

# Belief Functions

## Statistical Inference

Thierry Denœux

Summer 2023



# Estimation vs. prediction

- Consider an urn with an unknown proportion  $\theta$  of black balls.
- Assume that we have drawn  $n$  balls with replacement from the urn,  $y$  of which were black.
- Problems:
  - 1 What can we say about  $\theta$ ? (**estimation**)
  - 2 What can we say about the color  $Z$  of the next ball to be drawn from the urn? (**prediction**)
- Classical approaches
  - **Frequentist**: gives an answer that is correct most the time (over infinitely many replications of the random experiment)
  - **Bayesian**: assumes prior knowledge on  $\theta$  and computes posterior probability distributions  $f(\theta|y)$  and  $P(\text{black}|y)$



# Criticism of the frequentist approach

- The frequentist approach makes a statement that is **correct, say, for 95% of the samples**
- The confidence level is often interpreted as a measure of confidence in the statement for a particular sample
- However, this interpretation poses some logical problems



## Example

- Suppose  $X_1$  and  $X_2$  are iid with probability mass function

$$\mathbb{P}_\theta(X_i = \theta - 1) = \mathbb{P}_\theta(X_i = \theta + 1) = \frac{1}{2}, \quad i = 1, 2, \quad (1)$$

where  $\theta \in \mathbb{R}$  is an unknown parameter.

- Consider the following confidence set for  $\theta$ ,

$$C(X_1, X_2) = \begin{cases} \frac{1}{2}(X_1 + X_2) & \text{if } X_1 \neq X_2 \\ X_1 - 1 & \text{otherwise.} \end{cases} \quad (2)$$

- The corresponding confidence level is  $P_\theta(\theta \in C(X_1, X_2)) = 0.75$
- Now, let  $(x_1, x_2)$  be a given realization of the random sample  $(X_1, X_2)$ .
  - If  $x_1 \neq x_2$ , we know for sure that  $\theta = (x_1 + x_2)/2$
  - If  $x_1 = x_2$ , we know for sure that  $\theta$  is either  $x_1 - 1$  or  $x_1 + 1$ , but we have no reason to favor any of these two hypotheses in particular.
- This problem is known as the **problem of relevant subsets** (there are recognizable situations in which the coverage probability is different from the stated one)



# The relevant subset problem

- This phenomenon happens in the usual problem of interval estimation of the mean of a normal sample: “wide” CIs in some sense have larger coverage probability than the stated confidence level, and vice versa for “short” intervals.
- Specifically, let  $X_1, \dots, X_n$  be an iid sample from  $\mathcal{N}(\mu, \sigma^2)$  with both parameters unknown, and

$$C = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \frac{s}{|\bar{x}|} > k \right\}$$

for some  $k$ , where  $\bar{x}$  is the sample mean and  $s$  is the sample standard deviation.

- The standard CI for  $\mu$  is  $\bar{x} \pm t_{n-1; 1-\alpha/2} s / \sqrt{n}$
- It can be shown that, for some  $\epsilon > 0$ ,

$$P(\mu \in CI | C) > (1 - \alpha) + \epsilon$$

for all  $\mu$  and  $\sigma$

- “The existence of certain relevant subsets is an embarrassment to confidence theory” (Lehmann, 1986)



# Criticism of the Bayesian approach

- In the Bayesian approach,  $y$ ,  $z$  and  $\theta$  are seen as **random variables**
- **Estimation**: compute the posterior pdf of  $\theta$  given  $y$

$$f(\theta|y) \propto p(y|\theta)f(\theta)$$

where  $f(\theta)$  is the prior pdf on  $\theta$

- **Prediction**: compute the predictive posterior distribution

$$p(z|y) = \int p(z|\theta)f(\theta|y)d\theta$$

- **We need the prior  $f(\theta)$ !**
- The uniform prior is dependent on the parameterization; consequently, it is not truly noninformative.
- Another solution: Jeffreys prior



# Jeffreys prior

- The **Jeffreys** prior is defined objectively as being proportional to the square root of the determinant of the Fisher information

$$f(\theta) \propto \sqrt{\det I(\theta)}$$

where the component  $(i, j)$  of the information matrix  $I(\theta)_{ij}$  is

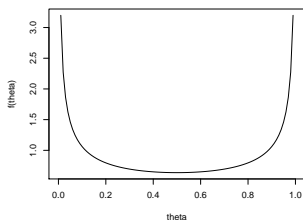
$$I(\theta)_{ij} = \mathbb{E}_{\theta} \left[ \frac{\partial \log f_{\theta}(x)}{\partial \theta_i} \frac{\partial \log f_{\theta}(x)}{\partial \theta_j} \right]$$

- The motivation for this definition is that the Jeffreys prior is invariant under reparameterization: if  $\varphi$  is a one-to-one transformation and  $\nu = \varphi(\theta)$ , then the Jeffreys prior on  $\nu$  is proportional to  $\sqrt{\det I(\nu)}$ .



# Problems with Jeffrey's prior

- However, there are still some issues with this approach:
  - First, the Jeffreys prior is sometimes improper.
  - Secondly, and maybe more importantly, the Jeffreys prior can hardly be considered to be truly noninformative.
- For instance, consider an iid sample  $X_1, \dots, X_n$  from a Bernoulli distribution  $B(\theta)$ . The Jeffreys prior on  $\theta$  is the beta distribution  $B(0.5, 0.5)$  whose pdf is displayed below. We can see that extreme values of  $\theta$  are considered a priori more probable than central values, which does represent non vacuous knowledge about  $\theta$ .





# Main ideas

- None of the classical approaches to statistical inference (frequentist and Bayesian) is fully satisfactory, from a conceptual point of view. This is why research on the foundations of statistical inference is going on.
- In this chapter, we describe an **approach based on belief functions**.
- The new approach boils down to Bayesian inference when a probabilistic prior is available, but **it does not require the user to provide such a prior**.
- We will apply this approach to different econometric models.
- Before applying belief functions to statistical inference, we need to define belief functions on infinite spaces.



# Outline

## 1 Belief functions on infinite spaces

- Definition
- Practical models
- Combination

## 2 Estimation

- Likelihood-based belief function
- Examples
- Consistency

## 3 Prediction

- Predictive belief function
- Examples



# Outline

## 1 Belief functions on infinite spaces

- Definition
- Practical models
- Combination

## 2 Estimation

- Likelihood-based belief function
- Examples
- Consistency

## 3 Prediction

- Predictive belief function
- Examples



# Belief function: general definition

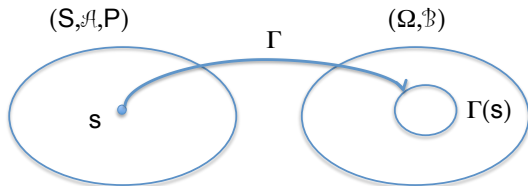
- Let  $\Omega$  be a set (finite or not) and  $\mathcal{B}$  be an algebra of subsets of  $\Omega$
- A **belief function (BF)** on  $\mathcal{B}$  is a mapping  $Bel : \mathcal{B} \rightarrow [0, 1]$  verifying  $Bel(\emptyset) = 0$ ,  $Bel(\Omega) = 1$  and the complete monotonicity property: for any  $k \geq 2$  and any collection  $B_1, \dots, B_k$  of elements of  $\mathcal{B}$ ,

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

- A function  $Pl : \mathcal{B} \rightarrow [0, 1]$  is a **plausibility function** iff  $Bel : \mathcal{B} \rightarrow [0, 1]$  defined by  $Bel(B) = 1 - Pl(\overline{B})$  is a belief function



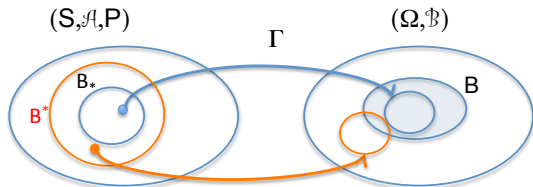
# Random set



- Let  $(S, \mathcal{A}, \mathbb{P})$  be a probability space, and  $(\Omega, \mathcal{B})$  a measurable space.
- Let  $\Gamma : S \rightarrow 2^\Omega$  be a **multivalued mapping**.
- The images  $\Gamma(s)$  are called the **focal sets** of  $\Gamma$ .
- Under measurability conditions (see next slide), this framework defines a **random set**.



# Strong measurability



- Lower and upper inverses: for all  $B \in \mathcal{B}$ ,

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

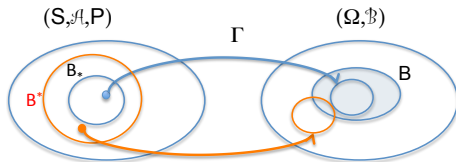
$$\Gamma^*(B) = B^* = \{s \in S : \Gamma(s) \cap B \neq \emptyset\}$$

- $\Gamma$  is **strongly measurable** wrt  $\mathcal{A}$  and  $\mathcal{B}$  if, for all  $B \in \mathcal{B}$ ,  $B^* \in \mathcal{A}$
- $(\forall B \in \mathcal{B}, B^* \in \mathcal{A}) \Leftrightarrow (\forall B \in \mathcal{B}, B_* \in \mathcal{A})$
- If  $\Gamma$  is strongly measurable, the tuple  $(S, \mathcal{A}, \mathbb{P}, \Omega, \mathcal{B}, \Gamma)$  is called a **random set**



# Belief function induced by a random set

## Lower and upper probabilities



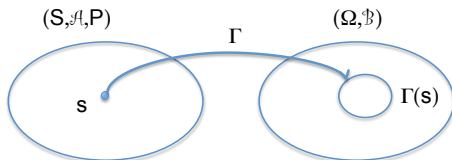
- Lower and upper probabilities:

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

- $\mathbb{P}_*$  is a BF, and  $\mathbb{P}^*$  is the dual plausibility function
- Conversely, for any belief function, there is a random set that induces it (Shafer's thesis, 1973)



# Interpretation



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





# Outline

## 1 Belief functions on infinite spaces

- Definition
- **Practical models**
- Combination

## 2 Estimation

- Likelihood-based belief function
- Examples
- Consistency

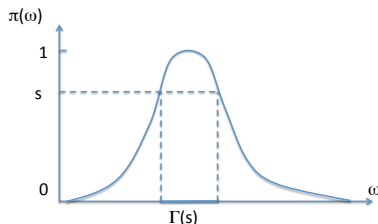
## 3 Prediction

- Predictive belief function
- Examples



# Consonant belief function

Random set



- Let  $\pi$  be a mapping from  $\Omega = \mathbb{R}^p$  to  $S = [0, 1]$  s.t.  $\sup \pi = 1$
- Let  $\Gamma$  be the multi-valued mapping from  $S$  to  $2^\Omega$  defined by

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

- Let  $\mathcal{B}([0, 1])$  be the Borel  $\sigma$ -field on  $[0, 1]$ , and  $P$  the uniform probability measure on  $[0, 1]$
- We consider the random set  $([0, 1], \mathcal{B}([0, 1]), P, \mathbb{R}, \mathcal{B}(\mathbb{R}), \Gamma)$ .



# Consonant belief function

- Let  $Bel$  and  $Pl$  be the belief and plausibility functions induced by the previous random set.
- The focal sets  $\Gamma(s)$  are nested, i.e., for any  $s$  and  $s'$ ,

$$s \geq s' \Rightarrow \Gamma(s) \subseteq \Gamma(s')$$

The belief function is said to be **consonant**.

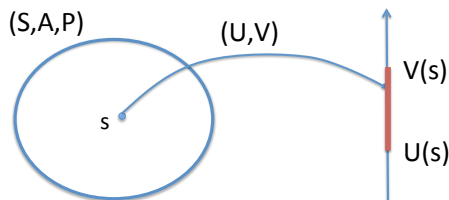
- The corresponding contour function  $pl$  is equal to  $\pi$ .
- The corresponding plausibility function is a **possibility measure**: for any  $B \subseteq \Omega$ ,

$$Pl(B) = \sup_{\omega \in B} pl(\omega)$$

$$Bel(B) = \inf_{\omega \notin B} (1 - pl(\omega))$$



# Random closed interval



- Let  $(U, V)$  be a two-dimensional random vector from a probability space  $(S, \mathcal{A}, \mathbb{P})$  to  $\mathbb{R}^2$  such that  $U \leq V$  a.s.
- Multi-valued mapping:

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

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



# Random closed interval

## Properties

- Lower/upper cdfs:

$$F_*(x) = \text{Bel}((-\infty, x]) = \mathbb{P}([U, V] \subseteq (-\infty, x]) = \mathbb{P}(V \leq x) = F_V(x)$$

$$F^*(x) = \text{Pl}((-\infty, x]) = \mathbb{P}([U, V] \cap (-\infty, x] \neq \emptyset) = \mathbb{P}(U \leq x) = F_U(x)$$

- Lower/upper expectation:

$$\mathbb{E}_*(\Gamma) = \int x dF^*(x) = \mathbb{E}(U)$$

$$\mathbb{E}^*(\Gamma) = \int x dF_*(x) = \mathbb{E}(V)$$

- Lower/upper quantiles

$$q_*(\alpha) = F_U^{-1}(\alpha),$$

$$q^*(\alpha) = F_V^{-1}(\alpha).$$



# Outline

## 1 Belief functions on infinite spaces

- Definition
- Practical models
- **Combination**

## 2 Estimation

- Likelihood-based belief function
- Examples
- Consistency

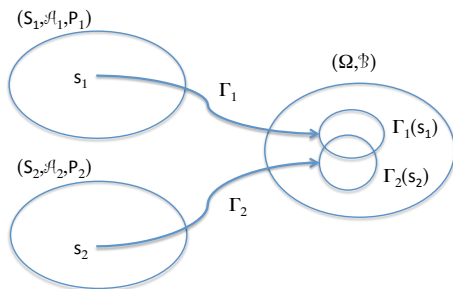
## 3 Prediction

- Predictive belief function
- Examples



# Dempster's rule

## Definition



- Let  $(S_i, \mathcal{A}_i, \mathbb{P}_i, \Omega, \mathcal{B}, \Gamma_i)$ ,  $i = 1, 2$  be two random sets 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 random set  $(S_1 \times S_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2, \Omega, \mathcal{B}, \Gamma_\cap)$  with

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



# Dempster's rule

## Definition

- For each  $B \in \mathcal{B}$ ,  $Bel(B)$  is the conditional probability that  $\Gamma_{\cap}(s) \subseteq B$ , given that  $\Gamma_{\cap}(s) \neq \emptyset$ :

$$Bel(B) = \frac{\mathbb{P}(\{(s_1, s_2) \in S_1 \times S_2 : \Gamma_{\cap}(s_1, s_2) \neq \emptyset, \Gamma_{\cap}(s_1, s_2) \subseteq B\})}{\mathbb{P}(\{(s_1, s_2) \in S_1 \times S_2 : \Gamma_{\cap}(s_1, s_2) \neq \emptyset\})}$$

- It is well defined iff the denominator is non null
- As in the finite case, the **degree of conflict** between the belief functions can be defined as one minus the denominator in the above equation.





# 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**

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

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



# Outline

## 1 Belief functions on infinite spaces

- Definition
- Practical models
- Combination

## 2 Estimation

- Likelihood-based belief function
- Examples
- Consistency

## 3 Prediction

- Predictive belief function
- Examples



# Parameter estimation

- Let  $\mathbf{y} \in \mathbb{Y}$  denote the observed data and  $f_{\theta}(\mathbf{y})$  the probability mass or density function describing the **data-generating mechanism**, where  $\theta \in \Theta$  is an unknown parameter
- Having observed  $\mathbf{y}$ , how to **quantify the uncertainty about  $\Theta$** , without specifying a prior probability distribution?
- Different approaches have been proposed by Dempster (1968), Shafer (1976) and more recently, Martin and Liu (2016)
- Here, I will emphasize Shafer's **Likelihood-based solution** (Shafer, 1976; Wasserman, 1990; Denœux, 2014), which is (much) simpler to implement, and connects nicely with the “likelihoodist” approach to statistical inference.



# Outline

- 1 Belief functions on infinite spaces
  - Definition
  - Practical models
  - Combination
- 2 Estimation
  - **Likelihood-based belief function**
  - Examples
  - Consistency
- 3 Prediction
  - Predictive belief function
  - Examples



# Likelihood-based belief function

## Requirements

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

- 1 **Likelihood principle**:  $Bel_{\mathbf{y}}^{\ominus}$  should be based only on the likelihood function

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

- 2 **Compatibility with Bayesian inference**: when a Bayesian prior  $P_0$  is available, combining it with  $Bel_{\mathbf{y}}^{\ominus}$  using Dempster's rule should yield the Bayesian posterior:

$$Bel_{\mathbf{y}}^{\ominus} \oplus P_0 = P(\cdot | \mathbf{y})$$

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



# Likelihood-based belief function

Solution (Dencœux, 2014)

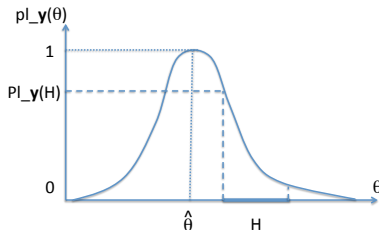
- $Bel_y^\Theta$  is the **consonant belief function** induced by the relative likelihood function

$$pl_y(\theta) = \frac{L_y(\theta)}{L_y(\hat{\theta})}$$

where  $\hat{\theta}$  is a MLE of  $\theta$ , and it is assumed that  $L_y(\hat{\theta}) < +\infty$

- Corresponding **plausibility function**

$$Pl_y^\Theta(H) = \sup_{\theta \in H} pl_y(\theta), \quad \forall H \subseteq \Theta$$

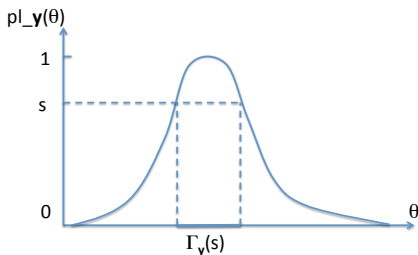


## Corresponding random set

- $Bel_y^\ominus$  can be seen as being induced by the following random set:

$$\Gamma_y(s) = \left\{ \theta \in \Theta \mid \frac{L_y(\theta)}{L_y(\hat{\theta})} \geq s \right\}$$

with  $s$  uniformly distributed in  $[0, 1]$ .



- If  $\Theta \subseteq \mathbb{R}$  and if  $L_y(\theta)$  is unimodal and upper-semicontinuous, then  $Bel_y^\ominus$  is induced by a **random closed interval**.

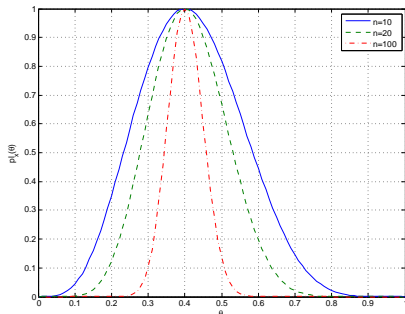


# Binomial example

In the urn model,  $Y \sim \mathcal{B}(n, \theta)$  and

$$pl_y(\theta) = \frac{\theta^y (1 - \theta)^{n-y}}{\widehat{\theta}^y (1 - \widehat{\theta})^{n-y}} = \left( \frac{\theta}{\widehat{\theta}} \right)^{n\widehat{\theta}} \left( \frac{1 - \theta}{1 - \widehat{\theta}} \right)^{n(1-\widehat{\theta})}$$

for all  $\theta \in \Theta = [0, 1]$ , where  $\widehat{\theta} = y/n$  is the MLE of  $\theta$



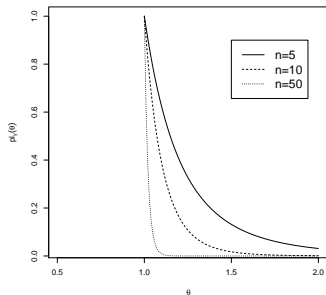


# Uniform example

- Let  $\mathbf{y} = (y_1, \dots, y_n)$  be a realization from an iid random sample from  $\mathcal{U}([0, \theta])$
- The likelihood function is

$$L_{\mathbf{y}}(\theta) = \theta^{-n} \mathbb{1}_{[y_{(n)}, +\infty)}(\theta)$$

- The likelihood-based BF is induced by the random closed interval  $[y_{(n)}, y_{(n)}S^{-1/n}]$ , with  $S \sim \mathcal{U}([0, 1])$



# Profile likelihood

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

$$pl_y(\xi) = Pl(\{\xi\} \times \Omega_\nu) = \sup_{\nu \in \Omega_\nu} pl_y(\xi, \nu),$$

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  $pl_y(\xi)$  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 here 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



# Relation with the likelihood-ratio test statistics

- We can also notice that  $Pl_{\mathbf{y}}^{\ominus}(H)$  is identical to the likelihood ratio statistic for  $H$
- From Wilk's theorem, we have asymptotically (under regularity conditions), when  $H$  holds,

$$-2 \ln Pl_{\mathbf{y}}(H) \sim \chi_r^2$$

where  $r$  is the number of restrictions imposed by  $H$

- Consequently,
  - rejecting hypothesis  $H$  if its plausibility is smaller than  $\exp(-\chi_{r;1-\alpha}^2/2)$  is a testing procedure with significance level approximately equal to  $\alpha$
  - The sets  $\Gamma(\exp(-\chi_{r;1-\alpha}^2/2))$  are approximate  $1 - \alpha$  confidence regions
- However, these properties are coincidental, as the approach outlined here is not based on frequentist inference



# Combination with a Bayesian prior

- The likelihood-based method described here does not require any prior knowledge of  $\theta$ .
- However, by construction, this approach boils down to Bayesian inference if a prior probability  $g(\theta)$  is provided and combined with  $Bel_y^\Theta$  by Dempster's rule.
- As it will usually not be possible to compute the analytical expression of the resulting posterior distribution, it can be approximated by Monte Carlo simulation. (see next slide)
- We can see that this is just the **rejection sampling** algorithm with the prior  $g(\theta)$  as proposal distribution.
- The rejection sampling algorithm can thus be seen, in this case, as a Monte Carlo approximation to Dempster's rule of combination.



## Combination with a Bayesian prior (continued)

Monte Carlo algorithm for combining the likelihood-based belief function with a Bayesian prior by Dempster's rule

**Require:** Desired number of focal sets  $N$

$i \leftarrow 0$

**while**  $i < N$  **do**

Draw  $s$  in  $[0, 1]$  from the uniform probability measure  $\lambda$  on  $[0, 1]$

Draw  $\theta$  from the prior probability distribution  $g(\theta)$

**if**  $p|_y(\theta) \geq s$  **then**

$i \leftarrow i + 1$

$\theta_i \leftarrow \theta$

**end if**

**end while**



# Outline

- 1 Belief functions on infinite spaces
  - Definition
  - Practical models
  - Combination
- 2 Estimation
  - Likelihood-based belief function
  - **Examples**
  - Consistency
- 3 Prediction
  - Predictive belief function
  - Examples



# Behrens-Fisher problem

- Let the observed data  $\mathbf{y}$  be composed of two independent normal samples  $\mathbf{y}_1 = (y_{11}, \dots, y_{1n_1})$  and  $\mathbf{y}_2 = (y_{21}, \dots, y_{2n_2})$  from  $\mathcal{N}(\mu_1, \sigma_1^2)$  and  $\mathcal{N}(\mu_2, \sigma_2^2)$ , respectively.
- We wish to compare the means  $\mu_1$  and  $\mu_2$ .
- Using the frequentist approach, this is done by computing a p-value for the hypothesis  $H_0 : \mu_1 = \mu_2$  of equality of means, or a confidence interval on  $\mu_1 - \mu_2$ . This problem, known as the Behrens-Fisher problem, only has approximate solutions
- Using our approach, the means are compared by computing the plausibility of  $H_0$  or, more generally, of  $H_\delta : \mu_1 - \mu_2 = \delta$





# Belief function solution

- The marginal contour function for  $(\mu_1, \mu_2)$  is

$$\begin{aligned} pl_{\mathbf{y}}(\mu_1, \mu_2) &= \sup_{\sigma_1, \sigma_2} pl_{\mathbf{y}}(\boldsymbol{\theta}) \\ &= \frac{\prod_{k=1}^2 L_{\mathbf{y}_k}(\mu_k, \hat{\sigma}_k(\mu_k))}{\prod_{k=1}^2 L_{\mathbf{y}_k}(\bar{y}_k, s_k)}, \end{aligned}$$

where

$$\hat{\sigma}_k(\mu_k) = \frac{1}{n_k} \sum_{i=1}^{n_k} (y_{ki} - \mu_k)^2.$$

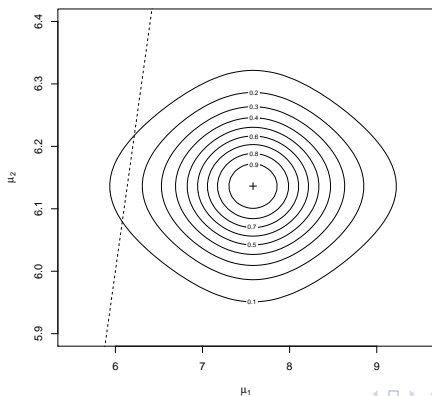
- The plausibility of  $H_\delta = \{(\mu_1, \mu_2) \in \mathbb{R}^2 \mid \mu_1 - \mu_2 = \delta\}$  can then be computed by maximizing  $pl_{\mathbf{y}}(\mu_1, \mu_2)$  under the constraint  $\mu_1 - \mu_2 = \delta$ , i.e.,

$$Pl_{\mathbf{y}}(H_\delta) = \max_{\mu_1} pl_{\mathbf{y}}(\mu_1, \mu_1 - \delta)$$



## Example (Lehman, 1975)

We consider the following driving times from a person's house to work measured for two different routes:  $\mathbf{y}_1 = (6.5, 6.8, 7.1, 7.3, 10.2)$  and  $\mathbf{y}_2 = (5.8, 5.8, 5.9, 6.0, 6.0, 6.0, 6.3, 6.3, 6.4, 6.5, 6.5)$ . Are the mean traveling times equal?



# Linear regression

## Model

We consider the following **standard regression model**

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

where

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



# Likelihood-based belief function

- The likelihood function for this model is

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

- The contour function can thus be readily calculated as

$$p_{\mathbf{y}}(\boldsymbol{\theta}) = \frac{L_{\mathbf{y}}(\boldsymbol{\theta})}{L_{\mathbf{y}}(\hat{\boldsymbol{\theta}})}$$

with  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma})'$ , where

- $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$  is the ordinary least squares estimate of  $\boldsymbol{\beta}$
- $\hat{\sigma}$  is the standard deviation of residuals



# Plausibility of linear hypotheses

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

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

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

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

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

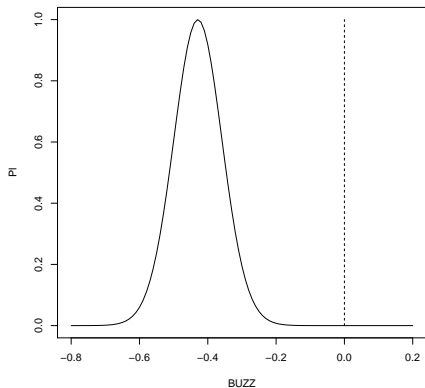
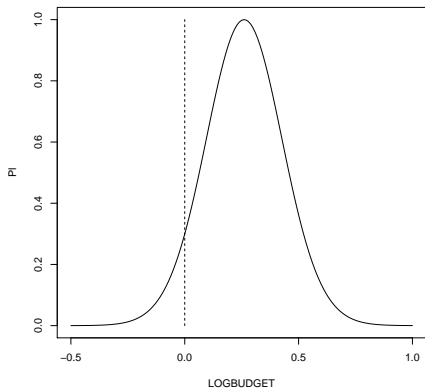


## Example: movie Box office data

- 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 (STARPOWER),
  - 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	$PI(\beta_j = 0)$
(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





# Outline

- 1 Belief functions on infinite spaces
  - Definition
  - Practical models
  - Combination
- 2 Estimation
  - Likelihood-based belief function
  - Examples
  - Consistency
- 3 Prediction
  - Predictive belief function
  - Examples



# Consistency of the likelihood-based belief function

- Assume that the observed data  $\mathbf{y} = (y_1, \dots, y_n)$  is a realization of an iid sample  $\mathbf{Y} = (Y_1, \dots, Y_n)$  from  $Y \sim f_{\theta}(y)$
- From Fraser (1968):

## Theorem

*If  $\mathbb{E}_{\theta_0}[\log f_{\theta}(Y)]$  exists, is finite for all  $\theta$ , and has a unique maximum at  $\theta_0$ , then, for any  $\theta \neq \theta_0$ ,  $p'_n(\theta) \rightarrow 0$  almost surely under the law determined by  $\theta_0$*



# Consistency of the likelihood-based belief function (continued)

- The property  $pI_n(\theta_0) \rightarrow 1$  a.s. does not hold in general (under regularity assumptions,  $-2 \log pI_n(\theta_0)$  converges in distribution to  $\chi_p^2$ )
- But we have the following theorem:

## Theorem

*Under some assumptions (Fraser, 1968), for any neighborhood  $N$  of  $\theta_0$ ,  $Bel_n^\ominus(N) \rightarrow 1$  and  $Pl_n^\ominus(N) \rightarrow 1$  almost surely under the law determined by  $\theta_0$*



# Outline

## 1 Belief functions on infinite spaces

- Definition
- Practical models
- Combination

## 2 Estimation

- Likelihood-based belief function
- Examples
- Consistency

## 3 Prediction

- Predictive belief function
- Examples



# Outline

- 1 Belief functions on infinite spaces
  - Definition
  - Practical models
  - Combination
- 2 Estimation
  - Likelihood-based belief function
  - Examples
  - Consistency
- 3 Prediction
  - Predictive belief function
  - Examples



# Prediction problem

- Observed (past) data:  $\mathbf{y}$  from  $\mathbf{Y} \sim f_{\theta}(\mathbf{y})$
- Future data:  $Z|\mathbf{y} \sim F_{\theta,\mathbf{y}}(z)$  (real random variable)
- Problem: quantify the uncertainty of  $Z$  using a predictive belief function



## Outline of the approach (1/2)

- Let us come back to the urn example
- Let  $Z \sim \mathcal{B}(\theta)$  be defined as

$$Z = \begin{cases} 1 & \text{if next ball is black} \\ 0 & \text{otherwise} \end{cases}$$

- We can write  $Z$  as a function of  $\theta$  and a **pivotal variable**  $W \sim \mathcal{U}([0, 1])$ ,

$$\begin{aligned} Z &= \begin{cases} 1 & \text{if } W \leq \theta \\ 0 & \text{otherwise} \end{cases} \\ &= \varphi(\theta, W) \end{aligned}$$



# Outline of the approach (2/2)

- The equality

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

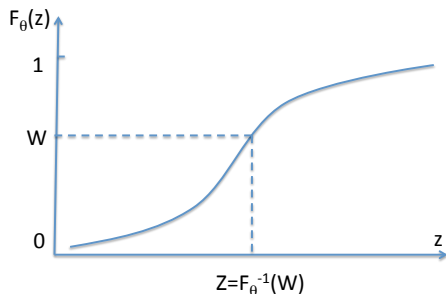
allows us to separate the two sources of uncertainty on  $Z$

- ① uncertainty on  $W$  (random/aleatory uncertainty)
- ② uncertainty on  $\theta$  (estimation/epistemic uncertainty)
- Two-step method:
  - ① Represent uncertainty on  $\theta$  using a likelihood-based belief function  $Bel_y^\theta$  constructed from the observed data  $y$  (estimation problem)
  - ② Combine  $Bel_y^\theta$  with the probability distribution of  $W$  to obtain a predictive belief function  $Bel_y^Z$





# $\varphi$ -equation



We can always write  $Z$  as a function of  $\theta$  and  $W$  as

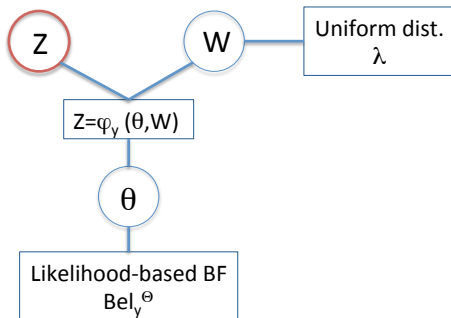
$$Z = F_{\theta, y}^{-1}(W) = \varphi_y(\theta, W)$$

where  $W \sim \mathcal{U}([0, 1])$  and  $F_{\theta, y}^{-1}$  is the generalized inverse of  $F_{\theta, y}$ ,

$$F_{\theta, y}^{-1}(W) = \inf\{z | F_{\theta, y}(z) \geq W\}$$



# Main result



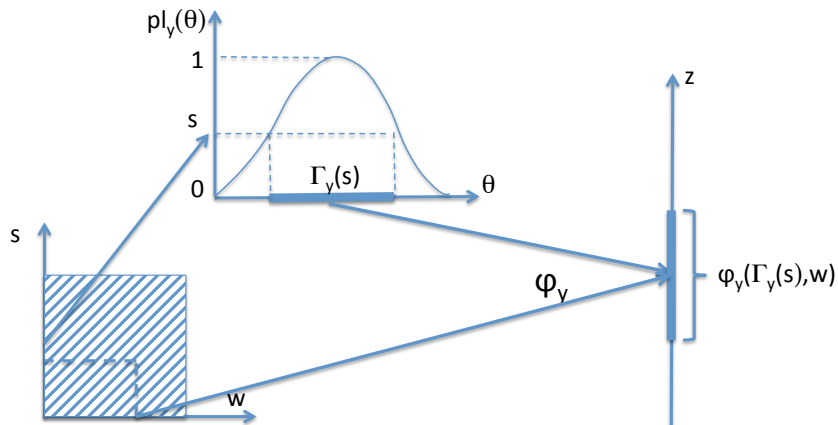
After combination by Dempster's rule and marginalization on  $\mathbb{Z}$ , we obtain the predictive BF on  $Z$  induced by the multi-valued mapping

$$(s, w) \rightarrow \varphi_y(\Gamma_y(s), w).$$

with  $(s, w)$  uniformly distributed in  $[0, 1]^2$



# Graphical representation



# Practical computation

- Analytical expression when possible (simple cases), or
- Monte Carlo simulation:
  - 1 Draw  $N$  pairs  $(s_i, w_i)$  independently from a uniform distribution
  - 2 compute (or approximate) the focal sets  $\varphi_{\mathbf{y}}(\Gamma_{\mathbf{y}}(s_i), w_i)$
- The predictive belief and plausibility of any subset  $A \subseteq \mathbb{Z}$  are then estimated by

$$\widehat{Bel}_{\mathbf{y}}^{\mathbb{Z}}(A) = \frac{1}{N} \#\{i \in \{1, \dots, N\} \mid \varphi_{\mathbf{y}}(\Gamma_{\mathbf{y}}(s_i), w_i) \subseteq A\}$$

$$\widehat{Pl}_{\mathbf{y}}^{\mathbb{Z}}(A) = \frac{1}{N} \#\{i \in \{1, \dots, N\} \mid \varphi_{\mathbf{y}}(\Gamma_{\mathbf{y}}(s_i), w_i) \cap A \neq \emptyset\}$$



## Example: the urn model

- Here,  $Y \sim \mathcal{B}(n, \theta)$ . The likelihood-based BF is induced by a random interval

$$\Gamma(s) = \{\theta : p|_y(\theta) \geq s\} = [\underline{\theta}(s), \bar{\theta}(s)]$$

- We have

$$Z = \varphi(\theta, W) = \begin{cases} 1 & \text{if } W \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

- Consequently,

$$\varphi(\Gamma(s), W) = \varphi([\underline{\theta}(s), \bar{\theta}(s)], W) = \begin{cases} \{1\} & \text{if } W \leq \underline{\theta}(s) \\ \{0\} & \text{if } \bar{\theta}(s) < W \\ \{0, 1\} & \text{otherwise} \end{cases}$$



# Example: the urn model

## Properties

We have

$$Bel_y(\{1\}) = \mathbb{E}(\underline{\theta}(S)) = \hat{\theta} - \int_0^{\hat{\theta}} pI_y(\theta)d\theta$$

$$Pl_y(\{1\}) = \mathbb{E}(\bar{\theta}(S)) = \hat{\theta} + \int_{\hat{\theta}}^1 pI_y(\theta)d\theta$$

So

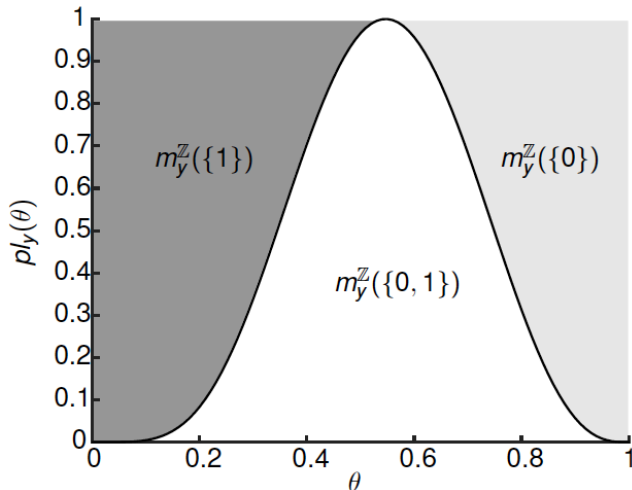
$$m(\{0, 1\}) = \int_0^1 pI_y(\theta)d\theta$$

As  $n \rightarrow \infty$ ,  $m(\{1\}) \rightarrow 1$  and  $m(\{0, 1\}) \rightarrow 0$  in probability.



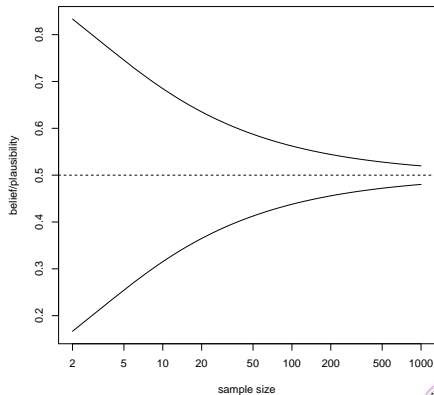
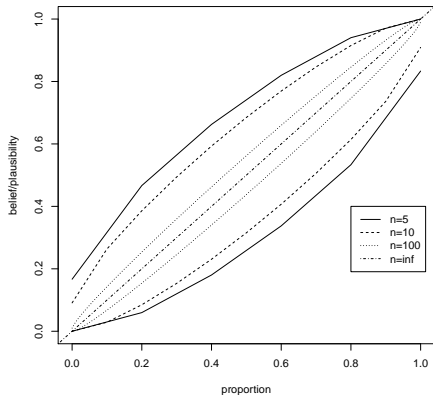
# Example: the urn model

## Geometric representation



# Example: the urn model

## Belief/plausibility intervals





# Uniform example

- Assume that  $Y_1, \dots, Y_n, Z$  is iid from  $\mathcal{U}([0, \theta])$
- Then  $F_\theta(z) = z/\theta$  for all  $0 \leq z \leq \theta$  and we can write  $Z = \theta W$  with  $W \sim \mathcal{U}([0, 1])$
- We have seen that the belief function  $Bel_{\mathbf{y}}^\Theta$  after observing  $\mathbf{Y} = \mathbf{y}$  is induced by the random interval  $[y_{(n)}, y_{(n)} S^{-1/n}]$
- Each focal set of  $Bel_{\mathbf{y}}^{\mathbb{Z}}$  is an interval

$$\varphi(\Gamma_{\mathbf{y}}(s), w) = [y_{(n)} w, y_{(n)} s^{-1/n} w]$$

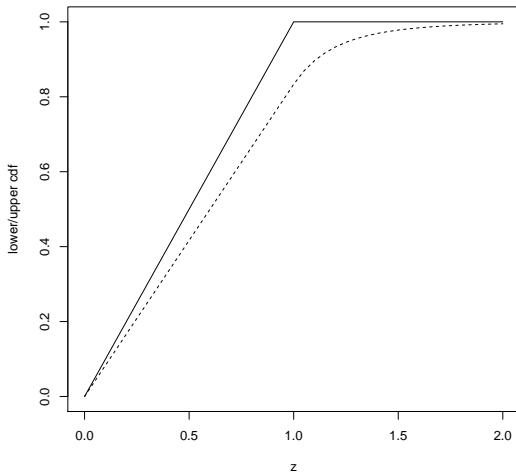
- The predictive belief function  $Bel_{\mathbf{y}}^{\mathbb{Z}}$  is induced by the random interval

$$[\widehat{Z}_{\mathbf{y}^*}, \widehat{Z}_{\mathbf{y}}^*] = [y_{(n)} W, y_{(n)} S^{-1/n} W]$$



# Uniform example

Lower and upper cdfs



# Uniform example

## Consistency

- From the consistency of the MLE,  $Y_{(n)}$  converges in probability to  $\theta_0$ , so

$$\widehat{Z}_{Y^*} = Y_{(n)}W \xrightarrow{d} \theta_0W = Z$$

- We have  $\mathbb{E}(S^{-1/n}) = n/(n-1)$ , and

$$\text{Var}(S^{-1/n}) = \frac{n}{(n-2)(n-1)^2}$$

- Consequently,  $\mathbb{E}(S^{-1/n}) \rightarrow 1$  and  $\text{Var}(S^{-1/n}) \rightarrow 0$ , so  $S^{-1/n} \xrightarrow{P} 1$
- Hence,

$$\widehat{Z}_{Y^*} = Y_{(n)}S^{-1/n}W \xrightarrow{d} \theta_0W = Z$$



# Consistency (general case)

- Assume that
  - The observed data  $\mathbf{y} = (y_1, \dots, y_n)$  is a realization of an iid sample  $\mathbf{Y} = (Y_1, \dots, Y_n)$
  - The likelihood function  $L_n(\boldsymbol{\theta})$  is unimodal and upper-semicontinuous, so that its level sets  $\Gamma_n(s)$  are closed and connected, and that function  $\varphi(\boldsymbol{\theta}, w)$  is continuous
- Under these conditions, the random set  $\varphi(\Gamma_n(S), W)$  is a closed random interval  $[\hat{Z}_{*n}, \hat{Z}_n^*]$
- Then:

## Theorem

*Assume that the conditions of the previous theorem hold, and that the predictive belief function  $Bel_n^Z$  is induced by a random closed interval  $[\hat{Z}_{*n}, \hat{Z}_n^*]$ . Then  $\hat{Z}_{*n}$  and  $\hat{Z}_n^*$  both converge in distribution to  $Z$  when  $n$  tends to infinity.*

# Outline

- 1 Belief functions on infinite spaces
  - Definition
  - Practical models
  - Combination
- 2 Estimation
  - Likelihood-based belief function
  - Examples
  - Consistency
- 3 Prediction
  - Predictive belief function
  - Examples



# Linear model: prediction

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

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

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

- We can write, equivalently,

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

where  $w$  has a standard uniform distribution

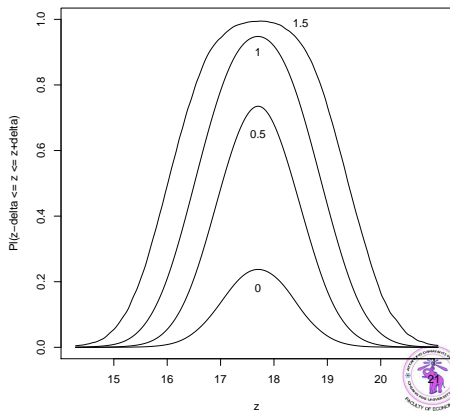
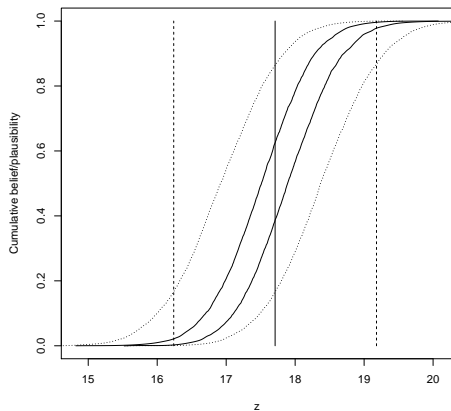
- The predictive belief function on  $z$  can then be approximated using Monte Carlo simulation



# Movie example

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

Lower and upper cdfs



# 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  $x_0$  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  $\mathbf{x}_0$  by a belief function  $Bel^{\mathbb{X}}$  in the sample space  $\mathbb{X}$  of  $\mathbf{x}_0$

- We then have

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

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

$$(\omega, s, w) \rightarrow \varphi_y(\Lambda(\omega), \Gamma_y(s), w)$$

- $Bel_y^{\mathbb{Z}}$  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  $\omega$  from  $\mathbb{P}^\Omega$

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

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

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

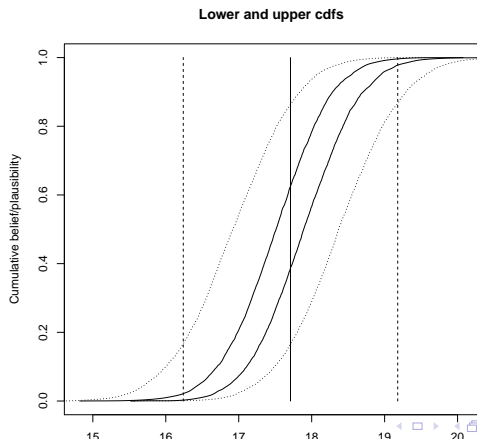
**end for**



# Movie example

## Lower and upper cdfs

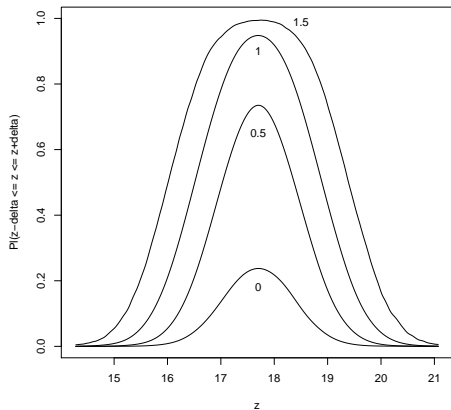
BO success of an action sequel film rated PG13 by MPAA, with LOGBUDGET=5.30, STARPOWER=23.62 and BUZZ= (0,2.81,5) (triangular possibility distribution)?



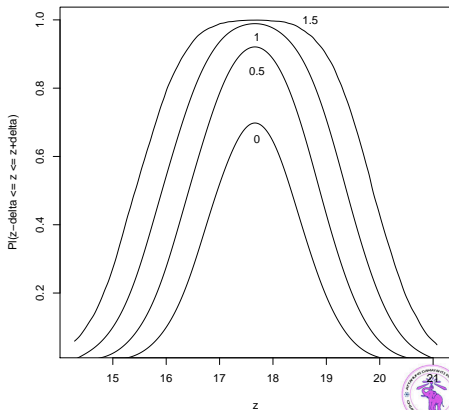
# Movie example

## PI-plots

### Certain inputs



### Uncertain inputs



# Innovation diffusion

- **Forecasting the diffusion of an innovation** has been a topic of considerable interest in marketing research
- 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
- Our approach 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



# Bass model

- Fundamental assumption (Bass, 1969): for eventual adopters, the probability  $f(t)$  of purchase at time  $t$ , given that no purchase has yet been made, is an affine function of the number of previous buyers

$$\frac{f(t)}{1 - F(t)} = p + qF(t)$$

where  $p$  is a **coefficient of innovation**,  $q$  is a **coefficient of imitation** and  $F(t) = \int_0^t f(u)du$ .

- Solving this differential equation, **the probability that an individual taken at random from the population will buy the product before time  $t$  is**

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

where  $c$  is the probability of eventually adopting the product and  $\theta = (p, q)$



# Parameter estimation

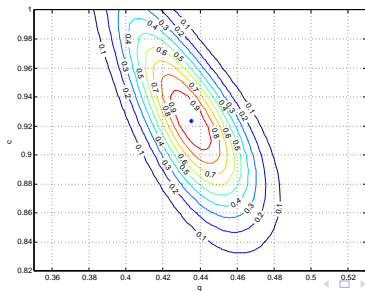
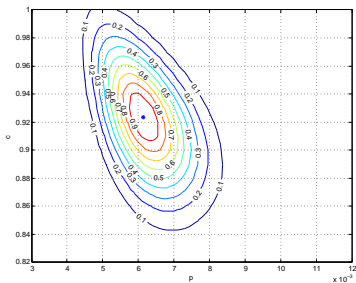
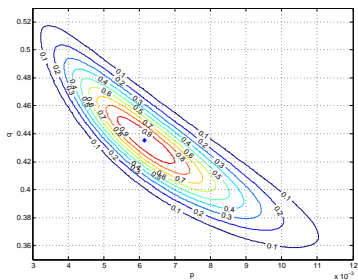
- Data:  $y_1, \dots, y_{T-1}$ , where  $y_i$  = observed number of adopters in time interval  $[t_{i-1}, t_i)$
- The number of individuals in the sample of size  $M$  who did not adopt the product at time  $t_{T-1}$  is  $y_T = M - \sum_{i=1}^{T-1} y_i$
- The probability of adopting the innovation between 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,  $\mathbf{y} = (y_1, \dots, y_T)$  is a realization of  $\mathbf{Y} \sim \mathcal{M}(M, p_1, \dots, p_T)$  and the **likelihood function** is

$$L_{\mathbf{y}}(\theta) \propto \prod_{i=1}^T p_i^{y_i} = \left( \prod_{i=1}^{T-1} [\Phi_\theta(t_i) - \Phi_\theta(t_{i-1})]^{y_i} \right) [1 - \Phi_\theta(t_{T-1})]^{y_T}$$

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



## Results





# Sales forecasting

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

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

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

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

and  $\mathbf{W} = (W_1, \dots, W_Q)$  has a uniform distribution in  $[0, 1]^Q$ .



# Predictive belief function

## Multi-valued mapping

- The **predictive belief function on  $Z$**  is induced by the multi-valued mapping  $(s, \mathbf{w}) \rightarrow \varphi(\Gamma_{\mathbf{y}}(s), \mathbf{w})$  with

$$\Gamma_{\mathbf{y}}(s) = \{\theta \in \Theta : p_{\mathbf{y}}(\theta) \geq s\}$$

- When  $\theta$  varies in  $\Gamma_{\mathbf{y}}(s)$ , the range of  $\pi_{\theta}$  is  $[\underline{\pi}_{\theta}(s), \bar{\pi}_{\theta}(s)]$ , with

$$\underline{\pi}_{\theta}(s) = \min_{\{\theta | p_{\mathbf{y}}(\theta) \geq s\}} \pi_{\theta}, \quad \bar{\pi}_{\theta}(s) = \max_{\{\theta | p_{\mathbf{y}}(\theta) \geq s\}} \pi_{\theta}$$

- We have

$$\varphi(\Gamma_{\mathbf{y}}(s), \mathbf{w}) = [\underline{Z}(s, \mathbf{w}), \bar{Z}(s, \mathbf{w})],$$

where  $\underline{Z}(s, \mathbf{w})$  and  $\bar{Z}(s, \mathbf{w})$  are, respectively, the number of  $w_i$ 's that are less than  $\underline{\pi}_{\theta}(s)$  and  $\bar{\pi}_{\theta}(s)$

- For fixed  $s$ ,  $\underline{Z}(s, \mathbf{W}) \sim \mathcal{B}(Q, \underline{\pi}_{\theta}(s))$  and  $\bar{Z}(s, \mathbf{W}) \sim \mathcal{B}(Q, \bar{\pi}_{\theta}(s))$



# Predictive belief function

## Calculation

- The belief and plausibilities that  $Z$  will be less than  $z$  are

$$Bel_y^{\mathbb{Z}}([0, z]) = \int_0^1 F_{Q, \underline{\pi}_\theta(s)}(z) ds$$

$$Pl_y^{\mathbb{Z}}([0, z]) = \int_0^1 F_{Q, \bar{\pi}_\theta(s)}(z) ds$$

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

- The contour function of  $Z$  is

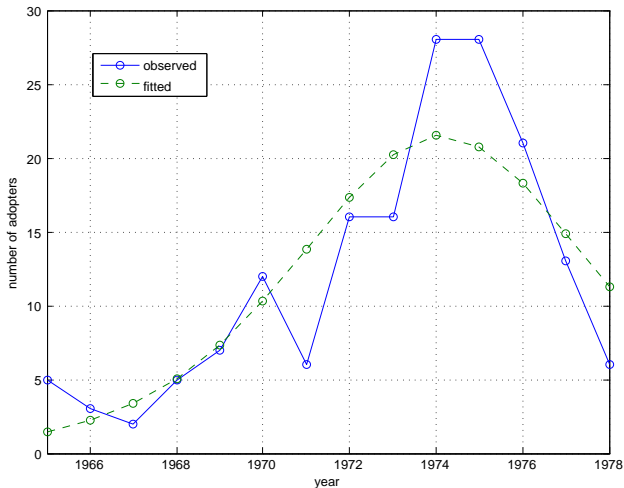
$$pl_y(z) = \int_0^1 (F_{Q, \underline{\pi}_\theta(s)}(z) - F_{Q, \bar{\pi}_\theta(s)}(z-1)) ds$$

- These integrals can be approximated by Monte-Carlo simulation



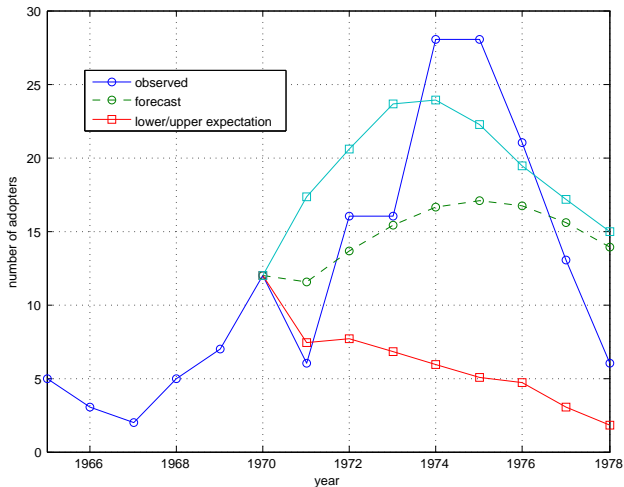
# Ultrasound data

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



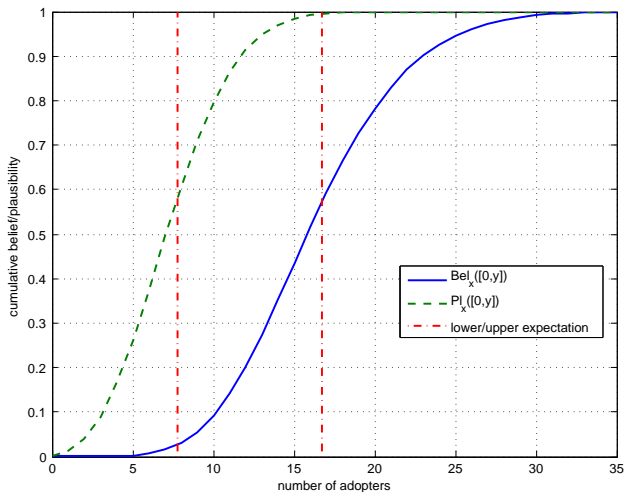
# Forecasting

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



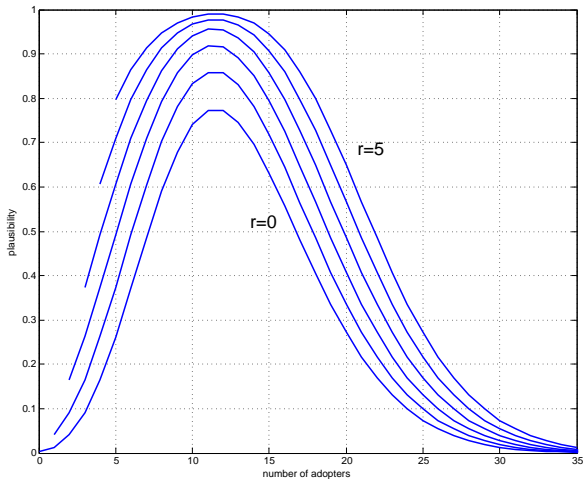
# Cumulative belief and plausibility functions

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



# PI-plot

Plausibilities  $P_{\mathbf{y}}^{\mathbb{Y}}([z - r, z + r])$  as functions of  $z$ , 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 lecture 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  $\theta$  is recovered when a prior on  $\theta$  is available
- The consistency of the method has been established under some conditions





# References

cf. <https://www.hds.utc.fr/~tdenoeux>



T. Denœux.

Likelihood-based belief function: justification and some extensions to low-quality data

*International Journal of Approximate Reasoning*, 55(7):1535–1547, 2014



O. Kanjanatarakul, S. Sriboonchitta and T. Denœux

Forecasting using belief functions. An application to marketing econometrics.

*International Journal of Approximate Reasoning*, 55(5):1113–1128, 2014



O. Kanjanatarakul, T. Denœux and S. Sriboonchitta

Prediction of future observations using belief functions: a likelihood-based approach

*International Journal of Approximate Reasoning*, Vol. 72, pages 71-94, 2016.

