Likelihood theory as a unified approach to uncertainty

Marco Cattaneo
Department of Statistics, LMU Munich

29 January 2014

likelihood approach

- statistical model: $(\Omega, \mathcal{A}, P_\theta)$ with $\theta \in \Theta$

likelihood approach

- statistical model: $(\Omega, \mathcal{A}, P_\theta)$ with $\theta \in \Theta$
- (normalized) **likelihood function** given the data $A \in \mathcal{A}$:

$$\lambda : \Theta \to [0,1] \quad \text{with} \quad \lambda(\theta) \propto P_\theta(A) \quad \text{and} \quad \sup_{\theta \in \Theta} \lambda(\theta) = 1$$

likelihood approach

- statistical model: $(\Omega, \mathcal{A}, P_\theta)$ with $\theta \in \Theta$
- (normalized) **likelihood function** given the data $A \in \mathcal{A}$:

$$\lambda : \Theta \to [0,1] \quad \text{with} \quad \lambda(\theta) \propto P_\theta(A) \quad \text{and} \quad \sup_{\theta \in \Theta} \lambda(\theta) = 1$$

- λ is central in all approaches to statistics:

likelihood approach

- statistical model: $(\Omega, \mathcal{A}, P_\theta)$ with $\theta \in \Theta$
- (normalized) **likelihood function** given the data $A \in \mathcal{A}$:

$$\lambda : \Theta \to [0,1] \quad \text{with} \quad \lambda(\theta) \propto P_\theta(A) \quad \text{and} \quad \sup_{\theta \in \Theta} \lambda(\theta) = 1$$

- λ is central in all approaches to statistics:
  - **Bayesian**: uncertain knowledge about θ described by posterior probability measure, obtained by combining λ with a prior probability measure, which cannot describe (complete) ignorance

likelihood approach

- statistical model: $(\Omega, \mathcal{A}, P_\theta)$ with $\theta \in \Theta$
- (normalized) **likelihood function** given the data $A \in \mathcal{A}$:

$$\lambda : \Theta \to [0,1] \quad \text{with} \quad \lambda(\theta) \propto P_\theta(A) \quad \text{and} \quad \sup_{\theta \in \Theta} \lambda(\theta) = 1$$

- λ is central in all approaches to statistics:
  - **Bayesian**: uncertain knowledge about θ described by posterior probability measure, obtained by combining λ with a prior probability measure, which cannot describe (complete) ignorance
  - **classical**: prior ignorance assumed, but no general description of uncertain knowledge: only single statistical methods, mostly based on λ and $\Lambda(B) = \sup_{\theta \in B} \lambda(\theta)$ with $B \subseteq \Theta$

likelihood approach

- statistical model: $(\Omega, \mathcal{A}, P_\theta)$ with $\theta \in \Theta$
- (normalized) **likelihood function** given the data $A \in \mathcal{A}$:

$$\lambda : \Theta \to [0,1] \quad \text{with} \quad \lambda(\theta) \propto P_\theta(A) \quad \text{and} \quad \sup_{\theta \in \Theta} \lambda(\theta) = 1$$

- λ is central in all approaches to statistics:
  - **Bayesian**: uncertain knowledge about θ described by posterior probability measure, obtained by combining λ with a prior probability measure, which cannot describe (complete) ignorance
  - **classical**: prior ignorance assumed, but no general description of uncertain knowledge: only single statistical methods, mostly based on λ and $\Lambda(B) = \sup_{\theta \in B} \lambda(\theta)$ with $B \subseteq \Theta$
  - **likelihood**: uncertain knowledge about θ described by λ, without need of prior information ("prior-free Bayesian" approach)

nonadditive measures

- the set function Λ is a nonadditive measure (imprecise probability measure)

nonadditive measures

- the set function Λ is a nonadditive measure (imprecise probability measure)
- more precisely, Λ is a **maxitive measure** (possibility measure):

$$\Lambda(B \cup C) = \Lambda(B) \vee \Lambda(C) \quad \text{for all (disjoint)} \;\; B, C \subseteq \Theta$$

nonadditive measures

- the set function Λ is a nonadditive measure (imprecise probability measure)
- more precisely, Λ is a **maxitive measure** (possibility measure):

$$\Lambda(B \cup C) = \Lambda(B) \vee \Lambda(C) \quad \text{for all (disjoint)} \;\; B, C \subseteq \Theta$$

- integrals with respect to (non)additive measures [Cattaneo, 2013b]:

nonadditive measures

- the set function Λ is a nonadditive measure (imprecise probability measure)
- more precisely, Λ is a **maxitive measure** (possibility measure):

$$\Lambda(B \cup C) = \Lambda(B) \vee \Lambda(C) \quad \text{for all (disjoint)} \;\; B, C \subseteq \Theta$$

- integrals with respect to (non)additive measures [Cattaneo, 2013b]:
  - **Bayesian approach**: decisions and inferences obtained by minimizing the integral of the loss (or error) with respect to the posterior probability measure

nonadditive measures

- the set function Λ is a nonadditive measure (imprecise probability measure)
- more precisely, Λ is a **maxitive measure** (possibility measure):

$$\Lambda(B \cup C) = \Lambda(B) \vee \Lambda(C) \quad \text{for all (disjoint)} \quad B, C \subseteq \Theta$$

- integrals with respect to (non)additive measures [Cattaneo, 2013b]:
  - **Bayesian approach**: decisions and inferences obtained by minimizing the integral of the loss (or error) with respect to the posterior probability measure
  - **likelihood approach**: decisions and inferences obtained by minimizing the integral of the loss (or error) with respect to the possibility measure Λ [Cattaneo, 2013a]

nonadditive measures

- the set function Λ is a nonadditive measure (imprecise probability measure)
- more precisely, Λ is a **maxitive measure** (possibility measure):

$$\Lambda(B \cup C) = \Lambda(B) \vee \Lambda(C) \quad \text{for all (disjoint)} \;\; B, C \subseteq \Theta$$

- integrals with respect to (non)additive measures [Cattaneo, 2013b]:
  - **Bayesian approach**: decisions and inferences obtained by minimizing the integral of the loss (or error) with respect to the posterior probability measure
  - **likelihood approach**: decisions and inferences obtained by minimizing the integral of the loss (or error) with respect to the possibility measure Λ [Cattaneo, 2013a]
- the likelihood function gives an interpretation to possibility measures and fuzzy sets [Cattaneo, 2008]

likelihood-based imprecise regression

- a general framework for regression with **imprecisely observed data** [Cattaneo and Wiencierz, 2012, 2014]

likelihood-based imprecise regression

- a general framework for regression with **imprecisely observed data** [Cattaneo and Wiencierz, 2012, 2014]

- instead of precisely observing the value of a variable, often only a set containing this value is observed (missing data, censored data, and precisely observed data are special cases)

likelihood-based imprecise regression

- a general framework for regression with **imprecisely observed data** [Cattaneo and Wiencierz, 2012, 2014]

- instead of precisely observing the value of a variable, often only a set containing this value is observed (missing data, censored data, and precisely observed data are special cases)

- the likelihood approach can cope with the uncertainty about the imprecisely observed values

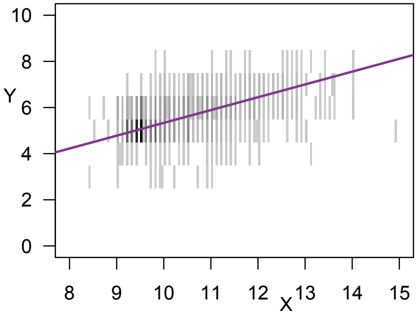
likelihood-based imprecise regression

- a general framework for regression with **imprecisely observed data** [Cattaneo and Wiencierz, 2012, 2014]

- instead of precisely observing the value of a variable, often only a set containing this value is observed (missing data, censored data, and precisely observed data are special cases)

- the likelihood approach can cope with the uncertainty about the imprecisely observed values

- **example**: relationship between the (subjective) sensory quality Y and the alcohol content X (in percent by volume) of $n = 1599$ Vinho Verde red wines from Portugal



probabilistic graphical models

- graphical models describe the (conditional) independences between the involved variables

probabilistic graphical models

- graphical models describe the (conditional) independences between the involved variables

- these independences and the corresponding (conditional) probabilities must be **learned from data**

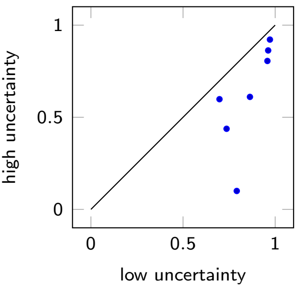
probabilistic graphical models

- graphical models describe the (conditional) independences between the involved variables
- these independences and the corresponding (conditional) probabilities must be **learned from data**
- the likelihood approach can describe the uncertainty about the probability estimates [Cattaneo, 2010]

probabilistic graphical models

- graphical models describe the (conditional) independences between the involved variables
- these independences and the corresponding (conditional) probabilities must be **learned from data**
- the likelihood approach can describe the uncertainty about the probability estimates [Cattaneo, 2010]
- **example**: classifier in which a class is returned only when the probabilities can be estimated with sufficient certainty [Antonucci, Cattaneo, and Corani, 2012]

accuracy of the classification:



present and future research

- likelihood approach and nonadditive measures:

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences
  - connection with risk measures and fuzzy sets

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences
  - connection with risk measures and fuzzy sets
- likelihood-based imprecise regression:

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences
  - connection with risk measures and fuzzy sets
- likelihood-based imprecise regression:
  - effects of stronger assumptions

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences
  - connection with risk measures and fuzzy sets
- likelihood-based imprecise regression:
  - effects of stronger assumptions
  - support vector machines with imprecisely observed data

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences
  - connection with risk measures and fuzzy sets
- likelihood-based imprecise regression:
  - effects of stronger assumptions
  - support vector machines with imprecisely observed data
- probabilistic graphical models:

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences
  - connection with risk measures and fuzzy sets
- likelihood-based imprecise regression:
  - effects of stronger assumptions
  - support vector machines with imprecisely observed data
- probabilistic graphical models:
  - simultaneous learning of independences and probabilities

present and future research

- likelihood approach and nonadditive measures:
  - representation theorems for likelihood-based preferences
  - connection with risk measures and fuzzy sets
- likelihood-based imprecise regression:
  - effects of stronger assumptions
  - support vector machines with imprecisely observed data
- probabilistic graphical models:
  - simultaneous learning of independences and probabilities
  - application to expert systems

references

- Cattaneo and Wiencierz (2014). **On the implementation of LIR: the case of simple linear regression with interval data**. *Computational Statistics*, in press.
- Cattaneo (2013a). **Likelihood decision functions**. *Electronic Journal of Statistics* 7, 2924–2946.
- Cattaneo (2013b). **On maxitive integration**. *Technical Report* 147, Department of Statistics, LMU Munich.
- Cattaneo and Wiencierz (2012). **Likelihood-based Imprecise Regression**. *International Journal of Approximate Reasoning* 53, 1137–1154.
- Antonucci, Cattaneo, and Corani (2012). **Likelihood-based robust classification with Bayesian networks**. *Advances in Computational Intelligence* 3, Springer, 491–500.
- Cattaneo (2010). **Likelihood-based inference for probabilistic graphical models: Some preliminary results**. *PGM 2010*, HIIT Publications, 57–64.
- Cattaneo (2008). **Fuzzy probabilities based on the likelihood function**. *Soft Methods for Handling Variability and Imprecision*, Springer, 43–50.