The likelihood approach to statistics

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

January 15, 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

foundations of statistics

- besides the Bayesian and frequentist approaches to statistics, there is also a likelihood approach (Fisher, Barnard, Edwards, . . .)

foundations of statistics

- besides the Bayesian and frequentist approaches to statistics, there is also a likelihood approach (Fisher, Barnard, Edwards, ...)

- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ set of probabilistic models:
  each $P_\theta \in \mathcal{P}$ is a probability measure on $(\Omega, \mathcal{A})$

foundations of statistics

- besides the Bayesian and frequentist approaches to statistics, there is also a likelihood approach (Fisher, Barnard, Edwards, ...)
- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ set of probabilistic models: each $P_\theta \in \mathcal{P}$ is a probability measure on $(\Omega, \mathcal{A})$
- when data $A \in \mathcal{A}$ are observed, the **likelihood function**

$$lik(\theta) \propto P_\theta(A)$$

  describes the *relative* ability of the models in $\mathcal{P}$ to forecast the observed data

foundations of statistics

- besides the Bayesian and frequentist approaches to statistics, there is also a likelihood approach (Fisher, Barnard, Edwards, ...)
- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ set of probabilistic models: each $P_\theta \in \mathcal{P}$ is a probability measure on $(\Omega, \mathcal{A})$
- when data $A \in \mathcal{A}$ are observed, the **likelihood function**

  $$\mathit{lik}(\theta) \propto P_\theta(A)$$

  describes the *relative* ability of the models in $\mathcal{P}$ to forecast the observed data
- the likelihood function is a central concept in statistics (both frequentist and Bayesian)

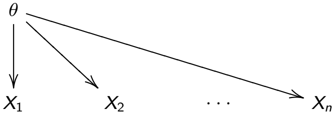
statistical inference

- the likelihood approach allows conditional inference without needing precise prior information

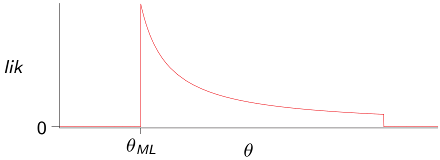
$$\begin{tikzcd}
\theta \arrow[d] \arrow[dr] \arrow[drrr] & & & \\
X_1 & X_2 & \cdots & X_n
\end{tikzcd}$$

statistical inference

- the likelihood approach allows conditional inference without needing precise prior information

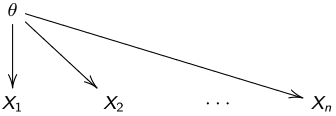


- likelihood-based inference methods are very successful, in particular **maximum likelihood** estimates (parametric or nonparametric)

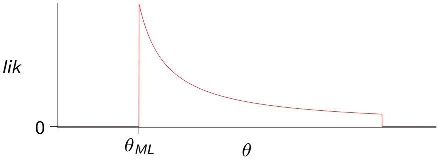


statistical inference

- the likelihood approach allows conditional inference without needing precise prior information



- likelihood-based inference methods are very successful, in particular **maximum likelihood** estimates (parametric or nonparametric)



- surprisingly, no likelihood-based decision making in the statistical literature

statistical decisions

- a statistical decision problem is described by a **loss function**

$$L : \Theta \times \mathcal{D} \to [0, \infty],$$

$L(\theta, d)$: loss incurred by making decision d, according to model P_θ

statistical decisions

- a statistical decision problem is described by a **loss function**

$$L : \Theta \times \mathcal{D} \to [0, \infty],$$

  $L(\theta, d)$: loss incurred by making decision d, according to model P_θ

- Bayesian decision criterion with posterior probability measure π on Θ (obtained from *lik*):

$$\text{minimize} \quad \int L(\theta, d) \, \mathrm{d}\pi(\theta)$$

statistical decisions

- a statistical decision problem is described by a **loss function**

$$L : \Theta \times \mathcal{D} \to [0, \infty],$$

  $L(\theta, d)$: loss incurred by making decision d, according to model P_θ

- Bayesian decision criterion with posterior probability measure π on Θ (obtained from *lik*):

$$\text{minimize} \quad \int L(\theta, d) \, \mathrm{d}\pi(\theta)$$

- in the likelihood approach (in particular in the likelihood ratio test), a **nonadditive** measure λ on Θ is obtained from *lik*:

$$\lambda(\mathcal{H}) = \sup_{\theta \in \mathcal{H}} \mathit{lik}(\theta) \quad \text{for all } \mathcal{H} \subseteq \Theta$$

statistical decisions

- a statistical decision problem is described by a **loss function**

$$L : \Theta \times \mathcal{D} \to [0, \infty],$$

  $L(\theta, d)$: loss incurred by making decision d, according to model P_θ

- Bayesian decision criterion with posterior probability measure π on Θ (obtained from *lik*):

$$\text{minimize} \quad \int L(\theta, d) \,\mathrm{d}\pi(\theta)$$

- in the likelihood approach (in particular in the likelihood ratio test), a **nonadditive** measure λ on Θ is obtained from *lik*:

$$\lambda(\mathcal{H}) = \sup_{\theta \in \mathcal{H}} lik(\theta) \quad \text{for all } \mathcal{H} \subseteq \Theta$$

- likelihood-based decision criterion:

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

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:

$$\text{minimize} \quad \int^{\mathrm{S}} L(\theta, d)\, \mathrm{d}\lambda(\theta) = \sup_{\theta \in \Theta} lik(\theta)\, L(\theta, d)$$

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:

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

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

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

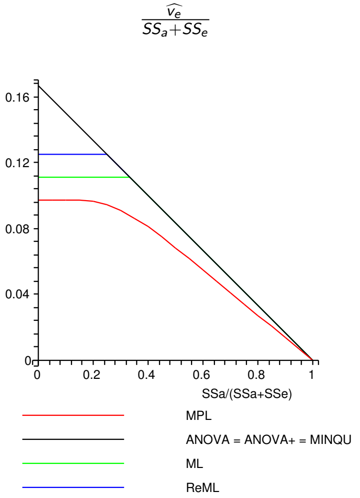
$$\bar{x}_{i\cdot} = \frac{1}{3}\sum_{j=1}^{3} x_{ij}, \quad \bar{x}_{\cdot\cdot} = \frac{1}{9}\sum_{i=1}^{3}\sum_{j=1}^{3} x_{ij},$$

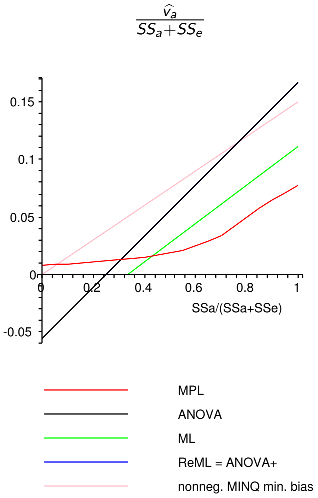
$$\frac{SS_e}{v_e} \sim \chi^2_6, \quad \text{and} \quad \frac{\frac{1}{3}\,SS_a}{v_a + \frac{1}{3}\,v_e} \sim \chi^2_2$$

- invariant loss functions:

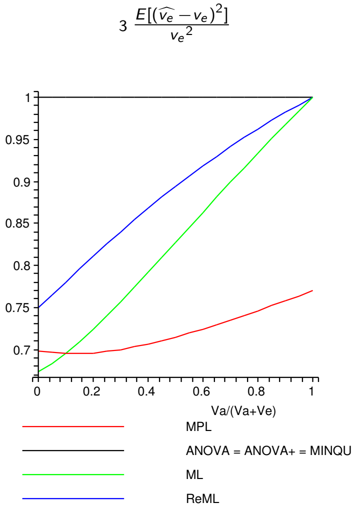
$$L(v_e, \widehat{v_e}) = 3\,\frac{(v_e - \widehat{v_e})^2}{{v_e}^2} \quad \text{and} \quad L(v_a, \widehat{v_a}) = \frac{(v_a - \widehat{v_a})^2}{(v_a + \frac{1}{3}\,v_e)^2}$$

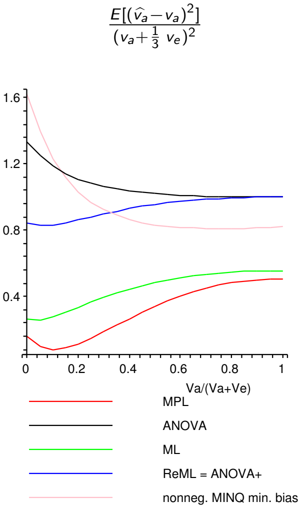
example: estimation of variance components





example: estimation of variance components





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)