Likelihood-Based Statistical Decisions

Marco Cattaneo
Seminar for Statistics
ETH Zürich, Switzerland

November 3, 2005

Statistical Models

Let $\mathcal{P}$ be a **set of statistical models**:

- $\mathcal{P}$ is a set of probability measures on a measurable space $(\Omega, \mathcal{A})$;
- no structure is imposed on $\mathcal{P}$.

Statistical Models

Let $\mathcal{P}$ be a **set of statistical models**:

- $\mathcal{P}$ is a set of probability measures on a measurable space $(\Omega, \mathcal{A})$;
- no structure is imposed on $\mathcal{P}$.

The observation of an event $A \in \mathcal{A}$ gives us some information about the models in $\mathcal{P}$. We want to use this information to infer something about the models in $\mathcal{P}$ or to evaluate possible decisions.

Likelihood Function

The **likelihood function** $lik : \mathcal{P} \to [0,1]$ based on the observation $A \in \mathcal{A}$:

- is defined by $lik(P) = P(A)$;
- measures the relative plausibility of the models $P \in \mathcal{P}$, on the basis of the observation A alone;
- is not calibrated: only ratios $lik(P)/lik(P')$ are well determined in a statistical sense.

Likelihood Function

The **likelihood function** $lik : \mathcal{P} \rightarrow [0,1]$ based on the observation $A \in \mathcal{A}$:

- is defined by $\boxed{lik(P) = P(A)}$;
- measures the relative plausibility of the models $P \in \mathcal{P}$, on the basis of the observation A alone;
- is not calibrated: only ratios $lik(P)/lik(P')$ are well determined in a statistical sense.

Likelihood-based inference methods:

- **maximum likelihood estimator** $\boxed{\widehat{P}_{ML} = \arg\max lik}$;
- tests and confidence regions based on the **likelihood ratio statistic** $\boxed{LR(\mathcal{H}) = \frac{\sup I_{\mathcal{H}}\, lik}{\sup lik}}$ for $\mathcal{H} \subseteq \mathcal{P}$.

Example: iid Bernoulli

Let $X_1, X_2, \ldots$ be independent random variables with distribution $Ber(\theta)$ under the model P_θ, and let $\mathcal{P} = \{P_\theta : 0 \leq \theta \leq 1\} \simeq [0, 1]$.

Example: iid Bernoulli

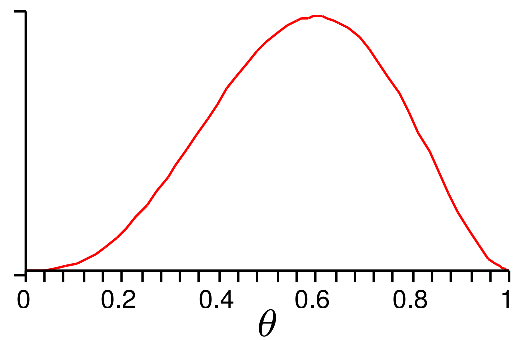
Let $X_1, X_2, \ldots$ be independent random variables with distribution $Ber(\theta)$ under the model P_θ, and let $\mathcal{P} = \{P_\theta : 0 \leq \theta \leq 1\} \simeq [0, 1]$.

$A_5 = \{X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0\} = \langle 10110 \rangle$

$lik(\theta) \propto \theta^3 \, (1-\theta)^2$

$\widehat{\theta}_{ML} = \frac{3}{5} = 0.6$

$LR([0, \frac{1}{2}]) \approx 0.904$



Example: iid Bernoulli

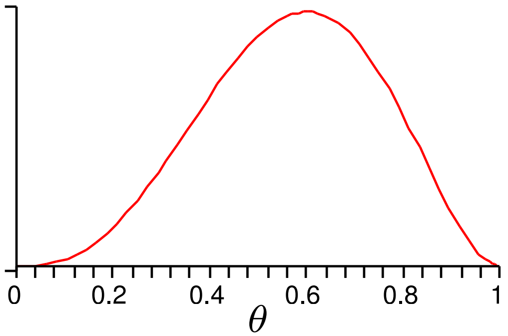
Let $X_1, X_2, \ldots$ be independent random variables with distribution $Ber(\theta)$ under the model P_θ, and let $\mathcal{P} = \{P_\theta : 0 \leq \theta \leq 1\} \simeq [0,1]$.

$A_5 = \{X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0\} = \langle 10110 \rangle$

$lik(\theta) \propto \theta^3 (1-\theta)^2$

$\widehat{\theta}_{ML} = \frac{3}{5} = 0.6$

$LR([0, \frac{1}{2}]) \approx 0.904$

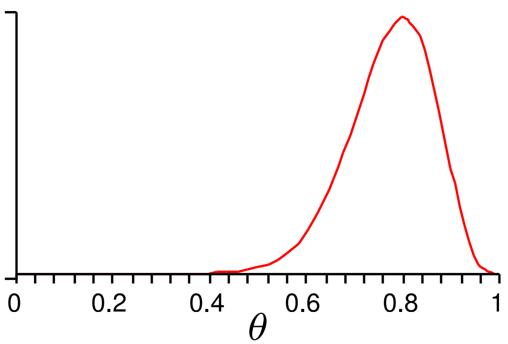


$A_{20} = \langle 10110111101111110111 \rangle$

$lik(\theta) \propto \theta^{16} (1-\theta)^4$

$\widehat{\theta}_{ML} = \frac{16}{20} = 0.8$

$LR([0, \frac{1}{2}]) \approx 0.021$



Likelihood-Based Statistical Inference

The likelihood-based inference methods:

- are asymptotic optimal (if some regularity conditions are satisfied);

Likelihood-Based Statistical Inference

The likelihood-based inference methods:

- are asymptotic optimal (if some regularity conditions are satisfied);
- are general and simple (and therefore widely applicable);
- are intuitive (in particular conditional and parametrization invariant);
- are supported by experience.

Statistical Decision Problem

A statistical decision problem is described by a **loss function** $L : \mathcal{P} \times \mathcal{D} \to [0, \infty)$:

- $\mathcal{D}$ is the set of possible decisions, and $\mathcal{P}$ is the set of considered statistical models;
- $L(P, d)$ is the loss we would incur, according to the model P, by making the decision d;
- the decision d is evaluated on the sole basis of $L_d : P \mapsto L(P, d)$.

Statistical Decision Problem

A statistical decision problem is described by a **loss function** $L : \mathcal{P} \times \mathcal{D} \to [0, \infty)$:

- $\mathcal{D}$ is the set of possible decisions, and $\mathcal{P}$ is the set of considered statistical models;
- $L(P, d)$ is the loss we would incur, according to the model P, by making the decision d;
- the decision d is evaluated on the sole basis of $L_d : P \mapsto L(P, d)$.

Inference can be seen as a special case of decision, for example:

- the estimation of P can be described by $\mathcal{D} = \mathcal{P}$ and L a metric on $\mathcal{P}$;
- a test of $\mathcal{H}_0 \subset \mathcal{P}$ against the alternative $\mathcal{H}_1 = \mathcal{P} \setminus \mathcal{H}_0$ can be described by $\mathcal{D} = \{r, nr\}$, $L_r = I_{\mathcal{H}_0}$ and $L_{nr} = c\, I_{\mathcal{H}_1}$, with $c \leq 1$.

Pre-Data Evaluation

Let $X : \Omega \rightarrow \mathcal{X}$ be a random object with $A = \{X = x_A\}$.

A **decision function** $\delta : \mathcal{X} \rightarrow \mathcal{D}$ is evaluated on the sole basis of the (pre-data) expected loss $R_\delta : P \mapsto E_P[L(P, \delta(X))]$.

Pre-Data Evaluation

Let $X : \Omega \to \mathcal{X}$ be a random object with $A = \{X = x_A\}$.

A **decision function** $\delta : \mathcal{X} \to \mathcal{D}$ is evaluated on the sole basis of the (pre-data) expected loss $R_\delta : P \mapsto E_P[L(P, \delta(X))]$.

To select a decision function δ, adopt a decision criterion, for example the **minimax criterion**: minimize $\sup R_\delta$.

To obtain the decision, apply the selected decision function δ to the observed realization x_A of X.

Post-Data (Conditional) Evaluation

The likelihood-based inference methods are **conditional**:

- they depend only on the observation $A = \{X = x_A\}$, not on the other possible realizations of X;
- they select a decision $d \in \mathcal{D}$, not a decision function $\delta : \mathcal{X} \to \mathcal{D}$.

Post-Data (Conditional) Evaluation

The likelihood-based inference methods are **conditional**:

- they depend only on the observation $A = \{X = x_A\}$, not on the other possible realizations of X;
- they select a decision $d \in \mathcal{D}$, not a decision function $\delta : \mathcal{X} \to \mathcal{D}$.

Since inference can be seen as a special case of decision, we can try to generalize the likelihood-based inference methods to a likelihood-based (conditional) decision criterion.

MPL Decision Criterion

The post-data evaluation of a decision d can only be based on L_d and lik.

A straightforward way to obtain a conditional decision criterion is to associate to every decision $d \in \mathcal{D}$ a nonnegative extended real number $F(L_d, lik)$ (an evaluation of L_d on the basis of lik) and to select d by minimizing $F(L_d, lik)$.

MPL Decision Criterion

The post-data evaluation of a decision d can only be based on L_d and lik.

A straightforward way to obtain a conditional decision criterion is to associate to every decision $d \in \mathcal{D}$ a nonnegative extended real number $F(L_d, lik)$ (an evaluation of L_d on the basis of lik) and to select d by minimizing $F(L_d, lik)$.

For example, by defining $F_{MPL}(L_d, lik) = \sup lik\, L_d$, we obtain the **MPL criterion**: minimize $\sup lik\, L_d$.

MPL Decision Criterion

The post-data evaluation of a decision d can only be based on L_d and lik.

A straightforward way to obtain a conditional decision criterion is to associate to every decision $d \in \mathcal{D}$ a nonnegative extended real number $F(L_d, lik)$ (an evaluation of L_d on the basis of lik) and to select d by minimizing $F(L_d, lik)$.

For example, by defining $F_{MPL}(L_d, lik) = \sup lik\, L_d$, we obtain the **MPL criterion**: minimize $\sup lik\, L_d$.

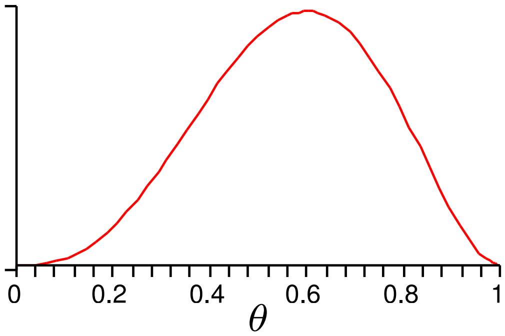
If lik is constant (i.e. we have no information about the models in $\mathcal{P}$), the MPL criterion reduces to: minimize $\sup L_d$ (conditional minimax criterion).

But if lik is not constant, it is used as a weighting of L_d (MPL means Minimax Plausibility-weighted Loss).

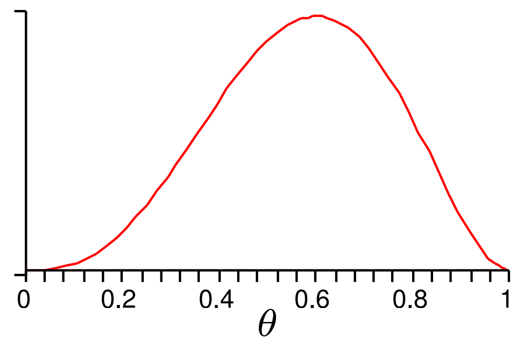
Example: iid Bernoulli

$$A_5 = \langle 10110 \rangle$$

$$lik(\theta) \propto \theta^3 \, (1-\theta)^2$$



Example: iid Bernoulli



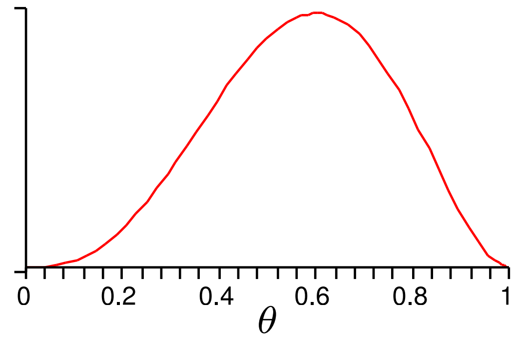
$A_5 = \langle 10110 \rangle$

$lik(\theta) \propto \theta^3 \, (1 - \theta)^2$

Estimation of $\theta \in [0, 1]$:

(likelihood-based inference: $\widehat{\theta}_{ML} = 0.6$)

Example: iid Bernoulli



$A_5 = \langle 10110 \rangle$

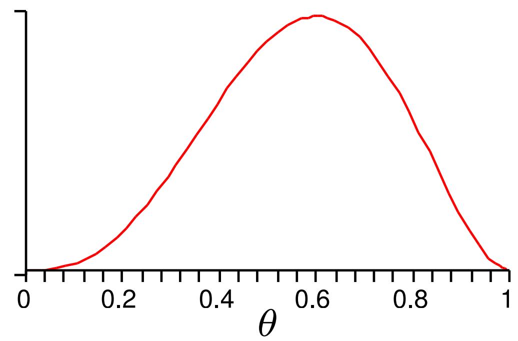
$lik(\theta) \propto \theta^3 (1-\theta)^2$

Estimation of $\theta \in [0,1]$:

(likelihood-based inference: $\widehat{\theta}_{ML} = 0.6$)

$$L(\theta, d) = \begin{cases} 0 & \text{if } |\theta - d| < \varepsilon \\ 1 & \text{if } |\theta - d| \geq \varepsilon \end{cases} \quad \Rightarrow \quad |\widehat{\theta}_{ML} - d_{MPL}| < \varepsilon$$

Example: iid Bernoulli



$A_5 = \langle 10110 \rangle$

$lik(\theta) \propto \theta^3 (1-\theta)^2$

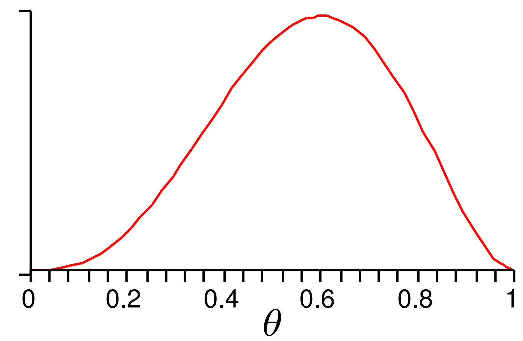
Estimation of $\theta \in [0,1]$:

(likelihood-based inference: $\widehat{\theta}_{ML} = 0.6$)

$$L(\theta, d) = \begin{cases} 0 & \text{if } |\theta - d| < \varepsilon \\ 1 & \text{if } |\theta - d| \geq \varepsilon \end{cases} \quad \Rightarrow \quad |\widehat{\theta}_{ML} - d_{MPL}| < \varepsilon$$

$$L(\theta, d) = |\theta - d| \quad \Rightarrow \quad d_{MPL} \approx 0.588$$

Example: iid Bernoulli



$A_5 = \langle 10110 \rangle$

$lik(\theta) \propto \theta^3 (1-\theta)^2$

Estimation of $\theta \in [0, 1]$:

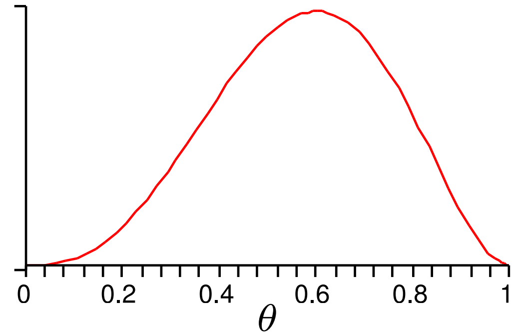
(likelihood-based inference: $\widehat{\theta}_{ML} = 0.6$)

$$L(\theta, d) = \begin{cases} 0 & \text{if } |\theta - d| < \varepsilon \\ 1 & \text{if } |\theta - d| \geq \varepsilon \end{cases} \quad \Rightarrow \quad |\widehat{\theta}_{ML} - d_{MPL}| < \varepsilon$$

$$L(\theta, d) = |\theta - d| \quad \Rightarrow \quad d_{MPL} \approx 0.588$$

$$L(\theta, d) = \begin{cases} 5\,|d - \theta| & \text{if } d \leq \theta \\ |d - \theta| & \text{if } d \geq \theta \end{cases} \quad \Rightarrow \quad d_{MPL} \approx 0.736$$

Example: iid Bernoulli



$A_5 = \langle 10110 \rangle$

$lik(\theta) \propto \theta^3 (1-\theta)^2$

Estimation of $\theta \in [0,1]$:
(likelihood-based inference: $\widehat{\theta}_{ML} = 0.6$)

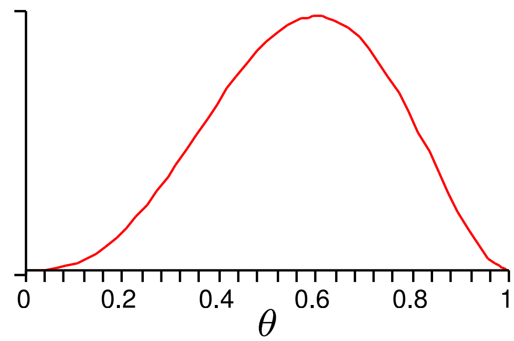
$$L(\theta, d) = \begin{cases} 0 & \text{if } |\theta - d| < \varepsilon \\ 1 & \text{if } |\theta - d| \geq \varepsilon \end{cases} \quad \Rightarrow \quad |\widehat{\theta}_{ML} - d_{MPL}| < \varepsilon$$

$$L(\theta, d) = |\theta - d| \quad \Rightarrow \quad d_{MPL} \approx 0.588$$

$$L(\theta, d) = \begin{cases} 5\,|d - \theta| & \text{if } d \leq \theta \\ |d - \theta| & \text{if } d \geq \theta \end{cases} \quad \Rightarrow \quad d_{MPL} \approx 0.736$$

Test of $\mathcal{H}_0$ against $\mathcal{H}_1$:
(likelihood-based inference: reject $\mathcal{H}_0$ if $LR(\mathcal{H}_0)$ is sufficiently small)

Example: iid Bernoulli



$A_5 = \langle 10110 \rangle$

$lik(\theta) \propto \theta^3 (1-\theta)^2$

Estimation of $\theta \in [0,1]$:

(likelihood-based inference: $\widehat{\theta}_{ML} = 0.6$)

$$L(\theta, d) = \begin{cases} 0 & \text{if } |\theta - d| < \varepsilon \\ 1 & \text{if } |\theta - d| \geq \varepsilon \end{cases} \quad \Rightarrow \quad |\widehat{\theta}_{ML} - d_{MPL}| < \varepsilon$$

$$L(\theta, d) = |\theta - d| \quad \Rightarrow \quad d_{MPL} \approx 0.588$$

$$L(\theta, d) = \begin{cases} 5\,|d - \theta| & \text{if } d \leq \theta \\ |d - \theta| & \text{if } d \geq \theta \end{cases} \quad \Rightarrow \quad d_{MPL} \approx 0.736$$

Test of $\mathcal{H}_0$ against $\mathcal{H}_1$:

(likelihood-based inference: reject $\mathcal{H}_0$ if $LR(\mathcal{H}_0)$ is sufficiently small)

$L_r = I_{\mathcal{H}_0}$ and $L_{nr} = c\, I_{\mathcal{H}_1} \quad \Rightarrow \quad (d_{MPL} = r \;\Leftrightarrow\; LR(\mathcal{H}_0) < c)$

Likelihood-Based Decision Functions

- sample space $(\Omega, \mathcal{A})$ (measurable space)
- set $\mathcal{P}$ of considered statistical models (probability measures on $(\Omega, \mathcal{A})$)
- set $\mathcal{D}$ of possible decisions
- loss function $L : \mathcal{P} \times \mathcal{D} \to [0, \infty)$

Likelihood-Based Decision Functions

- sample space $(\Omega, \mathcal{A})$ (measurable space)
- set $\mathcal{P}$ of considered statistical models (probability measures on $(\Omega, \mathcal{A})$)
- set $\mathcal{D}$ of possible decisions
- loss function $L : \mathcal{P} \times \mathcal{D} \to [0, \infty)$

If $A \in \mathcal{A}$ is observed, define $lik : P \mapsto P(A)$ and select the decision d_{mF} minimizing $F(L(\cdot, d), lik)$.

Likelihood-Based Decision Functions

- sample space $(\Omega, \mathcal{A})$ (measurable space)
- set $\mathcal{P}$ of considered statistical models (probability measures on $(\Omega, \mathcal{A})$)
- set $\mathcal{D}$ of possible decisions
- loss function $L : \mathcal{P} \times \mathcal{D} \rightarrow [0, \infty)$

If $A \in \mathcal{A}$ is observed, define $lik : P \mapsto P(A)$ and select the decision d_{mF} minimizing $F(L(\cdot, d), lik)$.

If $X : \Omega \rightarrow \mathcal{X}$ is a random object, we can select a decision d_{mF} for every possible realization x of X (by setting $A = \{X = x\}$), obtaining a decision function $\delta_{mF} : \mathcal{X} \rightarrow \mathcal{D}$.

δ_{mF} can be compared with other decision functions δ on the basis of the pre-data expected loss $R_\delta : P \mapsto E_P[L(P, \delta(X))]$.

Example: Mixed Effects Models

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

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad \forall\, i, j \in \{1, 2, 3\}$$

Example: Mixed Effects Models

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

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad \forall\, i, j \in \{1, 2, 3\}$$

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: Mixed Effects Models

The estimates $\widehat{v_e}$ and $\widehat{v_a}$ of the 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: Mixed Effects Models

The estimates $\widehat{v_e}$ and $\widehat{v_a}$ of the 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

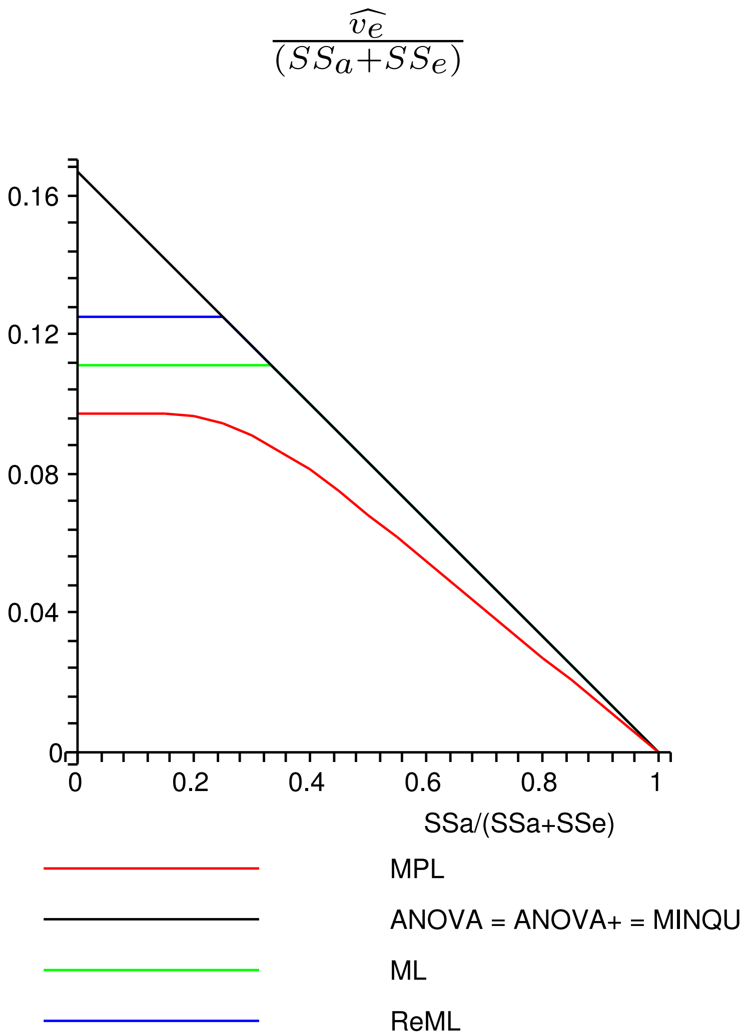
$$\bar{x}_{i\cdot} = \frac{1}{3}\sum_{j=1}^{3} x_{ij} , \quad \bar{x}_{\cdot\cdot} = \frac{1}{9}\sum_{i=1}^{3}\sum_{j=1}^{3} x_{ij} ,$$

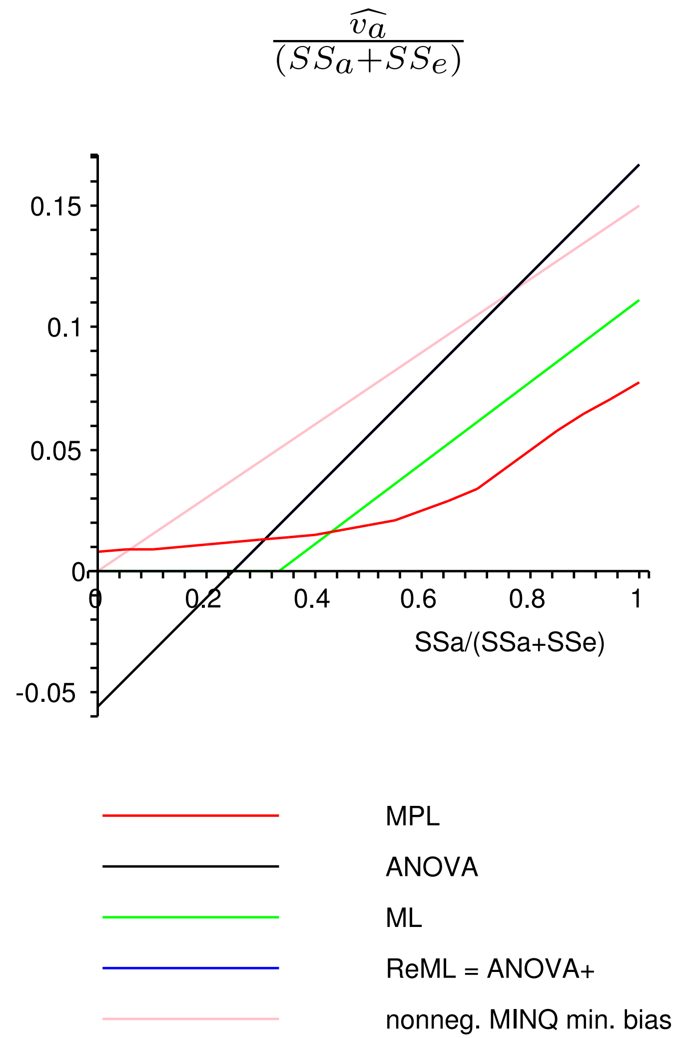
$$\frac{SS_e}{v_e} \sim \chi^2_6 \quad \text{and} \quad \frac{\frac{1}{3}\,SS_a}{v_a + \frac{1}{3}\,v_e} \sim \chi^2_2 .$$

The considered loss functions are

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

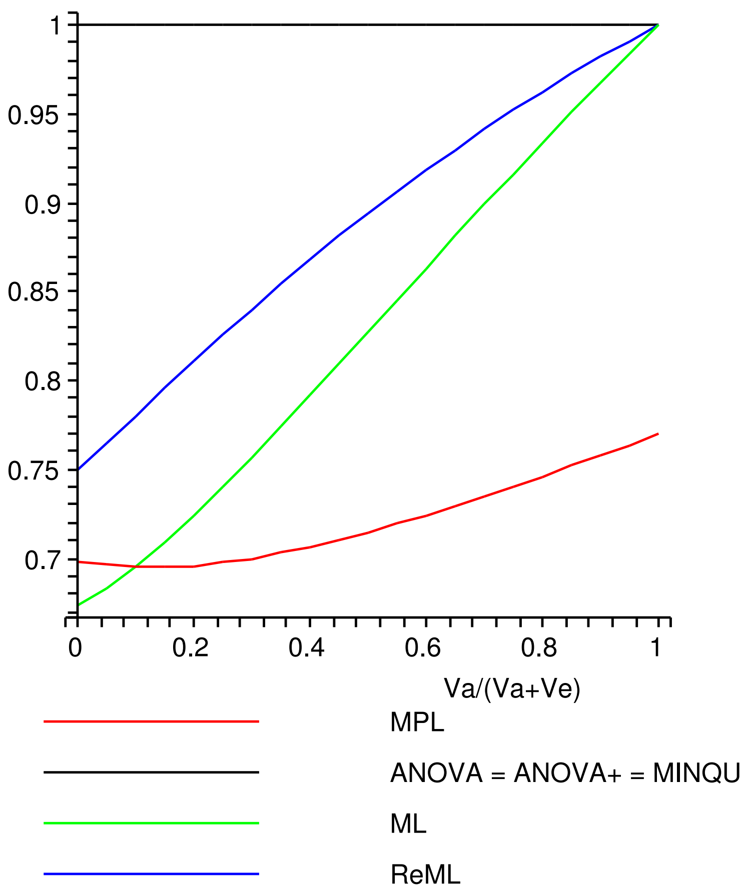
Example: Mixed Effects Models



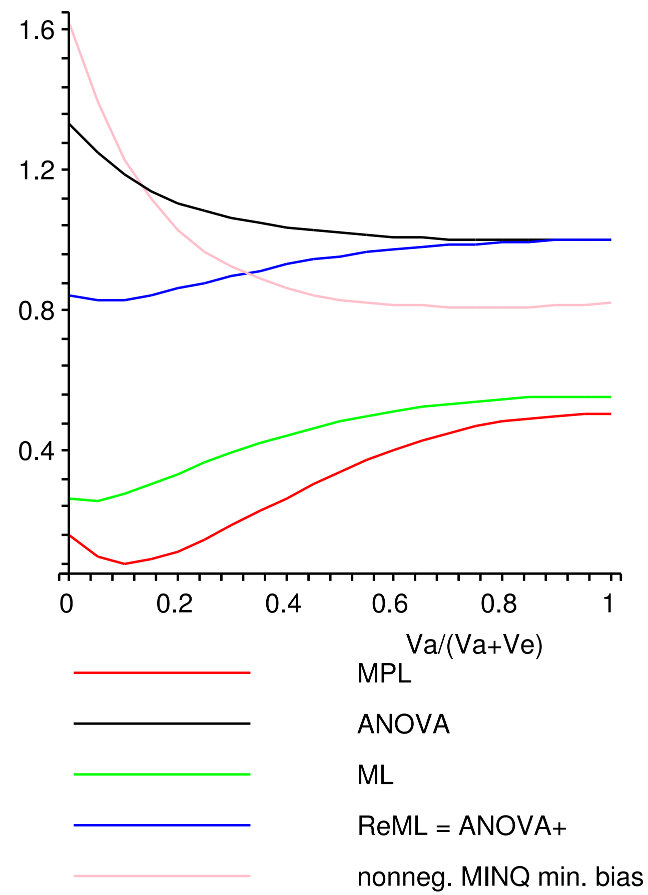


Example: Mixed Effects Models

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



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



Conditional Decision Criteria

Necessary properties of $F(L_d, lik)$?

Conditional Decision Criteria

Necessary properties of $F(L_d, lik)$?

Since lik and $c\,lik$ are equivalent (for $c > 0$), it suffices to consider $F(L_d, \overline{lik})$ (where $\overline{lik} \propto lik$ and $\sup \overline{lik} = 1$).

$F(L_d, \overline{lik})$ is an evaluation of L_d on the basis of $\overline{lik}$.

Conditional Decision Criteria

Necessary properties of $F(L_d, lik)$?

Since lik and $c\,lik$ are equivalent (for $c > 0$), it suffices to consider $F(L_d, \overline{lik})$ (where $\overline{lik} \propto lik$ and $\sup \overline{lik} = 1$).

$F(L_d, \overline{lik})$ is an evaluation of L_d on the basis of $\overline{lik}$.

Calibration: $F(c, \overline{lik}) = c$

Monotonicity: $L_d \geq L_{d'} \;\Rightarrow\; F(L_d, \overline{lik}) \geq F(L_{d'}, \overline{lik})$

Scale Invariance: $F(L_d, \overline{lik}) \geq F(L_{d'}, \overline{lik}) \Rightarrow F(c\,L_d, \overline{lik}) \geq F(c\,L_{d'}, \overline{lik})$

Conditional Decision Criteria

Necessary properties of $F(L_d, lik)$?

Since lik and $c\,lik$ are equivalent (for $c > 0$), it suffices to consider $F(L_d, \overline{lik})$ (where $\overline{lik} \propto lik$ and $\sup \overline{lik} = 1$).

$F(L_d, \overline{lik})$ is an evaluation of L_d on the basis of $\overline{lik}$.

Calibration: $F(c, \overline{lik}) = c$

Monotonicity: $L_d \geq L_{d'} \;\Rightarrow\; F(L_d, \overline{lik}) \geq F(L_{d'}, \overline{lik})$

Scale Invariance: $F(L_d, \overline{lik}) \geq F(L_{d'}, \overline{lik}) \Rightarrow F(c\,L_d, \overline{lik}) \geq F(c\,L_{d'}, \overline{lik})$

If Calibration is valid, Scale Invariance is equivalent to:

Homogeneity: $F(c\,L_d, \overline{lik}) = c\,F(L_d, \overline{lik})$

Maximum Likelihood Decision Criterion

$$F_{ML}(L_d, \overline{lik}) = \sup_{\lim \overline{lik}(P_n)=1} \limsup L_d(P_n)$$

The conditional decision criterion based on F_{ML} generalizes the idea: select the decision which is optimal under the model $\widehat{P}_{ML}$. In particular, in estimation problems the selected decision is the maximum likelihood estimate (if it exists).

Maximum Likelihood Decision Criterion

$$F_{ML}(L_d, \overline{lik}) = \sup_{\lim \overline{lik}(P_n)=1} \limsup L_d(P_n)$$

The conditional decision criterion based on F_{ML} generalizes the idea: select the decision which is optimal under the model $\widehat{P}_{ML}$. In particular, in estimation problems the selected decision is the maximum likelihood estimate (if it exists).

F_{ML} satisfies the 3 necessary properties Calibration, Monotonicity and Homogeneity. The supremum in the definition of F_{ML} has been chosen in order to satisfy:

Conditional Minimax: $F(L_d, I_{\mathcal{H}}) = \sup I_{\mathcal{H}} \, L_d$

Generalization of Inference Methods

In order to generalize the likelihood-based inference methods, it is necessary that F satisfies $F(I_{\mathcal{H}}, \overline{lik}) = f\left(\frac{\sup I_{\mathcal{H}}\, lik}{\sup I_{\mathcal{P}\setminus\mathcal{H}}\, lik}\right)$, where $f : [0, \infty] \to [0, \infty]$.

Generalization of Inference Methods

In order to generalize the likelihood-based inference methods, it is necessary that F satisfies $F(I_{\mathcal{H}}, \overline{lik}) = f\left(\frac{\sup I_{\mathcal{H}}\, lik}{\sup I_{\mathcal{P}\setminus\mathcal{H}}\, lik}\right)$, where $f : [0, \infty] \to [0, \infty]$.

From Monotonicity it follows that f is increasing, and from Calibration it follows that $f(0) = 0$ and $f(\infty) = 1$.
To generalize the maximum likelihood estimator, f must be strictly increasing in 1 (that is, f can not be constant on any neighborhood of 1).
To generalize the tests and confidence regions based on the likelihood ratio statistic, f must be continuous on $[0, \infty)$ and strictly increasing on $[0, 1]$.

Generalization of Inference Methods

In order to generalize the likelihood-based inference methods, it is necessary that F satisfies $F(I_{\mathcal{H}}, \overline{lik}) = f\left(\frac{\sup I_{\mathcal{H}}\, lik}{\sup I_{\mathcal{P}\setminus\mathcal{H}}\, lik}\right)$, where $f : [0,\infty] \to [0,\infty]$.

From Monotonicity it follows that f is increasing, and from Calibration it follows that $f(0) = 0$ and $f(\infty) = 1$.
To generalize the maximum likelihood estimator, f must be strictly increasing in 1 (that is, f can not be constant on any neighborhood of 1).
To generalize the tests and confidence regions based on the likelihood ratio statistic, f must be continuous on $[0,\infty)$ and strictly increasing on $[0,1]$.

$$F_{ML}(I_{\mathcal{H}}, \overline{lik}) = f_{ML}\left(\frac{\sup I_{\mathcal{H}}\, lik}{\sup I_{\mathcal{P}\setminus\mathcal{H}}\, lik}\right) \text{ with } f_{ML}(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}.$$

Thus F_{ML} generalizes the maximum likelihood estimator, but it does not generalize the tests and confidence regions based on the likelihood ratio statistic.

Likelihood Ratio Statistic

If $F(I_{\mathcal{H}}, \overline{lik}) = LR(\mathcal{H})$ for all $\mathcal{H} \subseteq \mathcal{P}$, F generalizes the likelihood-based inference methods, since $LR(\mathcal{H}) = \sup I_{\mathcal{H}} \overline{lik} = f_{LR}\left(\frac{\sup I_{\mathcal{H}} lik}{\sup I_{\mathcal{P}\setminus\mathcal{H}} lik}\right)$ with $f_{LR}(x) = \begin{cases} x & \text{if } x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$.

Likelihood Ratio Statistic

If $F(I_{\mathcal{H}}, \overline{lik}) = LR(\mathcal{H})$ for all $\mathcal{H} \subseteq \mathcal{P}$, F generalizes the likelihood-based inference methods, since $LR(\mathcal{H}) = \sup I_{\mathcal{H}} \overline{lik} = f_{LR}\left(\frac{\sup I_{\mathcal{H}} lik}{\sup I_{\mathcal{P}\setminus\mathcal{H}} lik}\right)$ with $f_{LR}(x) = \begin{cases} x & \text{if } x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$.

The likelihood ratio statistic LR is a **possibility measure** (completely maxitive measure) on $\mathcal{P}$:

$$LR(\mathcal{P}) = 1 \quad \text{and} \quad LR\left(\bigcup_{j \in J} \mathcal{H}_j\right) = \sup_{j \in J} LR(\mathcal{H}_j) \ .$$

Subadditive Integrals

Monotonicity: $L_d \geq L_{d'} \;\Rightarrow\; F(L_d, \overline{lik}) \geq F(L_{d'}, \overline{lik})$

Homogeneity: $F(c\, L_d, \overline{lik}) = c\, F(L_d, \overline{lik})$

Indicator Property: $F(I_{\mathcal{H}}, \overline{lik}) = LR(\mathcal{H})$

If F satisfies this 3 properties (Calibration follows from Homogeneity and Indicator Property), it can be considered an **integral** with respect to the possibility measure LR:

$$\boxed{F(L_d, \overline{lik}) = \int L_d \,\mathrm{d}LR} \,.$$

Subadditive Integrals

Monotonicity: $L_d \geq L_{d'} \;\Rightarrow\; F(L_d, \overline{lik}) \geq F(L_{d'}, \overline{lik})$

Homogeneity: $F(c\, L_d, \overline{lik}) = c\, F(L_d, \overline{lik})$

Indicator Property: $F(I_{\mathcal{H}}, \overline{lik}) = LR(\mathcal{H})$

If F satisfies this 3 properties (Calibration follows from Homogeneity and Indicator Property), it can be considered an **integral** with respect to the possibility measure LR:

$$\boxed{F(L_d, \overline{lik}) = \int L_d \,\mathrm{d}LR}\,.$$

If $\mathcal{H} \cap \mathcal{H}' = \emptyset$, then $\int (I_{\mathcal{H}} + I_{\mathcal{H}'}) \,\mathrm{d}LR = \max\{\int I_{\mathcal{H}} \,\mathrm{d}LR, \int I_{\mathcal{H}'} \,\mathrm{d}LR\}$.
Thus the integral F can not be additive, but we can require:

Subadditivity: $F(L_d + L_{d'}, \overline{lik}) \leq F(L_d, \overline{lik}) + F(L_{d'}, \overline{lik})$

Asymptotic Optimality

- $X_1, X_2, \ldots$ iid (discrete) random objects
- $\mathcal{P}$ and L satisfy some regularity conditions
- d_n is the conditional decision based on a subadditive integral F after having observed $X_1, \ldots, X_n$

$$\Rightarrow \; L(P, d_n) \to \inf L(P, \cdot) \;\; P\text{-a.s, for all } P \in \mathcal{P}$$

Asymptotic Optimality

- $X_1, X_2, \dots$ iid (discrete) random objects
- $\mathcal{P}$ and L satisfy some regularity conditions
- d_n is the conditional decision based on a subadditive integral F after having observed $X_1, \dots, X_n$

$$\Rightarrow \;\; L(P, d_n) \to \inf L(P, \cdot) \;\; P\text{-a.s, for all } P \in \mathcal{P}$$

The properties of F sufficient for this asymptotic optimality are much weaker than those defining the subadditive integrals. In particular, F_{ML} satisfies them.

In estimation problems with symmetric loss functions, asymptotic efficiency can be obtained, but stronger properties are necessary.

Equivariance

An integral can be expected to satisfy $\int (L_d \circ T)\,\mathrm{d}LR = \int L_d\,\mathrm{d}(LR \circ T^{-1})$, when $T : \mathcal{P} \to \mathcal{T}$ and $L_d : \mathcal{T} \to [0, \infty)$.

If the integral F satisfies this property, it follows in particular that the decision function is equivariant when the decision problem is invariant (to have equivariance it suffices that the property is valid for bijections $T : \mathcal{P} \to \mathcal{P}'$).

Equivariance

An integral can be expected to satisfy $\int (L_d \circ T)\,\mathrm{d}LR = \int L_d\,\mathrm{d}(LR \circ T^{-1})$, when $T : \mathcal{P} \to \mathcal{T}$ and $L_d : \mathcal{T} \to [0, \infty)$.

If the integral F satisfies this property, it follows in particular that the decision function is equivariant when the decision problem is invariant (to have equivariance it suffices that the property is valid for bijections $T : \mathcal{P} \to \mathcal{P}'$).

The above transformation property is valid if and only if the integral depends only on $LR \circ L_d^{-1} : x \mapsto LR\{L_d = x\}$.

If in addition the integral F depends only on $x \mapsto LR\{L_d > x\}$, it satisfies also the Conditional Minimax property.

Choquet and Shilkret Integrals

A subadditive integral depending only on $x \mapsto LR\{L_d > x\}$ satisfies all the properties considered. Examples are:

Choquet integral: $\displaystyle\int L_d \, \mathrm{d}LR = \int_0^\infty LR\{L_d > x\} \, \mathrm{d}x$

Shilkret integral:
$$\int L_d \, \mathrm{d}LR = \sup_{x>0} x \, LR\{L_d > x\} = \\ = \sup \, lik \, L_d = F_{MPL}(L_d, lik)$$

Choquet and Shilkret Integrals

A subadditive integral depending only on $x \mapsto LR\{L_d > x\}$ satisfies all the properties considered. Examples are:

Choquet integral: $$\int L_d \,\mathrm{d}LR = \int_0^\infty LR\{L_d > x\}\,\mathrm{d}x$$

Shilkret integral: $$\begin{aligned} \int L_d \,\mathrm{d}LR &= \sup_{x>0} x\, LR\{L_d > x\} = \\ &= \sup\, lik\, L_d = F_{MPL}(L_d, lik) \end{aligned}$$

The Choquet integral has the advantage of being translation invariant: $\int (L_d + c)\,\mathrm{d}LR = \int L_d \,\mathrm{d}LR + c$.

The Shilkret integral has the advantage of being simple and intuitive.

Likelihood-Based Statistical Decisions

The likelihood-based decision criteria, and in particular the MPL criterion:

- are asymptotic optimal (if some regularity conditions are satisfied);
- are general and simple (and therefore widely applicable);
- are intuitive (in particular conditional and parametrization invariant).