Introduction and fundamentals of probability

Arto Klami

Adaptive Informatics Research Centre
Helsinki University of Technology

19.9.2006

Outline

Outline

Philosophy: Probability?

- Here: Subjectivist view, formal treatment
- “Objectivist” alternatives
  - Classical - physical symmetries
  - Logical - similarities in formal structure
  - Frequentist - independency + classical
- “Useless per se, but turn out to be useful when included as such in the subjectivist theory”

Bayestian statistics

- A rationalist theory of personalistic beliefs in contexts of uncertainty: How to act to avoid undesirable consequences?
- Maximizing expected utility provides basis for rational decision making
- Not descriptive theory: Doesn't model behaviour
- This book answers to "Why?", while the other two books about "How?" and "What?" were never written
- The book seems to have something about those anyway

Structure

- Foundations : Axiomatic basis of probabilities and utilities
- Modelling : Justifications of familiar concepts based on simple assumptions
- Remodelling : Model selection in various perspectives
- Non-Bayesian theories

Outline

1. Introduction
   - Overview of concepts
   - The rest of the book
2. Foundations
   - Decision problems
   - Probabilities
   - Bayes' theorem
3. Exercises

Beliefs and actions

- Uncertainty: individual feeling of incomplete knowledge of a specific situation
- Bayesian theory is a bout the logical process of decision making (=taking an action) in situations of uncertainty
- Do not consider technical difficulties in making decisions with complete information (complex combinatory problems etc)
- Summarizing beliefs is also a decision: a future decision will be made based on the summary

Decision problems

- Actions: $\{a_i, i \in I\}$
- Uncertain events: for each a_i we have $\{E_j, j \in J\}$
- Consequences : for each E_j we have consequence c_j
- Sets of E_j partition the the set of all possibilities
- Actions written as sets of event-consequence pairs : $a_i = \{c_j | E_j, j \in J\}$ denotes that if action a_i is taken and event E_j happens there will be consequence c_j
- Actions may include hypothetical scenarions, hence also called options

Preference

- Preference relation $\leq$ defined for actions: $a_1 \leq a_2 \Rightarrow a_1$ is not preferred to a_2 (induces $\geq, <, >$ and $\sim$)
- Leads to preferences for consequences by considering (hypothetical) actions that always lead to certain consequences : $\{c_1|\Omega\} \leq \{c_2|\Omega\}$
- Also leads to uncertainty relation between events: an event E is consider to be not more likely than F if $\{c_2|E, c_1|E^c\} \leq \{c_2|F, c_1|F^c\}$ for all $c_1 < c_2$
- All can be conditional to some event: $a_1 \leq_G a_2$ iff for all a we have $\{a_1|G, a|G^c\} \leq \{a_2|G, a|G^c\}$

Coherence and quantification

Axioms:

1. Comparability of consequences and actions: If we want to make rational decisions we have to be willing to express (some) preferences
2. Transitivity: Intransitivity equals willingness to suffer unnecessary certain loss
3. Consistency: Preferences between pure consequences do not change, leads to monotonicity for events ($E \subset F \Rightarrow E \leq F$)
4. Existence of standard events S: measure attached to events, analogous to comparison to measure sticks
5. Precise measurement of preferences and uncertainties

Probability

- The probability $P(E)$ of an event is the real number $\mu(S)$ associated with any standard event S such that $E \sim S$
- Personal beliefs because the uncertainty relation is based on the preference relation given for actions
- $P(E)$ exists and is unique
- Finitely additive structure of values in $[0, 1] \Rightarrow$ coherent degrees of beliefs are probabilities
- Complete comparability of events due to axiom 5, even though only partial preference relation was required (?)

Independence

- Events are pairwise independent, $E \perp F$, iff for all c, c_1, c_2

$$c < \{c_2|E, c_1|E^c\} \Rightarrow c <_F \{c_2|E, c_1|E^c\}$$

  and the same for swapping E and F
- Judgements about E are not affected by additional information F
- Equal to the traditional definition: $P(E \bigcap F) = P(E)P(F)$
- Mutual independence: $P(\bigcap_{i \in I} E_i) = \prod_{i \in I} P(E_i)$ for all $I \subset J$

Revision of beliefs

- Conditional probability $P(D|H) = \frac{P(D \bigcap H)}{P(H)}$ corresponds to conditional degrees of belief (derived from axioms!)
- Revision of beliefs by Bayes' theorem:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_j P(D|H_j)P(H_j)}$$

- Standard terminology
  - $P(H_j)$ is the prior probability of H_j
  - $P(D|H_j)$ is the likelihood of H_j given D
  - $P(H_j|D)$ is the posterior probability of H_j
  - $P(D) = \sum_j P(D|H_j)P(H_j)$ is the predictive probability of D

Prior/posterior

- Prior and posterior are not to be interpreted chronologically!
- Coherent degress of beliefs must satisfy the relationship given by Bayes' theorem, but we may for example specify the posterior and conditional likelihood and then find out what our prior was
- In practice it usually is easier to specify the terms in the chronological order

Sequential revision

- If we are given two pieces of data it doesn't matter in which order we update our beliefs, and we can directly consider the combined data
- However, we need to specify conditional likelihoods of the form $P(D_2|H \bigcap D_1)$, unless making independence assumptions
- Conditional independence for D_i given H enables studying the whole collection of events with single likelihood
- Posterior odds could also be used

Outline

Exercise 1

Consider the game of Blackjack in a situation where you have been dealt two tens and the dealer has a nine. Describe the decision problem related to this: What are your possible actions, what are the events for those actions, and what are the consequences? Give also your preference relation of the actions and explain what it tells about your subjective probability estimates.

See http://en.wikipedia.org/wiki/Blackjack#Rules for rules summary if you are not familiar with Blackjack. Remember that the dealer is deterministic, so the only decision is how you act.

Exercise 2

A Construct a counter-example showing that pairwise independence does not imply mutual independence.

B Can you construct a counterexample showing that having $P(\bigcap_{i=1}^{N} E_i) = \prod_{i=1}^{N} P(E_i)$ is not sufficient for mutual independence of N events? Remember that the definition was that $P(\bigcap_{i \in I} E_i) = \prod_{i \in I} P(E_i)$ should hold for all subsets $I \subset \{1..N\}$.