Likelihood-Based Statistical Decisions

Marco Cattaneo
Seminar for Statistics
ETH Zürich, Switzerland

July 23, 2005

Likelihood Function

set of statistical **models** $\{P_\theta : \theta \in \Theta\}$
observation A

$\leadsto$ **likelihood function** $lik : \theta \mapsto P_\theta(A)$

Likelihood Function

set of statistical **models** $\{P_\theta : \theta \in \Theta\}$
observation A

$\rightsquigarrow$ **likelihood function** $lik : \theta \mapsto P_\theta(A)$

The likelihood function lik measures the *relative plausibility of the models* P_θ, on the basis of the observation A alone.
The likelihood function lik is *not calibrated*: only ratios $lik(\theta_1)/lik(\theta_2)$ are well determined.

Likelihood Function

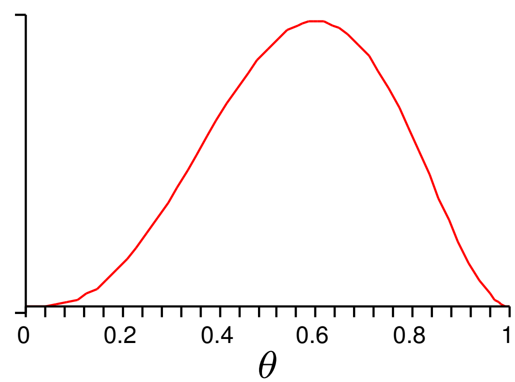
set of statistical **models** $\{P_\theta : \theta \in \Theta\}$
observation A

$\leadsto$ **likelihood function** $lik : \theta \mapsto P_\theta(A)$

The likelihood function lik measures the *relative plausibility of the models* P_θ, on the basis of the observation A alone.
The likelihood function lik is *not calibrated*: only ratios $lik(\theta_1)/lik(\theta_2)$ are well determined.

Example.
$X \sim Binomial\,(n, \theta)$
$n = 5, \quad \theta \in \Theta = [0, 1]$

$x = 3 \quad \Rightarrow \quad lik(\theta) \propto \theta^3\,(1-\theta)^2$



Statistical Decision Problem

set of statistical **models** $\{P_\theta : \theta \in \Theta\}$
set of possible **decisions** $\mathcal{D}$
loss function $L : \Theta \times \mathcal{D} \to [0, \infty)$

$L(\theta, d)$ is the loss we would incur, according to the model P_θ, by making the decision d.

Statistical Decision Problem

set of statistical **models** $\{P_\theta : \theta \in \Theta\}$
set of possible **decisions** $\mathcal{D}$
loss function $L : \Theta \times \mathcal{D} \to [0, \infty)$

$L(\theta, d)$ is the loss we would incur, according to the model P_θ, by making the decision d.

observation $A \rightsquigarrow$ likelihood function lik on Θ

MPL criterion: $\boxed{\text{minimize} \quad \boldsymbol{\sup_\theta lik(\theta)\, L(\theta, d)}}$

Statistical Decision Problem

set of statistical **models** $\{P_\theta : \theta \in \Theta\}$
set of possible **decisions** $\mathcal{D}$
loss function $L : \Theta \times \mathcal{D} \to [0, \infty)$

$L(\theta, d)$ is the loss we would incur, according to the model P_θ, by making the decision d.

observation $A \rightsquigarrow$ likelihood function lik on Θ

MPL criterion: $\boxed{\text{minimize} \quad \boldsymbol{\sup_\theta lik(\theta)\, L(\theta, d)}}$

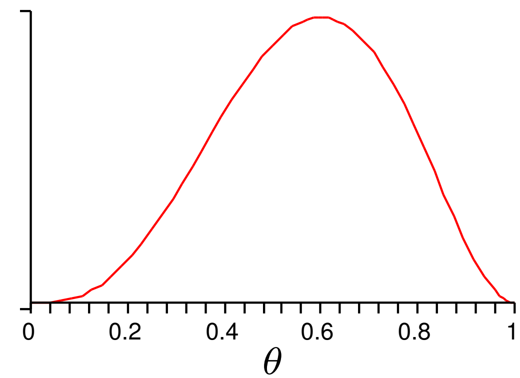
minimax criterion: minimize $\sup_\theta L(\theta, d)$

MPL = minimax if lik is constant (i.e., *complete ignorance* about Θ)
MPL: Minimax Plausibility-weighted Loss

Example

$$lik(\theta) \propto \theta^3 (1-\theta)^2$$
$$L(\theta, d) = |d - \theta^2|$$

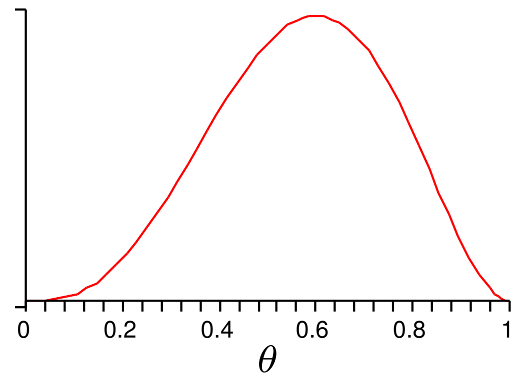
$$d_{ML} = 0.36, \quad d_{MPL} \approx 0.385, \quad d_{BU} \approx 0.335$$



Example

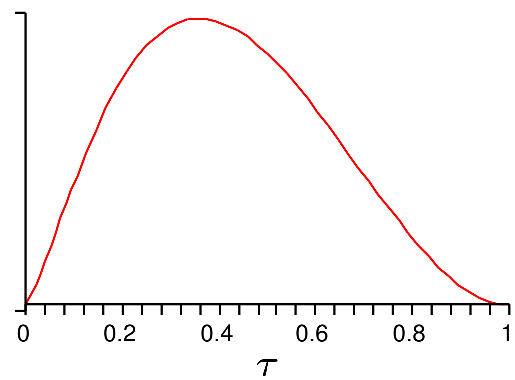
$$lik(\theta) \propto \theta^3 (1-\theta)^2$$
$$L(\theta, d) = |d - \theta^2|$$

$$d_{ML} = 0.36, \quad d_{MPL} \approx 0.385, \quad d_{BU} \approx 0.335$$



$$\tau = \theta^2, \quad lik(\tau) \propto \tau^{\frac{3}{2}} (1-\sqrt{\tau})^2$$
$$L(\tau, d) = |d - \tau|$$

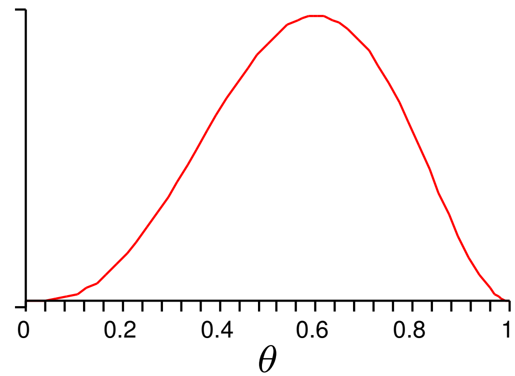
$$d_{ML} = 0.36, \quad d_{MPL} \approx 0.385, \quad d_{BU} \approx 0.404$$



Example

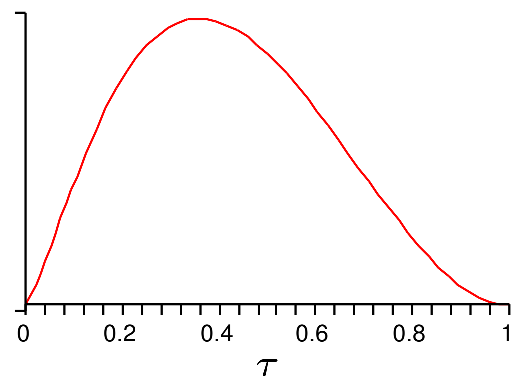
$$lik(\theta) \propto \theta^3 (1-\theta)^2$$
$$L(\theta, d) = |d - \theta^2|$$

$$d_{ML} = 0.36, \quad d_{MPL} \approx 0.385, \quad d_{BU} \approx 0.335$$



$$\tau = \theta^2, \quad lik(\tau) \propto \tau^{\frac{3}{2}} (1 - \sqrt{\tau})^2$$
$$L(\tau, d) = |d - \tau|$$

$$d_{ML} = 0.36, \quad d_{MPL} \approx 0.385, \quad d_{BU} \approx 0.404$$



$$L(\tau, d) = \begin{cases} 2\,|d-\tau| & \text{if } d \le \tau \\ |d-\tau| & \text{if } d \ge \tau \end{cases}$$

$d_{ML} = 0.36, \quad d_{MPL} \approx 0.468, \quad d_{BU} \approx 0.502$ ($d_{BU} \approx 0.435$ using θ)

Relative Plausibility

The likelihood function can be *easily updated* by multiplying it with the (conditional) likelihood functions based on the new observations.

Prior information can be encoded in a "prior likelihood function" assumed to be based on past (independent) observations.

Relative Plausibility

The likelihood function can be *easily updated* by multiplying it with the (conditional) likelihood functions based on the new observations.

Prior information can be encoded in a "prior likelihood function" assumed to be based on past (independent) observations.

The **relative plausibility** is the extension of the likelihood function to the subsets $\mathcal{H}$ of Θ by means of the supremum: $rp(\mathcal{H}) \propto \sup_{\theta \in \mathcal{H}} lik(\theta)$.

Relative Plausibility

The likelihood function can be *easily updated* by multiplying it with the (conditional) likelihood functions based on the new observations.

Prior information can be encoded in a "prior likelihood function" assumed to be based on past (independent) observations.

The **relative plausibility** is the extension of the likelihood function to the subsets $\mathcal{H}$ of Θ by means of the supremum: $rp(\mathcal{H}) \propto \sup_{\theta \in \mathcal{H}} lik(\theta)$.

The relative plausibility is thus a *quantitative description of the uncertain knowledge about the models* P_θ, that can start with complete ignorance or with prior information, that can be easily updated when new data are observed, and that can be used for inference and decision making.

Imprecise Probabilities

The relative plausibility is a *non-calibrated possibility measure* on Θ.

Imprecise Probabilities

The relative plausibility is a *non-calibrated possibility measure* on Θ.

MPL criterion: minimize $\underbrace{\sup_\theta rp\{\theta\}\, L(\theta, d)}_{\textit{Shilkret integral} \text{ of } L(\cdot, d) \text{ with respect to } rp}$

Imprecise Probabilities

The relative plausibility is a *non-calibrated possibility measure* on Θ.

MPL criterion: minimize $\underbrace{\sup_\theta rp\{\theta\}\, L(\theta, d)}_{}$

Shilkret integral of $L(\cdot, d)$ with respect to rp

If Γ is a set of probability measures on Θ, the consideration of the (second-order) relative plausibility on Γ leads to a *non-calibrated possibilistic hierarchical model*, which allows non-vacuous conclusions even if Γ is the set of all probability measures on Θ.

Properties

The relative plausibility and the MPL criterion:

Properties

The relative plausibility and the MPL criterion:

- are simple and intuitive.

Properties

The relative plausibility and the MPL criterion:

- are simple and intuitive.
- are parametrization invariant.

Properties

The relative plausibility and the MPL criterion:

- are simple and intuitive.
- are parametrization invariant.
- lead to decision functions that are equivariant (if the problem is invariant) and asymptotic optimal (if some regularity conditions are satisfied).

Properties

The relative plausibility and the MPL criterion:

- are simple and intuitive.
- are parametrization invariant.
- lead to decision functions that are equivariant (if the problem is invariant) and asymptotic optimal (if some regularity conditions are satisfied).
- satisfy the strong likelihood principle.

Properties

The relative plausibility and the MPL criterion:

- are simple and intuitive.
- are parametrization invariant.
- lead to decision functions that are equivariant (if the problem is invariant) and asymptotic optimal (if some regularity conditions are satisfied).
- satisfy the strong likelihood principle.
- can use pseudo likelihood functions.

Properties

The relative plausibility and the MPL criterion:

- are simple and intuitive.
- are parametrization invariant.
- lead to decision functions that are equivariant (if the problem is invariant) and asymptotic optimal (if some regularity conditions are satisfied).
- satisfy the strong likelihood principle.
- can use pseudo likelihood functions.
- can represent complete (or partial) ignorance.

Properties

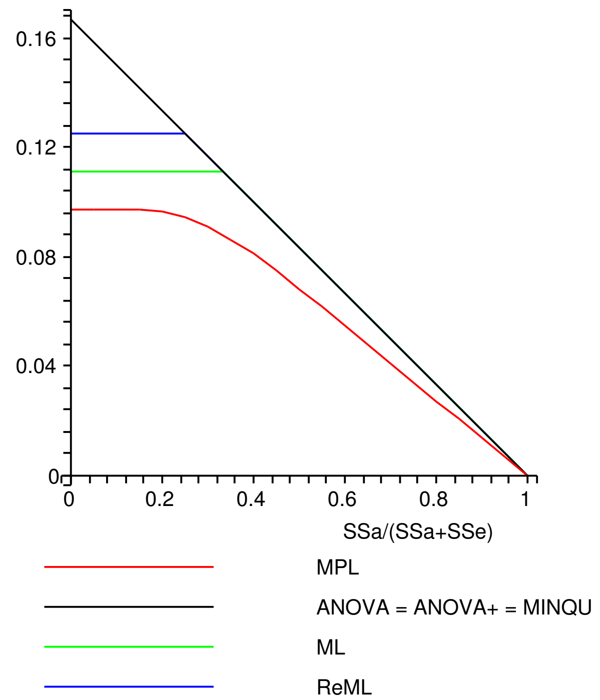
The relative plausibility and the MPL criterion:

- are simple and intuitive.
- are parametrization invariant.
- lead to decision functions that are equivariant (if the problem is invariant) and asymptotic optimal (if some regularity conditions are satisfied).
- satisfy the strong likelihood principle.
- can use pseudo likelihood functions.
- can represent complete (or partial) ignorance.
- can handle prior information in a natural way.

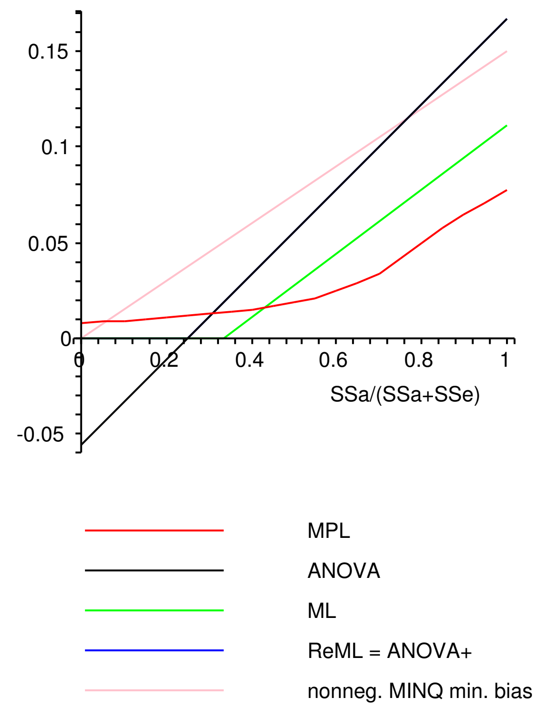
Example

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

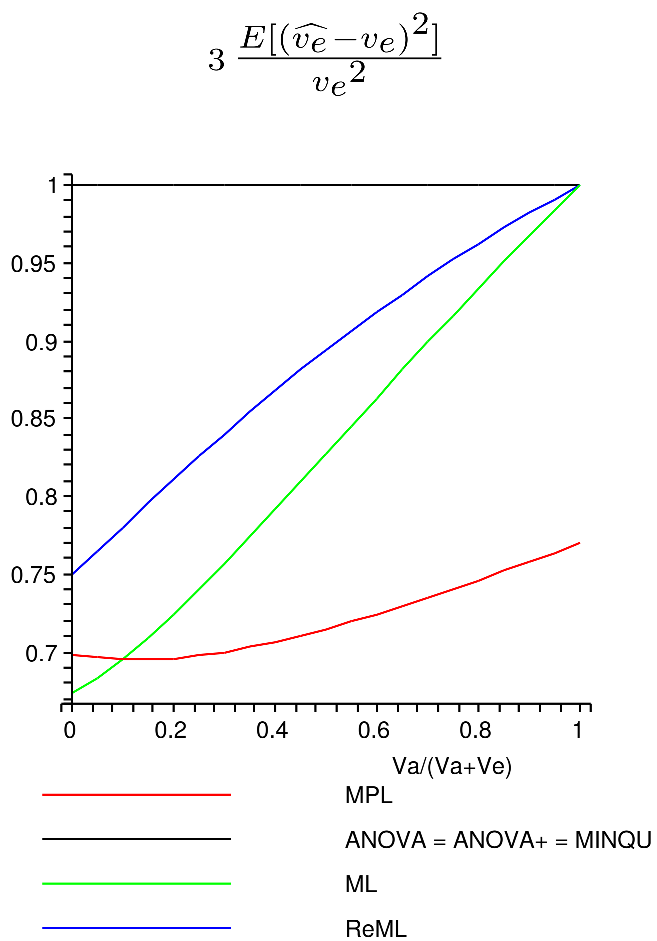
$$\frac{\widehat{v_e}}{(SS_a+SS_e)}$$



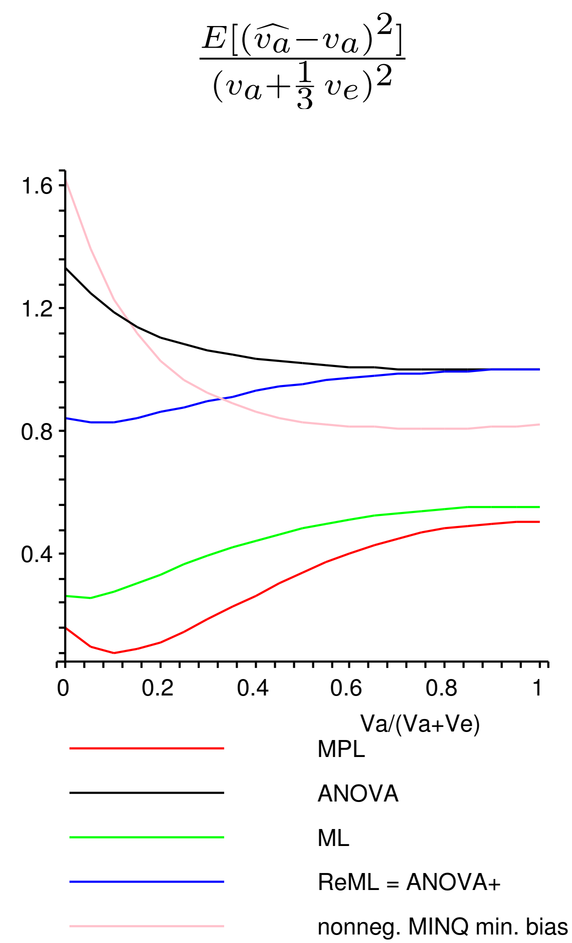
$$\frac{\widehat{v_a}}{(SS_a+SS_e)}$$



Example



$3\,\frac{E[(\widehat{ve}-ve)^2]}{ve^2}$
1
0.95
0.9
0.85
0.8
0.75
0.7
0
0.2
0.4
0.6
0.8
1
Va/(Va+Ve)
MPL
ANOVA = ANOVA+ = MINQU
ML
ReML



$\frac{E[(\widehat{va}-va)^2]}{(va+\frac{1}{3}\,ve)^2}$
1.6
1.2
0.8
0.4
0
0.2
0.4
0.6
0.8
1
Va/(Va+Ve)
MPL
ANOVA
ML
ReML = ANOVA+
nonneg. MINQ min. bias