A\beta = \boldsymbol{q}$, where A is a $r \times (p+1)$ constant matrix and $\boldsymbol{q}$ is a constant vector of length r, for some $r \leq p+1$
- Special cases: $\{\beta_j = 0\}$, $\{\beta_j = 0, \forall j \in \{1, \ldots, p\}\}$, or $\{\beta_j = \beta_k\}$, etc.
- The plausibility of H is

$$Pl_{\boldsymbol{y}}^{\Theta}(H) = \sup_{A\beta=\boldsymbol{q}} pl_{\boldsymbol{y}}(\boldsymbol{\theta}) = \frac{L_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}}_*)}{L_{\boldsymbol{y}}(\widehat{\boldsymbol{\theta}})}$$

where $\widehat{\boldsymbol{\theta}}_* = (\widehat{\boldsymbol{\beta}}_*', \widehat{\sigma}_*)'$ (restricted LS estimates) with

$$\widehat{\boldsymbol{\beta}}_* = \widehat{\boldsymbol{\beta}} - (X'X)^{-1}A'[A(X'X)^{-1}A']^{-1}(A\widehat{\boldsymbol{\beta}} - \boldsymbol{q})$$

$$\widehat{\sigma}_* = \sqrt{(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}_*)'(\boldsymbol{y} - X\widehat{\boldsymbol{\beta}}_*)/n}$$

Linear model: prediction

- Let z be a not-yet observed value of the dependent variable for a vector $\boldsymbol{x}_0$ of covariates:

$$z = \boldsymbol{x}_0'\beta + \epsilon_0,$$

with $\epsilon_0 \sim \mathcal{N}(0, \sigma^2)$

- We can write, equivalently,

$$z = \boldsymbol{x}_0'\beta + \sigma\Phi^{-1}(w) = \varphi_{\boldsymbol{x}_0,\boldsymbol{y}}(\boldsymbol{\theta}, w),$$

where w has a standard uniform distribution

- The predictive belief function on z can then be approximated using Monte Carlo simulation

Linear model: prediction

- Let z be a not-yet observed value of the dependent variable for a vector $\boldsymbol{x}_0$ of covariates:

$$z = \boldsymbol{x}_0' \boldsymbol{\beta} + \epsilon_0,$$

  with $\epsilon_0 \sim \mathcal{N}(0, \sigma^2)$
- We can write, equivalently,

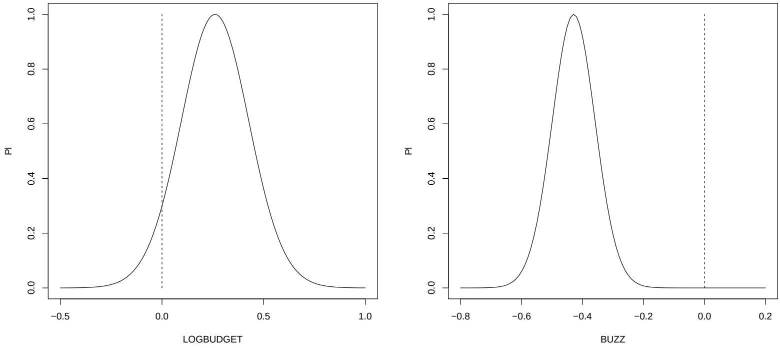
$$z = \boldsymbol{x}_0' \boldsymbol{\beta} + \sigma \Phi^{-1}(w) = \varphi_{\boldsymbol{x}_0, \boldsymbol{y}}(\boldsymbol{\theta}, w),$$

  where w has a standard uniform distribution
- The predictive belief function on z can then be approximated using Monte Carlo simulation

Example: movie Box office data

- Dataset about 62 movies released in 2009 (from Greene, 2012)
- Dependent variable: logarithm of Box Office receipts
- 11 covariates:
  - 3 dummy variables (G, PG, PG13) to encode the MPAA (Motion Picture Association of America) rating, logarithm of budget (LOGBUDGET), star power (STARPOWR),
  - a dummy variable to indicate if the movie is a sequel (SEQUEL),
  - four dummy variables to describe the genre (ACTION, COMEDY, ANIMATED, HORROR)
  - one variable to represent internet buzz (BUZZ)

Some marginal contour functions

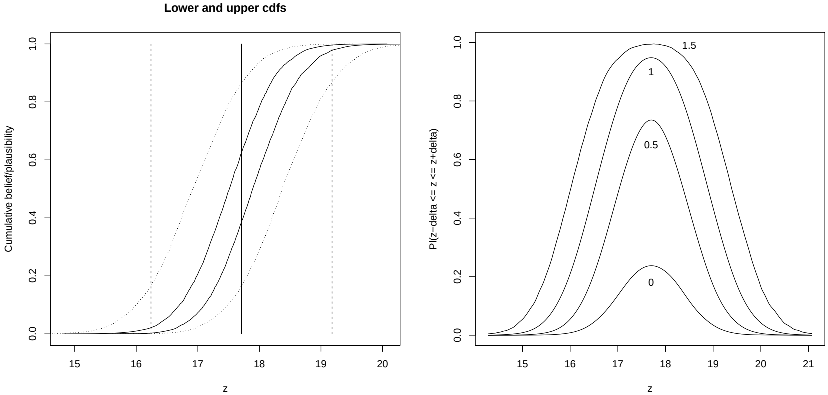


Regression coefficients

	Estimate	Std. Error	t-value	p-value	$Pl(\beta_j = 0)$
(Intercept)	15.400	0.643	23.960	$<$ 2e-16	1.0e-34
G	0.384	0.553	0.695	0.49	0.74
PG	0.534	0.300	1.780	0.081	0.15
PG13	0.215	0.219	0.983	0.33	0.55
LOGBUDGET	0.261	0.185	1.408	0.17	0.30
STARPOWR	4.32e-3	0.0128	0.337	0.74	0.93
SEQUEL	0.275	0.273	1.007	0.32	0.54
ACTION	-0.869	0.293	-2.964	4.7e-3	6.6e-3
COMEDY	-0.0162	0.256	-0.063	0.95	0.99
ANIMATED	-0.833	0.430	-1.937	0.058	0.11
HORROR	0.375	0.371	1.009	0.32	0.54
BUZZ	0.429	0.0784	5.473	1.4e-06	4.8e-07

Movie example

BO success of an action sequel film rated PG13 by MPAA, with LOGBUDGET=5.30, STARPOWER=23.62 and BUZZ= 2.81?



Ex ante forecasting
Problem and classical approach

- Consider the situation where some explanatory variables are unknown at the time of the forecast and have to be estimated or predicted
- Classical approach: assume that $\boldsymbol{x}_0$ has been estimated with some variance, which has to be taken into account in the calculation of the forecast variance
- According to Green (Econometric Analysis, 7th edition, 2012)
  - "*This vastly complicates the computation. Many authors view it as simply intractable*"
  - "*analytical results for the correct forecast variance remain to be derived except for simple special cases*"

Ex ante forecasting

Belief function approach

- In contrast, this problem can be handled very naturally in our approach by modeling partial knowledge of $\boldsymbol{x}_0$ by a belief function $Bel^{\mathbb{X}}$ in the sample space $\mathbb{X}$ of $\boldsymbol{x}_0$
- We then have

$$Bel_{\boldsymbol{y}}^{\mathbb{Z}} = \left(Bel_{\boldsymbol{y}}^{\Theta} \oplus Bel_{\boldsymbol{y}}^{\mathbb{Z}\times\Theta} \oplus Bel^{\mathbb{X}}\right)^{\downarrow\mathbb{Z}}$$

- Assume that the belief function $Bel^{\mathbb{X}}$ is induced by a source $(\Omega, \mathcal{A}, \mathbb{P}^{\Omega}, \Lambda)$, where Λ is a multi-valued mapping from Ω to $2^{\mathbb{X}}$
- The predictive belief function $Bel_{\boldsymbol{y}}^{\mathbb{Z}}$ is then induced by the multi-valued mapping

$$(\omega, s, w) \rightarrow \varphi_{y}(\Lambda(\omega), \Gamma_{\boldsymbol{y}}(s), w)$$

- $Bel_{\boldsymbol{y}}^{\mathbb{Z}}$ can be approximated by Monte Carlo simulation

Monte Carlo algorithm

Require: Desired number of focal sets N

for $i = 1$ **to** N **do**

Draw (s_i, w_i) uniformly in $[0, 1]^2$

Draw ω from $\mathbb{P}^\Omega$

Search for $z_{*i} = \min_{\boldsymbol{\theta}} \varphi_{\boldsymbol{y}}(\boldsymbol{x}_0, \boldsymbol{\theta}, w_i)$ such that $pl_{\boldsymbol{y}}(\boldsymbol{\theta}) \geq s_i$ and $\boldsymbol{x}_0 \in \Lambda(\omega)$.

Search for $z_i^* = \max_{\boldsymbol{\theta}} \varphi_{\boldsymbol{y}}(\boldsymbol{x}_0, \boldsymbol{\theta}, w_i)$ such that $pl_{\boldsymbol{y}}(\boldsymbol{\theta}) \geq s_i$ and $\boldsymbol{x}_0 \in \Lambda(\omega)$.

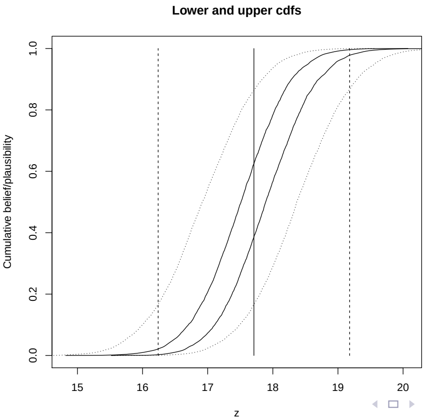
$B_i \leftarrow [z_{*i}, z_i^*]$

end for

Movie example

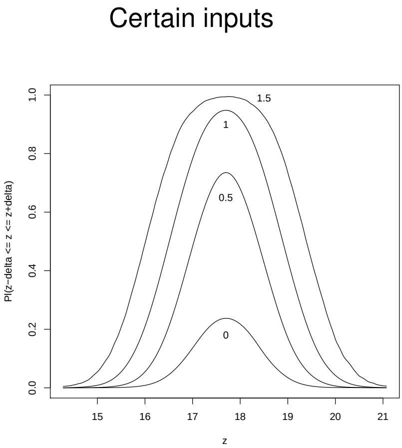
Lower and upper cdfs

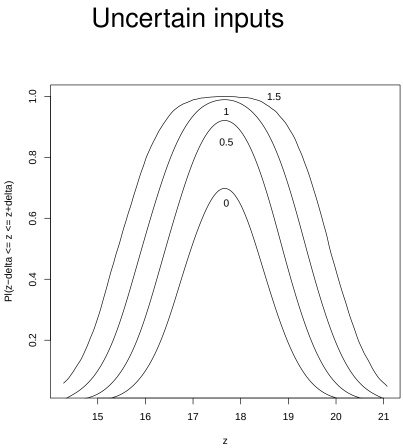
BO success of an action sequel film rated PG13 by MPAA, with LOGBUDGET=5.30, STARPOWER=23.62 and BUZZ= (0,2.81,5) (triangular possibility distribution)?



Movie example

Pl-plots





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 proposed method is conceptually simple and computationally tractable
- The belief function formalism makes it possible to combine information from several sources (such as expert opinions and statistical data)
- The Bayesian predictive probability distribution is recovered when a prior on θ is available

Some open questions

- Is there a way to compare this approach with other prediction methods (such as prediction intervals or Bayesian posterior distributions), other than by discussing the underlying principles?
- How to account for the partial inadequacy of the parametric generative model? Non-parametric approach?
- Under which conditions can we guarantee the consistency of the method (convergence, in some sense, of the predictive belief function to the true distribution as the sample size tends to infinity)

Papers and Matlab software available at:

`https://www.hds.utc.fr/~tdenoeux`

THANK YOU!