Uncertainty in Statistics: A round table

Didier Dubois
IRIT - CNRS &Université Paul Sabatier -
31062 Toulouse Cedex (France)
e-mail: dubois@irit.fr

Aims and Scope

- In the last 40 years some formalisms have emerged for handling uncertainty and/or human-originated information, that seem to compete with the probabilistic tradition.
  - Fuzzy set theory (Zadeh)
  - Random sets (Kendall, Matheron)
  - Belief functions (Shafer Smets)
  - Possibility theory (Shackle, Zadeh)
  - Imprecise probabilities (Dempster, Walley)
  - etc.
- All these formalisms put forward the use of sets as opposed to or as complementary to the use of probability distributions.
- The aim of the round table is to better understand their impact on statistics and information processing.

Outcome of the workshop

- Collect basic papers by speakers
- Collect significant comments and rejoinders
- Collect replies to questions posed in the round tables
- Based on this material write a joint position paper clarifying the various approaches to uncertainty in statistics and formalized canonical problems

The role of sets in information processing

- Data can take the form of sets (intervals, fuzzy intervals, fuzzy compact sets)
- Results of an information processing procedure can take the form of sets (confidence intervals, partitions, set-valued functions)

The key issue is to understand what these sets represent, what they mean: a set valued model or result is more complex than a usual mathematical model.

What do set-valued data mean ?

- A set can represent an actual object (ontic) or information about an ill-known value (epistemic)
- A set can represent an ill-known constant or an ill-known random variable
- Information represented by a set can be objective (sensor error interval) or subjective (expert opinion)
- can we represent knowledge about complex concepts by linguistic variables on a numerical scale ?

Can we provide a typology of scenarios or canonical example where data appear in the form of sets (or fuzzy sets) ?

Sets vs. probability distributions

- A probability distribution often represents a random phenomenon and captures variability.
  - A subjective probability represents an agent's beliefs.
  - However subjective probabilities are subject to paradoxes (the ignorance paradox, Ellsberg paradox, etc.)
- Sets offer a more natural tool for handling incomplete information
- Mixing sets and probabilities sounds the right way of handling uncertainty in a flexible way
  - Possibility distributions (fuzzy sets)
  - Belief functions (random sets)
  - Imprecise probabilities (convex probability sets)
  - fuzzy random variables

What becomes of basic statistical notions under set-valued representations

- Expectation, median, mode.
- Variance : set-valued or not ?
- Information measures : beyond entropy?
- Conditioning
- Gaussian distributions : do they still make sense for set-valued

Provide a typology of definitions of such notions, plus canonical examples where they make sense and can be useful.

Fuzzy random variables

- Kwakernaak (fuzzy sets of random variables)
- Puri Ralescu (fuzzy random sets)
- Kruse and Meyer (Possibility distribution over random variables)
- Couso and Sanchez 1 (conditional possibility measure $\rightarrow$ convex sets of probabilities)
- Couso and Sanchez 2 (Possibility distribution over probabilities)

Which definitions for which canonical situations ?

Statistical problems with set-valued information

- Regression
- Classification
- Estimation

Claim : it is useless to directly adapt classical techniques or algorithms to (fuzzy) set-valued data.

- Such an approach yields solutions but it is not clear to which problem...

Statistical problems with set-valued information

Question: what do standard information processing problem become under set-valued data ?

- Regression and classification : typology of new formulations of such problems?
- What does it mean to learn an imprecise probability model ?
- Connections between the problem of scarce data and set-valued models: the less data the more imprecise the model ?

-
-
-