Theory of Belief Functions: Application to machine learning and statistical inference

Lecture 2: Decision analysis

Thierry Denœux

Summer 2023

Outline

1. Decision-making under complete ignorance
2. Decision-making with probabilities
3. Decision-making with belief functions

Example of decision problem under uncertainty

Act (Purchase)	Good Economic Conditions	Poor Economic Conditions
Apartment building	50,000	30,000
Office building	100,000	-40,000
Warehouse	30,000	10,000

Formal framework

Acts, outcomes, states of nature

- A decision problem can be seen as a situation in which a decision-maker (DM) has to choose a course of action (an act) in some set $\mathcal{F} = \{f_1, \ldots, f_n\}$
- An act may have different consequences (outcomes), depending on the state of nature
- Denoting by $\Omega = \{\omega_1, \ldots, \omega_r\}$ the set of states of nature and by $\mathcal{C}$ the set of consequences (or outcomes), an act can be formalized as a mapping f from Ω to $\mathcal{C}$
- In this lecture, the three sets Ω, $\mathcal{C}$ and $\mathcal{F}$ will be assumed to be finite

Formal framework

Utilities

- The desirability of the consequences can often be modeled by a numerical utility function $u : \mathcal{C} \to \mathbb{R}$, which assigns a numerical value to each consequence
- The higher this value, the more desirable is the consequence for the DM
- In some problems, the consequences can be evaluated in terms of monetary value. The utilities can then be defined as the payoffs, or a function thereof
- If the actions are indexed by i and the states of nature by j, we will denote by u_{ij} the quantity $u[f_i(\omega_j)]$
- The $n \times r$ matrix $U = (u_{ij})$ will be called a payoff or utility matrix

Payoff matrix

Act (Purchase)	Good Economic Conditions (ω_1)	Poor Economic Conditions (ω_2)
Apartment building (f_1)	50,000	30,000
Office building (f_2)	100,000	-40,000
Warehouse (f_3)	30,000	10,000

Formal framework

Preferences

- If the true state of nature ω is known, the desirability of an act f can be deduced from that of its consequence $f(\omega)$
- Typically, the state of nature is unknown. Based on partial information, it is usually assumed that the DM can express preferences among acts, which may be represented mathematically by a preference relation $\succcurlyeq$ on $\mathcal{F}$
- This relation is interpreted as follows: given two acts f and g, $f \succcurlyeq g$ means that f is found by the DM to be at least as desirable as g
- We also define
  - The strict preference relation as $f \succ g$ iff $f \succcurlyeq g$ and not$(g \succcurlyeq f)$ (meaning that f is strictly more desirable than g) and
  - The indifference relation $f \sim g$ iff $f \succcurlyeq g$ and $g \succcurlyeq f$ (meaning that f and g are equally desirable)

Decision problems

- The decision problem can be formalized as building a preference relation among acts, from a utility matrix and some description of uncertainty, and finding the maximal elements of this relation
- Depending on the nature of the available information, different decision problems arise:
  1. Decision-making under ignorance
  2. Decision-making with probabilities
  3. Decision-making with belief functions

Outline

1. Decision-making under complete ignorance
2. Decision-making with probabilities
3. Decision-making with belief functions

Problem and non-domination principle

- We assume that the DM is totally ignorant of the state of nature: all the information given to the DM is the utility matrix U
- A act f_i is said to be dominated by f_k if the outcomes of f_k are at least as desirable as those of f_i for all states, and strictly more desirable for at least one state

$$\forall j, \; u_{kj} \geq u_{ij} \text{ and } \exists j, \; u_{kj} > u_{ij}$$

- Non-domination principle: an act cannot be chosen if it is dominated by another one

Example of a dominated act

Act (Purchase)	Good Economic Conditions (ω_1)	Poor Economic Conditions (ω_2)
Apartment building (f_1)	50,000	30,000
Office building (f_2)	100,000	-40,000
~~Warehouse (f_3)~~	~~30,000~~	~~10,000~~

Criteria for rational choice

- After all dominated acts have been removed, there remains the problem of ordering them by desirability, and of finding the set of most desirable acts
- Several criteria of “rational choice” have been proposed to derive a preference relation over acts, including:
  1. Laplace criterion
$$f_i \succeq f_k \text{ iff } \frac{1}{r}\sum_j u_{ij} \geq \frac{1}{r}\sum_j u_{kj}.$$
  2. Maximax criterion
$$f_i \succeq f_k \text{ iff } \max_j u_{ij} \geq \max_j u_{kj}.$$
  3. Maximin (Wald) criterion
$$f_i \succeq f_k \text{ iff } \min_j u_{ij} \geq \min_j u_{kj}.$$

Example

Act	ω_1	ω_2	ave	max	min
Apartment (f_1)	50,000	30,000	**40,000**	50,000	**30,000**
Office (f_2)	100,000	-40,000	30,000	**100,00**	-40,000

Hurwicz criterion

- Hurwicz criterion: $f_i \succeq f_k$ iff

$$\alpha \min_j u_{ij} + (1-\alpha) \max_j u_{ij} \geq \alpha \min_j u_{kj} + (1-\alpha) \max_j u_{kj}$$

  where α is a parameter in $[0, 1]$, called the pessimism index
- Boils down to
  - the maximax criterion if $\alpha = 0$
  - the maximin criterion if $\alpha = 1$
- α describes the DM's attitude toward ambiguity.
- Formal justification given by Arrow and Hurwicz (1972).

Outline

1. Decision-making under complete ignorance
2. Decision-making with probabilities
3. Decision-making with belief functions

Lottery

- Let us now consider the situation where uncertainty about the state of nature is quantified by a probability distribution π on Ω.
- These probabilities can be objective (decision under risk) or subjective.
- An act $f : \Omega \rightarrow \mathcal{C}$ induces a probability measure p on the set $\mathcal{C}$ of consequences (assumed to be finite), called a lottery:

$$\forall c \in \mathcal{C}, \quad p(c) = \sum_{\{\omega : f(\omega) = c\}} \pi(\omega).$$

Maximum Expected Utility principle

- Given a utility function $u : \mathcal{C} \to \mathbb{R}$, the expected utility for a lottery p is

$$\mathbb{E}_p(u) = \sum_{c \in \mathcal{C}} u(c)p(c).$$

- Maximum Expected Utility (MEU) principle: a lottery p_i is more desirable than a lottery p_k if it has a higher expected utility:

$$p_i \succeq p_k \Leftrightarrow \mathbb{E}_{p_i}(u) \geq \mathbb{E}_{p_k}(u).$$

- The MEU principle was first axiomatized by von Neumann and Morgenstern (1944).

Example

Act	ω_1	ω_2
Apartment (f_1)	50,000	30,000
Office (f_2)	100,000	-40,000

- Assume that there is 60% chance that the economic situation will be poor (ω_2).
- Act f_1 induces the lottery p_1 such that $p_1(50,000) = 0.4$ and $p_1(30,000) = 0.6$. Act f_2 induces the lottery p_2 such that $p_2(100,000) = 0.4$ and $p_2(-40,000) = 0.6$.
- The expected utilities are

$$\mathbb{E}_{p_1}(u) = 50,000 \times 0.4 + 30,000 \times 0.6 = 38,000$$
$$\mathbb{E}_{p_2}(u) = 100,000 \times 0.4 - 40,000 \times 0.6 = 16,000$$

- Act f_1 is thus more desirable according to the maximum expected utility criterion.

Outline

How belief functions come into the picture

Belief functions become components of a decision problem in any of the following two situations (or both)

1. The decision maker's subjective beliefs concerning the state of nature are described by a belief function Bel^{Ω} on Ω
2. The DM is not able to precisely describe the outcomes of some acts under each state of nature

Case 1: uncertainty described by a belief function

- Let m^Ω be a mass function on Ω
- Any act $f : \Omega \to \mathcal{C}$ carries m^Ω to the set $\mathcal{C}$ of consequences, yielding a mass function $m_f^\mathcal{C}$, which quantifies the DM's beliefs about the outcome of act f
- Each mass $m^\Omega(A)$ is transfered to $f(A)$

$$m_f^\mathcal{C}(B) = \sum_{\{A \subseteq \Omega : f(A) = B\}} m^\Omega(A)$$

  for any $B \subseteq \mathcal{C}$
- $m_f^\mathcal{C}$ is a credibilistic lottery corresponding to act f

Case 2: partial knowledge of outcomes

- In that case, an act may formally be represented by a multi-valued mapping $f : \Omega \rightarrow 2^{\mathcal{C}}$, assigning a set of possible consequences $f(\omega) \subseteq \mathcal{C}$ to each state of nature ω
- Given a probability measure P on Ω, f then induces the following mass function $m_f^{\mathcal{C}}$ on $\mathcal{C}$,

$$m_f^{\mathcal{C}}(B) = \sum_{\{\omega \in \Omega : f(\omega) = B\}} p(\omega)$$

for all $B \subseteq \mathcal{C}$

Example

- Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and m^Ω the following mass function

$$\begin{array}{ll} m^\Omega(\{\omega_1, \omega_2\}) = 0.3, & m^\Omega(\{\omega_2, \omega_3\}) = 0.2 \\ m^\Omega(\{\omega_3\}) = 0.4, & m^\Omega(\Omega) = 0.1 \end{array}$$

- Let $\mathcal{C} = \{c_1, c_2, c_3\}$ and f the act

$$f(\omega_1) = \{c_1\}, \quad f(\omega_2) = \{c_1, c_2\}, \quad f(\omega_3) = \{c_2, c_3\}$$

- To compute $m_f^\mathcal{C}$, we transfer the masses as follows

$$\begin{aligned} m^\Omega(\{\omega_1, \omega_2\}) = 0.3 &\rightarrow f(\omega_1) \cup f(\omega_2) = \{c_1, c_2\} \\ m^\Omega(\{\omega_2, \omega_3\}) = 0.2 &\rightarrow f(\omega_2) \cup f(\omega_3) = \{c_1, c_2, c_3\} \\ m^\Omega(\{\omega_3\}) = 0.4 &\rightarrow f(\omega_3) = \{c_2, c_3\} \\ m^\Omega(\Omega) = 0.1 &\rightarrow f(\omega_1) \cup f(\omega_2) \cup f(\omega_3) = \{c_1, c_2, c_3\} \end{aligned}$$

- Finally, we obtain the following mass function on $\mathcal{C}$

$$m^\mathcal{C}(\{c_1, c_2\}) = 0.3, \quad m^\mathcal{C}(\{c_2, c_3\}) = 0.4, \quad m^\mathcal{C}(\mathcal{C}) = 0.3$$

Decision problem

- In the two situations considered above, we can assign to each act f a credibilistic lottery, defined as a mass function on $\mathcal{C}$
- Given a utility function u on $\mathcal{C}$, we then need to extend the MEU model
- Several such extensions will now be reviewed

Upper and lower expectations

- Let m be a mass function on $\mathcal{C}$, and u a utility function $\mathcal{C} \to \mathbb{R}$
- The lower and upper expectations of u are defined, respectively, as the averages of the minima and the maxima of u within each focal set of m

$$\underline{\mathbb{E}}_m(u) = \sum_{A \subseteq \mathcal{C}} m(A) \min_{c \in A} u(c)$$

$$\overline{\mathbb{E}}_m(u) = \sum_{A \subseteq \mathcal{C}} m(A) \max_{c \in A} u(c)$$

- It is clear that $\underline{\mathbb{E}}_m(u) \leq \overline{\mathbb{E}}_m(u)$, with the inequality becoming an equality when m is Bayesian, in which case the lower and upper expectations collapse to the usual expectation
- If $m = m_A$ is logical with focal set A, then $\underline{\mathbb{E}}_m(u)$ and $\overline{\mathbb{E}}_m(u)$ are, respectively, the minimum and the maximum of u in A

Corresponding decision criteria

- Having defined the notions of lower and upper expectations, we can define two preference relations among credibilistic lotteries as

$$m_1 \underline{\succcurlyeq} m_2 \text{ iff } \underline{\mathbb{E}}_{m_1}(u) \geq \underline{\mathbb{E}}_{m_2}(u)$$

  and

$$m_1 \overline{\succcurlyeq} m_2 \text{ iff } \overline{\mathbb{E}}_{m_1}(u) \geq \overline{\mathbb{E}}_{m_2}(u)$$

- Relation $\underline{\succcurlyeq}$ corresponds to a pessimistic (or conservative) attitude of the DM. When m is logical, it corresponds to the maximin criterion
- Symmetrically, $\overline{\succcurlyeq}$ corresponds to an optimistic attitude and extends the maximax criterion
- Both criteria boil down to the MEU criterion when m is Bayesian

Generalized Hurwicz criterion

- The Hurwicz criterion can be generalized as

$$\mathbb{E}_{m,\alpha}(u) = \sum_{A \subseteq \mathcal{C}} m(A) \left(\alpha \min_{c \in A} u(c) + (1-\alpha) \max_{c \in A} u(c) \right)$$
$$= \alpha \underline{\mathbb{E}}_m(u) + (1-\alpha) \overline{\mathbb{E}}(u)$$

  where $\alpha \in [0,1]$ is a pessimism index
- This criterion was first introduced and justified axiomatically by Jaffray (1988)

Transferable belief model

- A completely different approach to decision-making with belief function was advocated by Smets, as part of the Transferable Belief Model
- Smets defended a two-level mental model
  1. A credal level, where an agent's beliefs are represented by belief functions, and
  2. A pignistic level, where decisions are made by maximizing the EU with respect to a probability measure derived from a belief function
- The rationale for introducing probabilities at the decision level is the avoidance of Dutch books
- Smets argued that the belief-probability transformation T should be linear, i.e., it should verify

$$T(\alpha m_1 + (1-\alpha)m_2) = \alpha T(m_1) + (1-\alpha)T(m_2),$$

for any mass functions m_1 and m_2 and for any $\alpha \in [0,1]$

Pignistic transformation

- The only linear belief-probability transformation T is the pignistic transformation, with $p_m = T(m)$ given by

$$p_m(c) = \sum_{\{A \subseteq \mathcal{C} : c \in A\}} \frac{m(A)}{|A|}, \quad \forall c \in \mathcal{C}$$

- The expected utility w.r.t. the pignistic probability is

$$\mathbb{E}_p(u) = \sum_{c \in \mathcal{C}} p_m(c) u(c) = \sum_{A \subseteq \mathcal{C}} m(A) \left(\frac{1}{|A|} \sum_{c \in A} u(c) \right)$$

- The maximum pignistic expected utility criterion thus extends the Laplace criterion

Summary

non-probabilized		belief functions	probabilized
maximin	$\longleftrightarrow$	lower expectation	
maximax	$\longleftrightarrow$	upper expectation	
Laplace	$\longleftrightarrow$	pignistic expectation	expected utility
Hurwicz	$\longleftrightarrow$	generalized Hurwicz	