Decision Making under Model Uncertainty

Marco Cattaneo
Department of Statistics, LMU Munich
cattaneo@stat.uni-muenchen.de

April 24, 2009

past and present research

- foundations of statistics
  (reasoning and decision making under uncertainty)

past and present research

- foundations of statistics
  (reasoning and decision making under uncertainty)

- PhD with Frank Hampel at ETH Zurich
  (November 2002 – March 2007):
  Statistical Decisions Based Directly on the Likelihood Function

past and present research

- foundations of statistics
  (reasoning and decision making under uncertainty)
- PhD with Frank Hampel at ETH Zurich
  (November 2002 – March 2007):
  Statistical Decisions Based Directly on the Likelihood Function
- complex uncertainty:
  simple randomness is superposed by non-stochastic aspects of uncertainty (model uncertainty)

past and present research

- foundations of statistics
  (reasoning and decision making under uncertainty)
- PhD with Frank Hampel at ETH Zurich
  (November 2002 – March 2007):
  Statistical Decisions Based Directly on the Likelihood Function
- complex uncertainty:
  simple randomness is superposed by non-stochastic aspects of uncertainty (model uncertainty)
- Postdoc with Thomas Augustin at LMU Munich
  (SNSF Research Fellowship, October 2007 – March 2009):
  Decision making on the basis of a probabilistic-possibilistic hierarchical description of uncertain knowledge

decision making

- models of reality: $m \in \mathcal{M}$

decision making

- models of reality: $m \in \mathcal{M}$
- consequences of decisions: $c \in \mathcal{C}$

decision making

- models of reality: $m \in \mathcal{M}$
- consequences of decisions: $c \in \mathcal{C}$
- **decisions**: $d : \mathcal{M} \longrightarrow \mathcal{C}$

decision making

- models of reality: $m \in \mathcal{M}$
- consequences of decisions: $c \in \mathcal{C}$
- **decisions**: $d : \mathcal{M} \longrightarrow \mathcal{C}$
- simplifying assumption: $\mathcal{C} = [0, \infty)$,
  $d(m)$ is the loss incurred by making decision d, according to model m

decision making

- models of reality: $m \in \mathcal{M}$
- consequences of decisions: $c \in \mathcal{C}$
- **decisions**: $d : \mathcal{M} \longrightarrow \mathcal{C}$
- simplifying assumption: $\mathcal{C} = [0, \infty)$,
  $d(m)$ is the loss incurred by making decision d, according to model m
- **minimax** decision criterion:

$$\text{minimize} \quad \sup_{m \in \mathcal{M}} d(m)$$

statistical information

- P_m describes the probabilistic forecasts associated with model $m \in \mathcal{M}$

statistical information

- P_m describes the probabilistic forecasts associated with model $m \in \mathcal{M}$

- if the event A describes the observed data, then a **likelihood** function $l : \mathcal{M} \longrightarrow [0, \infty)$ with

$$l(m) \propto P_m(A) \quad \text{for all } m \in \mathcal{M}$$

  describes the (relative) forecasting ability of the models in $\mathcal{M}$

statistical information

- P_m describes the probabilistic forecasts associated with model $m \in \mathcal{M}$

- if the event A describes the observed data, then a **likelihood** function $l : \mathcal{M} \longrightarrow [0, \infty)$ with

$$l(m) \propto P_m(A) \quad \text{for all } m \in \mathcal{M}$$

  describes the (relative) forecasting ability of the models in $\mathcal{M}$

- **maximum likelihood** decision criterion:

$$\text{minimize} \quad d(m_{ML}),$$

  where m_{ML} is the only absolute maximum point of l

statistical information

- P_m describes the probabilistic forecasts associated with model $m \in \mathcal{M}$
- if the event A describes the observed data, then a **likelihood** function $l : \mathcal{M} \longrightarrow [0, \infty)$ with

$$l(m) \propto P_m(A) \quad \text{for all } m \in \mathcal{M}$$

describes the (relative) forecasting ability of the models in $\mathcal{M}$

- **maximum likelihood** decision criterion:

$$\text{minimize} \quad d(m_{ML}),$$

where m_{ML} is the only absolute maximum point of l

- **Bayesian** decision criterion with prior probability measure π on $\mathcal{M}$:

$$\text{minimize} \quad \int d(m)\, l(m)\, \mathrm{d}\pi(m)$$

nonadditive measures and integrals

- in the likelihood approach to statistics, the likelihood function is usually extended to the nonadditive measure λ on $\mathcal{M}$ defined by

$$\lambda(M) = \sup_{m \in M} l(m) \quad \text{for all } M \subseteq \mathcal{M}$$

nonadditive measures and integrals

- in the likelihood approach to statistics, the likelihood function is usually extended to the nonadditive measure λ on $\mathcal{M}$ defined by

$$\lambda(M) = \sup_{m \in M} l(m) \quad \text{for all } M \subseteq \mathcal{M}$$

- λ is a completely maxitive measure (**possibility** measure):

$$\lambda\left(\bigcup_{M \in \mathcal{S}} M\right) = \sup_{M \in \mathcal{S}} \lambda(M) \quad \text{for all } \mathcal{S} \subseteq 2^{\mathcal{M}}$$

nonadditive measures and integrals

- in the likelihood approach to statistics, the likelihood function is usually extended to the nonadditive measure λ on $\mathcal{M}$ defined by

$$\lambda(M) = \sup_{m \in M} l(m) \quad \text{for all } M \subseteq \mathcal{M}$$

- λ is a completely maxitive measure (**possibility** measure):

$$\lambda\left(\bigcup_{M \in \mathcal{S}} M\right) = \sup_{M \in \mathcal{S}} \lambda(M) \quad \text{for all } \mathcal{S} \subseteq 2^{\mathcal{M}}$$

- **likelihood-based** decision criterion:

$$\text{minimize} \quad \int d(m) \, \mathrm{d}\lambda(m)$$

likelihood-based statistical decisions

- the likelihood-based decisions share the properties of the likelihood-based inferences, in particular:
  - asymptotic optimality (consistency)
  - asymptotic efficiency
  - parametrization invariance
  - equivariance

likelihood-based statistical decisions

- the likelihood-based decisions share the properties of the likelihood-based inferences, in particular:
  - asymptotic optimality (consistency)
  - asymptotic efficiency
  - parametrization invariance
  - equivariance

- a simple property (sure-thing principle) characterizes the likelihood-based decision criterion based on the Shilkret integral: the **MPL** (Minimax Plausibility-weighted Loss) decision criterion:

$$\text{minimize} \quad \int^{\mathrm{S}} d(m)\,\mathrm{d}\lambda(m) = \sup_{m \in \mathcal{M}} d(m)\, l(m)$$

example: estimation of variance components

- estimation of the variance components in the 3×3 random effect one-way layout, under normality assumptions and weighted squared error loss

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad \text{for all } i, j \in \{1, 2, 3\}$$

example: estimation of variance components

- estimation of the variance components in the 3×3 random effect one-way layout, under normality assumptions and weighted squared error loss

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad \text{for all } i, j \in \{1, 2, 3\}$$

- normality assumptions:

$$\alpha_i \sim \mathcal{N}(0, v_a), \;\; \varepsilon_{ij} \sim \mathcal{N}(0, v_e), \;\; \text{all independent}$$

$$\Rightarrow \;\; X_{ij} \sim \mathcal{N}(\mu, v_a + v_e) \;\text{ dependent}, \;\; \mu \in \mathbb{R}, \;\; v_a, v_e \in (0, \infty)$$

example: estimation of variance components

- estimation of the variance components in the 3×3 random effect one-way layout, under normality assumptions and weighted squared error loss

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad \text{for all } i, j \in \{1, 2, 3\}$$

- normality assumptions:

$$\alpha_i \sim \mathcal{N}(0, v_a), \;\; \varepsilon_{ij} \sim \mathcal{N}(0, v_e), \;\; \text{all independent}$$

$$\Rightarrow \;\; X_{ij} \sim \mathcal{N}(\mu, v_a + v_e) \;\text{ dependent}, \;\; \mu \in \mathbb{R}, \;\; v_a, v_e \in (0, \infty)$$

- probabilistic models:

$$\mathcal{M} = \{m_{\mu, v_a, v_e} : \mu \in \mathbb{R},\, v_a, v_e \in (0, \infty)\}$$

example: estimation of variance components

- estimates $\widehat{v_e}$ and $\widehat{v_a}$ of variance components v_e and v_a are functions of

$$SS_e = \sum_{i=1}^{3} \sum_{j=1}^{3} (x_{ij} - \bar{x}_{i\cdot})^2 \quad \text{and} \quad SS_a = 3 \sum_{i=1}^{3} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2,$$

where

$$\bar{x}_{i\cdot} = \frac{1}{3} \sum_{j=1}^{3} x_{ij}, \quad \bar{x}_{\cdot\cdot} = \frac{1}{9} \sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij},$$

$$\frac{SS_e}{v_e} \sim \chi^2_6, \quad \text{and} \quad \frac{\frac{1}{3}\, SS_a}{v_a + \frac{1}{3}\, v_e} \sim \chi^2_2$$

example: estimation of variance components

- estimates $\widehat{v_e}$ and $\widehat{v_a}$ of variance components v_e and v_a are functions of

$$SS_e = \sum_{i=1}^{3} \sum_{j=1}^{3} (x_{ij} - \bar{x}_{i\cdot})^2 \quad \text{and} \quad SS_a = 3 \sum_{i=1}^{3} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2,$$

where

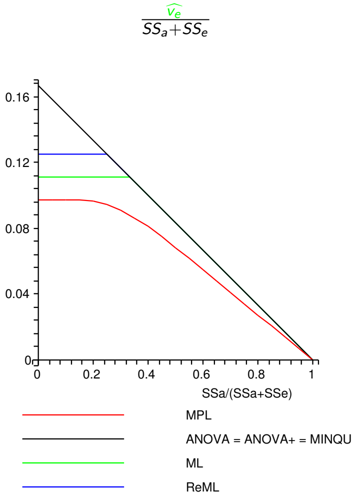
$$\bar{x}_{i\cdot} = \frac{1}{3} \sum_{j=1}^{3} x_{ij}, \quad \bar{x}_{\cdot\cdot} = \frac{1}{9} \sum_{i=1}^{3} \sum_{j=1}^{3} x_{ij},$$

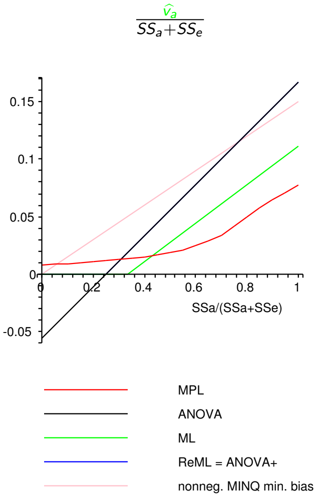
$$\frac{SS_e}{v_e} \sim \chi^2_6, \quad \text{and} \quad \frac{\frac{1}{3}\, SS_a}{v_a + \frac{1}{3}\, v_e} \sim \chi^2_2$$

- decisions (estimates $\widehat{v_e}, \widehat{v_a} \in (0, \infty)$ with invariant loss):

$$d_{\widehat{v_e}}(m_{\mu, v_a, v_e}) = 3\, \frac{(v_e - \widehat{v_e})^2}{{v_e}^2} \quad \text{and} \quad d_{\widehat{v_a}}(m_{\mu, v_a, v_e}) = \frac{(v_a - \widehat{v_a})^2}{(v_a + \frac{1}{3}\, v_e)^2}$$

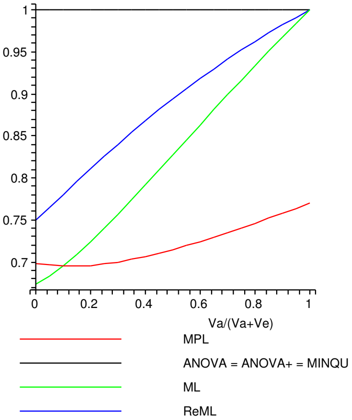
example: estimation of variance components



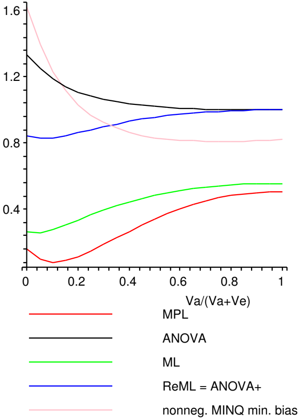


example: estimation of variance components

$$3\ \frac{E[(\widehat{v_e} - v_e)^2]}{{v_e}^2}$$



$$\frac{E[(\widehat{v_a} - v_a)^2]}{(v_a + \frac{1}{3}\ v_e)^2}$$



present and future research

- comparison/combination with other approaches in various applications

present and future research

- comparison/combination with other approaches in various applications
- application to robustness problems
  (relaxation of i.i.d. assumption, robust likelihood)

present and future research

- comparison/combination with other approaches in various applications
- application to robustness problems
  (relaxation of i.i.d. assumption, robust likelihood)
- application to imprecise probability theory
  (better updating, possibilistic previsions, convex set of non-normalized measures)

present and future research

- comparison/combination with other approaches in various applications
- application to robustness problems
  (relaxation of i.i.d. assumption, robust likelihood)
- application to imprecise probability theory
  (better updating, possibilistic previsions, convex set of non-normalized measures)
- application to graphical models
  (probabilistic and non-probabilistic aspects of uncertainty)

present and future research

- comparison/combination with other approaches in various applications
- application to robustness problems
  (relaxation of i.i.d. assumption, robust likelihood)
- application to imprecise probability theory
  (better updating, possibilistic previsions, convex set of non-normalized measures)
- application to graphical models
  (probabilistic and non-probabilistic aspects of uncertainty)
- application to financial risk measures
  (derivation and interpretation of convex risk measures)