Part 1: Some Additional Remarks

Part 2: Some (of the Many) Challenges

Part 3: Summary View

Part 1: Some Additional Remarks

1. Interpretations

$\surd$ Walley

($\surd$ formal)

- frequentist:

  * lim inf, lim sup of partially diverging relative frequencies
    e.g. Papamarcou & Fine (1991, Ann. Prob.)

  * Law of Large Numbers: Miranda & de Cooman (2004, in preparation)

- logical
  Kyburg, Levi, Weichselberger

2. Further notes on reliability

- $A(n)$: nonparametric predictive inference
  (e.g., Arts, Coolen & van der Laan (2004, Qual. Techn. Quant. Management); Augustin & Coolen (2004, JSPI); Coolen & Yan (2004, JSPI); Coolen & Coolen-Schrijner (2004, Operational Res. Soc.))

  Exchangeable observations $x_1, \ldots x_n$ produce natural division of the real line into n intervals: Probability of the next observation to fall into a certain interval: $\frac{1}{n+1}$

- Survey by Lev V. Utkin: www.levvu.narod.ru

- Igor Kozine

3. Finance

- Pelessoni & Vicig (e.g., 2003a, Int. J. Fuzz. KBS; 2003b, Reliable Computing; 2004, Int. J. Approx. Reas.)
  → close relation between coherent risk measures and lower previsions
  → prudential extension (= cautions alternative to natural extension)
- Schied: optimal investment strategies
- Jaffray, Cohen, Machina
- Insurance: Jeleva

4. On the power of linear programming

- four views of linear programming
- Applying dualization gives "natural understanding" of avoiding sure loss, coherence and natural extension; and easy proofs of lower envelope theorems.

Part 2: Some (of the Many) Challenges

a) Still many foundations are unclear
→ sequential aspects
→ appropriate concept**s** (plural!) of conditional probability

Conditional probability and independence are closely related.
Many independence concepts are needed.

$\Longrightarrow$ Many concepts of conditional probability are needed.
For references see: Walley & de Cooman (1999, Int.J.Approx.Reas.); Weichselberger & Augustin (2003, ISIPTA03)

b) Regression Models: ””in principle”” solved!?

- Rieder
- qualitative robustness
- imprecise observations
- Generalized Likelihood
- Kullback-Leibler information and least favorable pairs
- Bayesian regression models

c) For inference a detailed understanding of problems of sequential information processing is indispensable.

d) Simulations and Monte Carlo Methods

- Simulations to get experience on the behavior of our models (The complexity-based approach by Fierens and Fine 2003, ISIPTA03, might by useful for this.)
- MCMC methods for imprecise probabilities!? (Meeden and Lazar 2003, ISIPTA03)
(use vertex reduction!?)

e) More real applications

Part 3: Summary View

Scott Ferson: Introduction to using imprecise probability in risk analysis

Motivation

"In Reliability the tail is where the action is"

- Oberguggenberger (Talk in Munich) $P(failure) = \in [10^{-3}, 10^{-589}]$
- Propagation of incertitude

Interval Probability

- Interval arithmetic for probabilities
- Probabilistic logic e.g.

$$P(A|\&|B) = [\underline{P}(A) \cdot \underline{P}(B); \overline{P}(A) \cdot \overline{P}(B)]\,,$$

where "|" separates independent parts

- cases of different dependence structures
- Fréchet bounds (best possible when no assumption about dependence)
- repeated variable problems
- software available!

Robust Bayes

- $\begin{array}{rl} & \text{Set of priors} \\ + & \text{(Set of) Likelihood(s)} \\ \hline & \text{Set of posteriors} \end{array} \quad \text{Bayes rule}$

- Problem of zero preservation
- How to define appropriate classes of priors?

Dempster Shafer theory

(Evidence theory, Random Sets, Theory of Hints, ...)

- Basic probability assignment (probability mass function on the power set)
- TA: by the way: why precise? Walley (1991), Miranda, de Cooman & Couso (2002, IPMU; 2004, JSPI, to appear), Augustin (2004b, Int. J. GenS, in Rev.)
- Dempster-Shafer structure
  $\{(A_1, m(A_1)), \ldots\}$
  {(focal element, BPA (basic probability assignment)), ...}
- $A_i = [\underline{a}_i, \overline{a}_i]$ "uncertain number" $\Rightarrow$ cumulative belief functions (Yager), also arithmetic for DS-structures
- handling of censoring and measurement error

Probability bounds analysis

- P-boxes
- Interval bounds on cumulative distribution functions
- Arithmetic under different dependence / independence assumptions, e.g. sum of two random variables
- handling of censoring

Imprecise Probabilities

Gert de Cooman: Imprecise probability models and their behavioral interpretation

Basic concepts:

- in contrast to classical probability: lower previsions are more expressive than lower probabilities
- rigorous and comprehensive theory based on behavioral rationality considerations

	$\underline{P}(X)$		$\overline{P}(X)$	
buy	|	indecision	|	sell

- gambles as basic entity; events via indicator functions
- desirability
- avoiding sure loss
- coherence
- natural extension
- lower envelope theorems give the relation to sets of probability / previsions

Decision Theory

- Maximality (generalization of admissibility: admissibility corresponds to maximality under a vacuous prior)

Building more complex models

- Joint lower previsions
- Conditioning
  - * separate and joint coherence
  - * generalized Bayes rule
- Marginal Extension

$$\underline{P}_Y \otimes \underline{P}(\cdot|Y)$$

Teddy Seidenfeld: Some decision theory with imprecise and indeterminate probability and utility

Static Decision Making

- Horse lotteries, Anscombe-Auman (AA) approach
- $AA1 - AA4$: When can preferences be described by subjective expected utility, i.e. when does a pair $< p, u >$ of probability p and utility u exist that represents the preferences?
- Cooperative Bayesian decision making:
  Assume AA + Pareto, then: If there is any difference in probability *and* utility, then only dictatorship is possible.
- How to relax the Anscombe-Auman, i.e. the subjective expected utility, theory?
  - Γ Maximin $\rightarrow$ Gilboa & Schmeidler (1989)
  - Maximality $\rightarrow$ Seidenfeld, Schervish, Kadane (1995)
  - E-admissibility (Coherence)
    - go beyond pairwise comparisons!
      $\Rightarrow$ choice rule, better: rejection rule
    - axiomatic characterization
    - choice rule unique to $\mathcal{M}$
    - Maximality and E-admissibility are equivalent for closed(!) (convex) sets of probabilities (and convex choice set(?)). What happens at the boundary may matter, however.

Dynamic Decision Making

- deFinetti's book argument
- called-off previsions
- Keeping 'Ordering' (Γ-Maximin) leads to sequential incoherence, while relying on E-admissibility leads to sequential coherence
- Value of information may become negative.

- *Dilation*
  $\underline{P}(A|B) < \underline{P}(A) \leq \overline{P}(A) < \overline{P}(A|B)$ *for* **all** B.

- Also possible for Dempster's rule of conditioning

- "Design issue": if dilation cannot be avoided then design the experiment such that dilation does not hurt you.

Serafin Moral: Independence and graphical models

Different independent concepts

a) unknown interaction

b) epistemic irrelevance (one-sided!)

c) epistemic independence

d) strong independence

e) repetition independence

f) random set independence

g) conditional independence

a) to c) typically associated with epistemic probability;
d) and e) usually for physical probability

Scores for learning in Bayesian networks

- Independence Test CHI
- $K2$ Score
- BIC
- Akaike Criterion
- Upper Entropy of Imprecise Estimation

$$\downarrow$$
$$IDM$$

(with "informative $\overrightarrow{\alpha}$")

Learning credal networks

Propagation in credal networks

Marco Zaffalon: Knowledge discovery from data sets under weak assumptions, application to classification

knowledge **d**iscovery from **d**ata under *tenable* assumptions

Classical approaches

- Bayesian approach in general: choose the class with the highest posterior
- Naive Bayesian classifier
  * unstructured
  * mutually independence
- Tree-augmented naive Bayes (TAN)
- Learn tree structure and probability
- Still allows a clear optimal solution

Further Aspects

- Empirical evaluation
- Generalized cross validation

Necessity of Imprecise Probabilities

Ignorance matters!

- Prior ignorance
- Missing data → partial ignorance on the likelihood,
  MAR ("missing at random") is often questionable

Credal Classification

based on the Imprecise Dirichlet Model (IDM): Walley (1996, JRSSB), Bernard (2004, Int. J. Approx. Reas.)

applying the principle of maximality $\Rightarrow$ credal classification:

- Mapping $f : \mathcal{A}_1 \times \ldots \times \mathcal{A}_n \to \mathcal{P}(\mathcal{C})$ instead of $\mathcal{C}$
- Sets shrink with sample size
- naive credal classifier
- manageable (same computational complexity as naive Bayesian classifier)
- Tree augmented credal classification
- Incomplete (missing, coarsening (e.g. censoring)) data need much care! Observation generating model

0	observed	$\in$	$\mathcal{P}(\mathcal{D}^N)$
D		$\in$	$(\mathcal{D})$

$$\mathbf{E}(f|O) = \inf_{priors} \quad \inf_{models} \mathbf{E}(f|d)$$

- derived as a theorem!
- (TA: "Minimax Likelihood for flat priors")
- conservative learning rules

Classification

- "Independence"
  $$P(0^+ \mid D^+) = P(0^+ \mid D^-)$$
  to avoid vacuous conclusions
- conservative updating rules
- mixed rules!

Thomas Augustin: Robust Neyman Pearson theory & summary view on imprecise probabilities

Some preliminaries

- Weichselberger (1995, 2000, 2001):
  - * R-probability, F-probability; closely related to 'avoiding sure loss' and 'coherence', but σ-additive classical probability as a primitive
  - * structure $\mathcal{M}$: set of all compatible classical probabilities
- Important special case: two-monotone probability
- Interval-valued expectation and/ or Choquet-Integral
- for finite spaces: $\mathcal{M}$ convex polyhedron; Vertex Reduction Lemma (Consideration of $\mathcal{E}(\mathcal{M})$ is sufficient to calculate expectations)

Robust Statistics and Neighborhood Models

- Many standard procedures are highly inrobust
- Idea: Protect yourself by an insurance contract:
- Neighborhood models (and generalizations)
- Huber-Strassen theorem: For two-monotone probabilities there exists a globally least favorable pair and therefore an optimal test.
  Dimension reduction for product spaces possible.
- Extensions to general interval probability

Decision Making I (No-Data Problem)

- Classical optimality criteria are unsatisfactory.
- Generalized expected loss
- Gamma-minimax, Choqeut expected utility: calculating optimal actions by transforming the optimization problem into a single linear programming problem
- Other representation of interval ordering: linear combinationsof lower and upper bounds lead to bilinear optimization

Decision Making II (Data Problem)

- utilizing additional information (data from a sample, expert opinions)
- Main Theorem of Bayesian decision theory is no longer valid
  $\Rightarrow$ two *different* ways to proceed:

    * update imprecise prior to obtain imprecise posterior (sequential view, Robust Bayesian analysis, also advocated by Walley, Teddy's way of applying the Gamma-minimax principle)
        – leads to possibly negative value of information, dilation

  OR

    * calculate risk minimizing decision function (i.e. make problem static by searching for an optimal strategy)
        – problem of counterfactuals: all possible observations matter (not only that outcome which actually occurred). Updating (by calculating conditional probability) and decision making do not coincide.

- debate between Bayesians ('updating', 'conditional view') and frequentists ('overall risk, frequency-based evaluation') gains fundamental importance, now the standpoint matters indeed!