Likelihood-based imprecise probabilities and decision making

Marco Cattaneo
Department of Statistics, LMU Munich

WPMSIIP 2012, Munich, Germany
14 September 2012

learning of IP models

- **statistical model**: $\{P_\theta : \theta \in \Theta\}$

learning of IP models

- **statistical model**: $\{P_\theta : \theta \in \Theta\}$
- (normalized) **likelihood function**: $lik : \theta \mapsto P_\theta(data)$ with $\sup_{\theta \in \Theta} lik(\theta) = 1$

learning of IP models

- **statistical model**: $\{P_\theta : \theta \in \Theta\}$

- (normalized) **likelihood function**: $lik : \theta \mapsto P_\theta(data)$ with $\sup_{\theta \in \Theta} lik(\theta) = 1$

- **general approach to the learning of IP models** (e.g., Antonucci, Cattaneo, and Corani, 2012; Cattaneo and Wiencierz, 2012):

$$\mathcal{P} = \{P_\theta : lik(\theta) > \beta\}$$

  for some $\beta \in (0, 1)$ (i.e., the IP model $\mathcal{P}$ is the likelihood-based confidence region for P_θ with cutoff point β)

learning of IP models

- **statistical model**: $\{P_\theta : \theta \in \Theta\}$

- (normalized) **likelihood function**: $lik : \theta \mapsto P_\theta(data)$ with $\sup_{\theta \in \Theta} lik(\theta) = 1$

- **general approach to the learning of IP models** (e.g., Antonucci, Cattaneo, and Corani, 2012; Cattaneo and Wiencierz, 2012):

$$\mathcal{P} = \{P_\theta : lik(\theta) > \beta\}$$

  for some $\beta \in (0, 1)$ (i.e., the IP model $\mathcal{P}$ is the likelihood-based confidence region for P_θ with cutoff point β)

- learning from data with prior (near) **ignorance** (about θ) is fundamental for statistical applications of IP models

learning of IP models

- **statistical model**: $\{P_\theta : \theta \in \Theta\}$

- (normalized) **likelihood function**: $lik : \theta \mapsto P_\theta(data)$ with $\sup_{\theta \in \Theta} lik(\theta) = 1$

- **general approach to the learning of IP models** (e.g., Antonucci, Cattaneo, and Corani, 2012; Cattaneo and Wiencierz, 2012):

$$\mathcal{P} = \{P_\theta : lik(\theta) > \beta\}$$

  for some $\beta \in (0, 1)$ (i.e., the IP model $\mathcal{P}$ is the likelihood-based confidence region for P_θ with cutoff point β)

- learning from data with prior (near) **ignorance** (about θ) is fundamental for statistical applications of IP models

- prior ignorance, learning, and (Walley-)coherence are **incompatible**: the above general approach relaxes (Walley-)coherence during learning

decision making

- the likelihood-based IP model $\mathcal{P}$ can be used for **decision making**: i.e., for choosing a decision $d \in \mathcal{D}$ according to its (uncertain/random) utility U_d

decision making

- the likelihood-based IP model $\mathcal{P}$ can be used for **decision making**: i.e., for choosing a decision $d \in \mathcal{D}$ according to its (uncertain/random) utility U_d

- in particular, the following **5 criteria** have been considered for IP-based decision making (e.g., Troffaes, 2007): Γ-maximin, Γ-maximax, maximality, E-admissibility, and interval dominance

decision making

- the likelihood-based IP model $\mathcal{P}$ can be used for **decision making**: i.e., for choosing a decision $d \in \mathcal{D}$ according to its (uncertain/random) utility U_d

- in particular, the following **5 criteria** have been considered for IP-based decision making (e.g., Troffaes, 2007): Γ-maximin, Γ-maximax, maximality, E-admissibility, and interval dominance

- the two steps of likelihood-based IP learning and IP-based decision making taken together can be interpreted as **likelihood-based decision making** (Cattaneo, 2007, 2012): in particular, the Γ-maximin criterion corresponds to the LRM (Likelihood-based Region Minimax) criterion

properties

- the decisions resulting from likelihood-based IP learning and IP-based decision making (with **any** of the following 5 criteria: Γ-maximin, Γ-maximax, maximality, E-admissibility, interval dominance) satisfy:

properties

- the decisions resulting from likelihood-based IP learning and IP-based decision making (with **any** of the following 5 criteria: Γ-maximin, Γ-maximax, maximality, E-admissibility, interval dominance) satisfy:
  - several **invariance** properties, e.g.:

properties

- the decisions resulting from likelihood-based IP learning and IP-based decision making (with **any** of the following 5 criteria: Γ-maximin, Γ-maximax, maximality, E-admissibility, interval dominance) satisfy:
  - several **invariance** properties, e.g.:
    - dependence only on sufficient statistics,

properties

- the decisions resulting from likelihood-based IP learning and IP-based decision making (with **any** of the following 5 criteria: Γ-maximin, Γ-maximax, maximality, E-admissibility, interval dominance) satisfy:
  - several **invariance** properties, e.g.:
    - dependence only on sufficient statistics,
    - parametrization invariance,

properties

- the decisions resulting from likelihood-based IP learning and IP-based decision making (with **any** of the following 5 criteria: Γ-maximin, Γ-maximax, maximality, E-admissibility, interval dominance) satisfy:
  - several **invariance** properties, e.g.:
    - dependence only on sufficient statistics,
    - parametrization invariance,
    - equivariance (when the decision problem is invariant),

properties

- the decisions resulting from likelihood-based IP learning and IP-based decision making (with **any** of the following 5 criteria: Γ-maximin, Γ-maximax, maximality, E-admissibility, interval dominance) satisfy:
  - several **invariance** properties, e.g.:
    - dependence only on sufficient statistics,
    - parametrization invariance,
    - equivariance (when the decision problem is invariant),
  - **consistency** (under weak regularity conditions): i.e., the decisions d_n based on the data $X_1, \ldots, X_n$ satisfy (for all $\theta \in \Theta$)

$$E_\theta(U_{d_n}) \xrightarrow{n \to \infty} \sup_{d \in \mathcal{D}} E_\theta(U_d) \quad P_\theta\text{-a.s.}$$

properties

- the decisions resulting from likelihood-based IP learning and IP-based decision making (with **any** of the following 5 criteria: Γ-maximin, Γ-maximax, maximality, E-admissibility, interval dominance) satisfy:
  - several **invariance** properties, e.g.:
    - dependence only on sufficient statistics,
    - parametrization invariance,
    - equivariance (when the decision problem is invariant),
  - **consistency** (under weak regularity conditions): i.e., the decisions d_n based on the data $X_1, \ldots, X_n$ satisfy (for all $\theta \in \Theta$)

$$E_\theta(U_{d_n}) \xrightarrow{n \to \infty} \sup_{d \in \mathcal{D}} E_\theta(U_d) \quad P_\theta\text{-a.s.}$$

- but **only** the Γ-maximin criterion (i.e., the LRM criterion) leads to decisions satisfying **efficiency** (under regularity conditions): i.e., the above convergence is as fast as possible

references

- Antonucci, Cattaneo, and Corani (2012). **Likelihood-based robust classification with Bayesian networks**. In: *Advances in Computational Intelligence*, Part 3, Springer, pp. 491–500.
- Cattaneo (2007). **Statistical Decisions Based Directly on the Likelihood Function**. *PhD thesis*, ETH Zurich.
- Cattaneo (2012). **Likelihood decision functions**. *Technical Report* 128, Department of Statistics, LMU Munich.
- Cattaneo and Wiencierz (2012). **Likelihood-based Imprecise Regression**. *International Journal of Approximate Reasoning* 53, 1137–1154.
- Troffaes (2007). **Decision making under uncertainty using imprecise probabilities**. *International Journal of Approximate Reasoning* 45, 17–29.