The Role of Epistemic uncertainty in risk analysis

D. Dubois

IRIT-CNRS, Université Paul Sabatier
31062 TOULOUSE FRANCE

Origins of uncertainty

- The variability of observed repeatable natural phenomena : « **randomness ».**
  - Coins, dice…: what about the outcome of the next throw?
- The lack of information: **incompleteness**
  - because of information is often lacking, knowledge about issues of interest is generally not perfect.
- Conflicting testimonies or reports: **inconsistency**
  - The more sources, the more likely the inconsistency

Modeling all facets of uncertainty

- Have a language that distinguishes between uncertainty due to variability from uncertainty due to lack of knowledge or missing information.
- **For describing variability: Probability distributions**
  - but information demanding, and paradoxical for ignorance
- **For representing incomplete information : Sets (intervals)** but a very crude representation of uncertainty
- ***Find representations that allow for both aspects of uncertainty*** *:* enrich set representations with confidence levels
- **Fusion methods** for merging testimonies coming from several experts (but not necessarily based on averaging, better conjunctions and disjunctions viz. Dempster rule of conditioning, or argumentation).

  Handling inconsistencies

LANDSCAPE OF UNCERTAINTY THEORIES

BAYESIAN/STATISTICAL PROBABILITY
Randomized points

UPPER-LOWER PROBABILITIES
Disjunctive sets of probabilities

DEMPSTER UPPER-LOWER PROBABILITIES
SHAFER-SMETS **BELIEF FUNCTIONS**
Random disjunctive sets

Quantitative Possibility theory
Fuzzy (nested disjunctive) sets

Classical and modal logic
Disjunctive sets

A risk analysis methodology

Risk can be defined as the combination of the likelihood of occurrence of an undesirable event and the severity of the damage that can be caused by this event.

- Information collection and representation
- Propagation of uncertainty through a mathematical model
- Extraction of useful information
- Decision step

Risk analysis : *Information collection step*

- **Faithfulness principle** : choose the type of representation in agreement with the quantity of available information
  - Remain faithful to available information, including information gaps.
- **Simple representations** (possibility, generalized p-boxes) naturally capture expert interval information with confidence levels, quantiles, means, mode, etc.
  - If *variability and enough statistical information*: probability distributions.
  - If *incomplete information on some value* : interval, possibility distribution (fuzzy interval)…
  - If *parameterized model with ill-known parameters* : p-box
- Devise elicitation procedures dedicated to imprecise probabilities, to query an expert on available information

How useful are these practical representations:

- **Cutting complexity:**
  - Convex sets of probability are very complex representations
  - Random sets are potentially exponential
  - P-boxes, possibility distributions and other extensions are linear, but still encode convex probability sets, often random sets.
- **Enriching the standard probability analysis** with meta-information and capabilities for reasoning about (lack of) knowledge in the risk analysis process, while remaining tractable on modern computers.

Information propagation step

- **Joint Monte-Carlo and interval analysis** to be carried out in the encompassing setting of random sets, with various independence assumptions.
- *Distinction between epistemic (in)dependence and stochastic independence*
  - Dependent sources of information and logical independence between variables
  - Independent sources and dependent variables (e.g. the same variable)
  - No assumption of independence (more difficult to compute)
- Simple representations cannot be preserved via propagation : *general random set outputs are obtained*.

Presentation of results:

how to interpret results?

- summarize too complex information by extracting p-boxes, possibility distributions, or statistical measures of imprecision, trend or variability
  - P-boxes can address questions about threshold violations ($x_{out} \geq a$??)
  - *questions of the form* $a \leq x_{out} \leq b$? are better addressed by possibility distributions or generalized p-boxes
- **Aim**: Lay bare the resulting information gap and the resulting knowledge on the variability of the output quantity of interest.

Decision with imprecise probability: 3 attitudes

- Accept incomparability when comparing imprecise utility evaluations of decisions.
  - Pareto optimality : decisions that dominate other choices for all probability functions
  - E-admissibility : decisions that dominate other choices for at least one probability function (Walley, etc…)
- Select a single probability measure for each decision.
  - Compare lower expectations of decisions (Gilboa)
  - Generalize Hurwicz criterion to focal sets with degree of optimism (Jaffray)
- Select a single probability measure and use expected utility
  - Shapley value = pignistic transformation(SMETS)
  - By picking a probability measure that achieves a compromise between pessimistic and optimistic attitudes

Conclusion

- *There exists a coherent range of uncertainty theories combining interval and probability representations, in the framework of imprecise probability.*
- *Many open problems, theoretical, and computational, remain.*
- *The distinction between epistemic and aleatory uncertainty is essential in risk analysis, so as to provide relevant decision support.*
- *New uncertainty theories complement the usual probabilistic approach, and lay bare this distinction.*
- *They reconcile probability and (modal) logics of incompleteness.*
- *They need to be extended to include inconsistency handling*

Comments on Destercke

- *The proposed approach comes down to turning a histogram of ratings into a collection of intervals and summarizing them by upper and lower expectations*, but
  - It does not question the idea that the average rating is the best summary (while this was questioned by Ben Naim and Prade). Why not the median or the mode of the imprecise probability distribution ???
  - Proposition 2 seems to compare rating profiles by their peakedness: what about using majorisation or entropy ?

Comments on Destercke

- *The proposed approach comes down to turning a histogram of ratings into a collection of intervals and summarizing them by upper and lower expectations*, but
  - The notion of trust seems to be central in electronic commerce and the like : could this notion be useful in risk analysis for discounting expert information?

Comments on Hansen et al.

- *The authors use standard probabilities for BOTH aleatory and epistemic uncertainties*
- *But they handle epistemic uncertainty separately as subjective distributions over scenarii defined by ill-known constant quantities involved in the aleatory space and the evaluation function.*
- *They compute subjective distributions of average output risk values over scenarii,*
  - How does this approach compare with a sensitivity analysis over ill-known constant quantities using p-boxes or imprecise probabilities ?

Comments on Hansen et al.

- *The authors use standard probabilities for BOTH aleatory and epistemic uncertainties*
- *But they handle epistemic uncertainty separately as subjective distributions over scenarii defined by ill-known constant quantities involved in the aleatory space and the evaluation function.*
- *They compute subjective distributions of average output risk values over scenarii,*
  - Does the gap between the lacunary information provided by experts and the subjective probability distributions provided by experts have any significant impact on results ?

Comments on Hansen et al.

- *The authors use standard probabilities for BOTH aleatory and epistemic uncertainties*
- *But they handle epistemic uncertainty separately as subjective distributions over scenarii defined by ill-known constant quantities involved in the aleatory space and the evaluation function.*
- *They compute subjective distributions of average output risk values over scenarii,*
  - Does it make sense to sample a subjective probability distribution describing a constant quantity ? Or is it a computation trick ?