Likelihood-based evaluations

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

January 8, 2008

hierarchical model

A hierarchical model $\overset{LR}{\mathcal{P}}$ consists of a set $\mathcal{P}$ of probability measures on a measurable space $(\Omega, \mathcal{A})$ and of a normalized possibility measure LR on $\mathcal{P}$.

hierarchical model

A hierarchical model $\overset{LR}{\mathcal{P}}$ consists of a set $\mathcal{P}$ of probability measures on a measurable space $(\Omega, \mathcal{A})$ and of a normalized possibility measure LR on $\mathcal{P}$.

Let *lik* be the density function of LR:

$$lik : P \mapsto LR\{P\} \text{ on } \mathcal{P}, \quad \text{and} \quad LR(\mathcal{H}) = \sup_{P \in \mathcal{H}} lik(P) \text{ for all } \mathcal{H} \subseteq \mathcal{P}.$$

hierarchical model

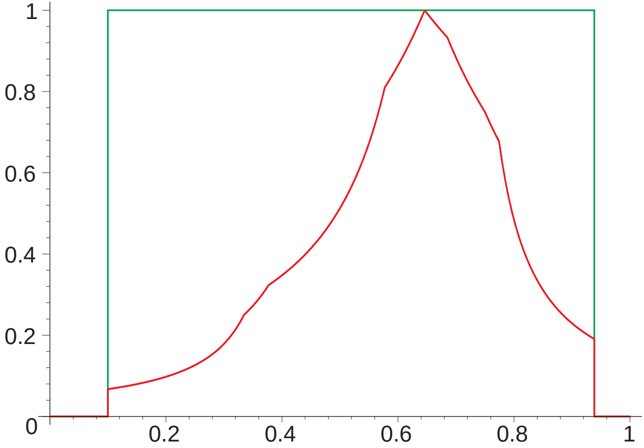
A hierarchical model $\overset{LR}{\mathcal{P}}$ consists of a set $\mathcal{P}$ of probability measures on a measurable space $(\Omega, \mathcal{A})$ and of a normalized possibility measure LR on $\mathcal{P}$.

Let *lik* be the density function of LR:

$$lik : P \mapsto LR\{P\} \text{ on } \mathcal{P}, \quad \text{and} \quad LR(\mathcal{H}) = \sup_{P \in \mathcal{H}} lik(P) \text{ for all } \mathcal{H} \subseteq \mathcal{P}.$$

Each element of $\mathcal{P}$ is interpreted as a statistical model of the reality under consideration, and the (normalized) likelihood function *lik* is interpreted as a measure of the relative plausibility of the statistical models in $\mathcal{P}$.

a "fuzzy probability" from an example by De Cooman and Zaffalon (2004)



evaluation of a "fuzzy number"

The uncertain knowledge about the value of a function $g : \mathcal{P} \mapsto \mathcal{G}$ is described by the (normalized) possibility measure $LR \circ g^{-1}$ induced on $\mathcal{G}$; if $\mathcal{G} = \mathbb{R}$, then $LR \circ g^{-1}$ corresponds to a "fuzzy number".

evaluation of a "fuzzy number"

The uncertain knowledge about the value of a function $g : \mathcal{P} \mapsto \mathcal{G}$ is described by the (normalized) possibility measure $LR \circ g^{-1}$ induced on $\mathcal{G}$; if $\mathcal{G} = \mathbb{R}$, then $LR \circ g^{-1}$ corresponds to a "fuzzy number".

Examples of functions $\mathcal{P} \to \mathbb{R}$:

- $g : P \mapsto E_P(X)$

evaluation of a "fuzzy number"

The uncertain knowledge about the value of a function $g : \mathcal{P} \mapsto \mathcal{G}$ is described by the (normalized) possibility measure $LR \circ g^{-1}$ induced on $\mathcal{G}$; if $\mathcal{G} = \mathbb{R}$, then $LR \circ g^{-1}$ corresponds to a "fuzzy number".

Examples of functions $\mathcal{P} \to \mathbb{R}$:

- $g : P \mapsto E_P(X)$
- $p : P \mapsto P(A)$

evaluation of a “fuzzy number”

The uncertain knowledge about the value of a function $g : \mathcal{P} \mapsto \mathcal{G}$ is described by the (normalized) possibility measure $LR \circ g^{-1}$ induced on $\mathcal{G}$; if $\mathcal{G} = \mathbb{R}$, then $LR \circ g^{-1}$ corresponds to a “fuzzy number”.

Examples of functions $\mathcal{P} \to \mathbb{R}$:

- $g : P \mapsto E_P(X)$
- $p : P \mapsto P(A)$
- $l_d : P \mapsto L(P, d)$, where $L : \mathcal{P} \times \mathcal{D} \to [0, \infty)$ is a loss function describing a statistical decision problem

evaluation of a "fuzzy number"

The uncertain knowledge about the value of a function $g : \mathcal{P} \mapsto \mathcal{G}$ is described by the (normalized) possibility measure $LR \circ g^{-1}$ induced on $\mathcal{G}$; if $\mathcal{G} = \mathbb{R}$, then $LR \circ g^{-1}$ corresponds to a "fuzzy number".

Examples of functions $\mathcal{P} \to \mathbb{R}$:

- $g : P \mapsto E_P(X)$
- $p : P \mapsto P(A)$
- $l_d : P \mapsto L(P, d)$, where $L : \mathcal{P} \times \mathcal{D} \to [0, \infty)$ is a loss function describing a statistical decision problem

  idea behind my thesis: select the decision $d \in \mathcal{D}$ minimizing the evaluation $V(l_d, lik)$ of the corresponding "fuzzy loss" $LR \circ l_d^{-1}$

examples of likelihood-based (upper) evaluations

- Likelihood-based Region Minimax, with $\beta \in (0, 1)$:

$$V_{\mathrm{LRM}_\beta}(l_d, lik) = \sup_{P \in \mathcal{P}:\, lik(P) > \beta} l_d(P)$$

examples of likelihood-based (upper) evaluations

- Likelihood-based Region Minimax, with $\beta \in (0, 1)$:

$$V_{\mathrm{LRM}_\beta}(l_d, lik) = \sup_{P \in \mathcal{P}:\, lik(P) > \beta} l_d(P)$$

- Maximum Likelihood Decision:

$$V_{\mathrm{MLD}}(l_d, lik) = \lim_{\beta \uparrow 1} V_{\mathrm{LRM}_\beta}(l_d, lik) \qquad (= l_d(\hat{P}_{ML}))$$

examples of likelihood-based (upper) evaluations

- Likelihood-based Region Minimax, with $\beta \in (0,1)$:

$$V_{\mathrm{LRM}_\beta}(l_d, lik) = \sup_{P \in \mathcal{P}:\, lik(P) > \beta} l_d(P)$$

- Maximum Likelihood Decision:

$$V_{\mathrm{MLD}}(l_d, lik) = \lim_{\beta \uparrow 1} V_{\mathrm{LRM}_\beta}(l_d, lik) \qquad (= l_d(\hat{P}_{ML}))$$

- Essential Minimax:

$$V_{\mathrm{EM}}(l_d, lik) = \lim_{\beta \downarrow 0} V_{\mathrm{LRM}_\beta}(l_d, lik) = \sup_{P \in \mathcal{P}:\, lik(P) > 0} l_d(P)$$

examples of likelihood-based (upper) evaluations

- Likelihood-based Region Minimax, with $\beta \in (0,1)$:

$$V_{\mathrm{LRM}_\beta}(l_d, lik) = \sup_{P \in \mathcal{P}:\, lik(P) > \beta} l_d(P)$$

- Maximum Likelihood Decision:

$$V_{\mathrm{MLD}}(l_d, lik) = \lim_{\beta \uparrow 1} V_{\mathrm{LRM}_\beta}(l_d, lik) \qquad (= l_d(\hat{P}_{ML}))$$

- Essential Minimax:

$$V_{\mathrm{EM}}(l_d, lik) = \lim_{\beta \downarrow 0} V_{\mathrm{LRM}_\beta}(l_d, lik) = \sup_{P \in \mathcal{P}:\, lik(P) > 0} l_d(P)$$

- Minimax Plausibility-weighted Loss:

$$V_{\mathrm{MPL}}(l_d, lik) = \sup_{P \in \mathcal{P}} lik(P)\, l_d(P)$$

integral representations

$$V_{\mathrm{MPL}}(l_d, lik) = \sup_{P \in \mathcal{P}} lik(P)\, l_d(P) = \int^{\mathsf{S}} l_d \,\mathrm{d}LR \qquad \text{(Shilkret integral)}$$

integral representations

$$V_{\mathsf{MPL}}(l_d, lik) = \sup_{P \in \mathcal{P}} lik(P)\, l_d(P) = \int^{\mathsf{S}} l_d \,\mathrm{d}LR \qquad \text{(Shilkret integral)}$$

$$V_{\mathsf{MPL}^*}(l_d, lik) = \int^{\mathsf{C}} l_d \,\mathrm{d}LR = \int_0^{\infty} LR\{l_d \geq x\} \,\mathrm{d}x \qquad \text{(Choquet integral)}$$

integral representations

$$V_{\mathrm{MPL}}(l_d, lik) = \sup_{P \in \mathcal{P}} lik(P)\, l_d(P) = \int^{\mathsf{S}} l_d \,\mathrm{d}LR \qquad \text{(Shilkret integral)}$$

$$V_{\mathrm{MPL}^*}(l_d, lik) = \int^{\mathsf{C}} l_d \,\mathrm{d}LR = \int_0^\infty LR\{l_d \geq x\}\,\mathrm{d}x \qquad \text{(Choquet integral)}$$

$$V_{\mathrm{LRM}_\beta}(l_d, lik) = \int^{\mathsf{S}} l_d \,\mathrm{d}(\mathrm{I}_{(\beta,1]} \circ LR) = \int^{\mathsf{C}} l_d \,\mathrm{d}(\mathrm{I}_{(\beta,1]} \circ LR)$$

$$V_{\mathrm{MLD}}(l_d, lik) = \int^{\mathsf{S}} l_d \,\mathrm{d}(\mathrm{I}_{\{1\}} \circ LR) = \int^{\mathsf{C}} l_d \,\mathrm{d}(\mathrm{I}_{\{1\}} \circ LR)$$

$$V_{\mathrm{EM}}(l_d, lik) = \int^{\mathsf{S}} l_d \,\mathrm{d}(\mathrm{I}_{(0,1]} \circ LR) = \int^{\mathsf{C}} l_d \,\mathrm{d}(\mathrm{I}_{(0,1]} \circ LR)$$

integral representations

$$V_{\mathrm{MPL}}(l_d, lik) = \sup_{P \in \mathcal{P}} lik(P)\, l_d(P) = \int^{\mathsf{S}} l_d \, \mathrm{d}LR \qquad \text{(Shilkret integral)}$$

$$V_{\mathrm{MPL}^*}(l_d, lik) = \int^{\mathsf{C}} l_d \, \mathrm{d}LR = \int_0^{\infty} LR\{l_d \geq x\} \, \mathrm{d}x \qquad \text{(Choquet integral)}$$

$$V_{\mathrm{LRM}_\beta}(l_d, lik) = \int^{\mathsf{S}} l_d \, \mathrm{d}(\mathrm{I}_{(\beta,1]} \circ LR) = \int^{\mathsf{C}} l_d \, \mathrm{d}(\mathrm{I}_{(\beta,1]} \circ LR)$$

$$V_{\mathrm{MLD}}(l_d, lik) = \int^{\mathsf{S}} l_d \, \mathrm{d}(\mathrm{I}_{\{1\}} \circ LR) = \int^{\mathsf{C}} l_d \, \mathrm{d}(\mathrm{I}_{\{1\}} \circ LR)$$

$$V_{\mathrm{EM}}(l_d, lik) = \int^{\mathsf{S}} l_d \, \mathrm{d}(\mathrm{I}_{(0,1]} \circ LR) = \int^{\mathsf{C}} l_d \, \mathrm{d}(\mathrm{I}_{(0,1]} \circ LR)$$

In general, if $\delta : [0,1] \to [0,1]$ is a nondecreasing function with $\delta(0) = 0$ and $\delta(1) = 1$, then the decision criterion based on

$$V(l_d, lik) = \int^{\mathsf{S}} l_d \, \mathrm{d}(\delta \circ LR) \quad \text{or} \quad V(l_d, lik) = \int^{\mathsf{C}} l_d \, \mathrm{d}(\delta \circ LR)$$

leads to the usual likelihood-based inference methods (when applied to some standard form of the corresponding decision problems) if and only if δ is bijective.

likelihood-based decision criteria

A **likelihood-based decision criterion** can be expressed as

$$\text{minimize} \quad V(l_d, lik),$$

where the evaluation $V(l_d, lik)$ depends only on $LR \circ l_d^{-1}$, is calibrated (that is, $V(c, lik) = c$ for all $c \in [0, \infty)$), and satisfies the following two conditions:

likelihood-based decision criteria

A **likelihood-based decision criterion** can be expressed as

$$\text{minimize} \quad V(l_d, lik),$$

where the evaluation $V(l_d, lik)$ depends only on $LR \circ l_d^{-1}$, is calibrated (that is, $V(c, lik) = c$ for all $c \in [0, \infty)$), and satisfies the following two conditions:

- monotonicity:

  $$l_d(P) \leq l_{d'}(P) \text{ for all } P \in \mathcal{P} \quad \Rightarrow \quad V(l_d, lik) \leq V(l_{d'}, lik),$$

likelihood-based decision criteria

A **likelihood-based decision criterion** can be expressed as

$$\text{minimize} \quad V(l_d, lik),$$

where the evaluation $V(l_d, lik)$ depends only on $LR \circ l_d^{-1}$, is calibrated (that is, $V(c, lik) = c$ for all $c \in [0, \infty)$), and satisfies the following two conditions:

- monotonicity:

  $$l_d(P) \leq l_{d'}(P) \text{ for all } P \in \mathcal{P} \quad \Rightarrow \quad V(l_d, lik) \leq V(l_{d'}, lik),$$

- scale invariance: for all $c \in (0, \infty)$

  $$V(l_d, lik) \leq V(l_{d'}, lik) \quad \Rightarrow \quad V(c\, l_d, lik) \leq V(c\, l_{d'}, lik).$$

consistency

A likelihood-based decision criterion is **consistent** if

$$\mathcal{P} = \{P_1, P_2\} \text{ and } \frac{lik(P_1)}{lik(P_2)} \to \infty \quad \Rightarrow \quad V(l_d, lik) \to l_d(P_1).$$

consistency

A likelihood-based decision criterion is **consistent** if

$$\mathcal{P} = \{P_1, P_2\} \text{ and } \frac{lik(P_1)}{lik(P_2)} \to \infty \quad \Rightarrow \quad V(l_d, lik) \to l_d(P_1).$$

The consistency of a likelihood-based decision criterion is necessary and sufficient to guarantee the (strong) asymptotic optimality of the resulting sequences of decisions (under the usual regularity conditions).

consistency

A likelihood-based decision criterion is **consistent** if

$$\mathcal{P} = \{P_1, P_2\} \text{ and } \frac{lik(P_1)}{lik(P_2)} \to \infty \quad \Rightarrow \quad V(l_d, lik) \to l_d(P_1).$$

The consistency of a likelihood-based decision criterion is necessary and sufficient to guarantee the (strong) asymptotic optimality of the resulting sequences of decisions (under the usual regularity conditions).

If $\delta : [0,1] \to [0,1]$ is a nondecreasing function with $\delta(0) = 0$ and $\delta(1) = 1$, then the decision criterion based on

$$V(l_d, lik) = \int^{\mathrm{S}} l_d \, \mathrm{d}(\delta \circ LR) \quad \text{or} \quad V(l_d, lik) = \int^{\mathrm{C}} l_d \, \mathrm{d}(\delta \circ LR)$$

is consistent if and only if δ is continuous at 0.

sure-thing principle

A likelihood-based decision criterion satisfies the **sure-thing principle** if when $\mathcal{P}', \mathcal{P}''$ build a partition of $\mathcal{P}$

$$\left.\begin{array}{l} V(l_d|_{\mathcal{P}'}, lik|_{\mathcal{P}'}) \leq V(l_{d'}|_{\mathcal{P}'}, lik|_{\mathcal{P}'}) \\ V(l_d|_{\mathcal{P}''}, lik|_{\mathcal{P}''}) \leq V(l_{d'}|_{\mathcal{P}''}, lik|_{\mathcal{P}''}) \end{array}\right\} \quad \Rightarrow \quad V(l_d, lik) \leq V(l_{d'}, lik).$$

sure-thing principle

A likelihood-based decision criterion satisfies the **sure-thing principle** if when $\mathcal{P}', \mathcal{P}''$ build a partition of $\mathcal{P}$

$$\left.\begin{array}{l} V(l_d|_{\mathcal{P}'}, lik|_{\mathcal{P}'}) \leq V(l_{d'}|_{\mathcal{P}'}, lik|_{\mathcal{P}'}) \\ V(l_d|_{\mathcal{P}''}, lik|_{\mathcal{P}''}) \leq V(l_{d'}|_{\mathcal{P}''}, lik|_{\mathcal{P}''}) \end{array}\right\} \quad \Rightarrow \quad V(l_d, lik) \leq V(l_{d'}, lik).$$

For bounded functions l_d, the only upper evaluations corresponding to likelihood-based decision criteria satisfying the sure-thing principle are

- $V_{\mathrm{EM}}(l_d, lik) = \int^{\mathrm{S}} l_d \, \mathrm{d}(\mathrm{I}_{(0,1]} \circ LR) = \sup_{P \in \mathcal{P} \colon lik(P) > 0} l_d(P)$,
- $V_{\alpha\text{-MPL}}(l_d, lik) = \int^{\mathrm{S}} l_d \, \mathrm{d}(LR^{\alpha}) = \sup_{P \in \mathcal{P}} lik(P)^{\alpha} \, l_d(P)$,
- $V_{\mathrm{MLD}}(l_d, lik) = \int^{\mathrm{S}} l_d \, \mathrm{d}(\mathrm{I}_{\{1\}} \circ LR) = \lim_{\beta \uparrow 1} \sup_{P \in \mathcal{P} \colon lik(P) > \beta} l_d(P)$,

where $\alpha \in (0, \infty)$ can be interpreted as a parameter expressing the confidence in the information provided by *lik*.

α-MPL decision criteria

A likelihood-based decision criterion satisfying the "sure-thing principle for infinite partitions" is called **decomposable**. The decomposability allows the introduction of a concept of "conditional evaluation" analogous to the concept of conditional expectation.

α-MPL decision criteria

A likelihood-based decision criterion satisfying the "sure-thing principle for infinite partitions" is called **decomposable**. The decomposability allows the introduction of a concept of "conditional evaluation" analogous to the concept of conditional expectation.

The only evaluations corresponding to decomposable, consistent likelihood-based decision criteria are

$$V_{\alpha\text{-MPL}}(l_d, lik) = \int^{\mathrm{S}} l_d \, \mathrm{d}(LR^\alpha) = \sup_{P \in \mathcal{P}} lik(P)^\alpha \, l_d(P), \quad \text{with } \alpha \in (0, \infty).$$

α-MPL decision criteria

A likelihood-based decision criterion satisfying the "sure-thing principle for infinite partitions" is called **decomposable**. The decomposability allows the introduction of a concept of "conditional evaluation" analogous to the concept of conditional expectation.

The only evaluations corresponding to decomposable, consistent likelihood-based decision criteria are

$$V_{\alpha\text{-MPL}}(l_d, lik) = \int^{\mathrm{S}} l_d \, \mathrm{d}(LR^\alpha) = \sup_{P \in \mathcal{P}} lik(P)^\alpha \, l_d(P), \quad \text{with } \alpha \in (0, \infty).$$

But these evaluations do not satisfy the following property:

- location invariance: for all $c \in (0, \infty)$

  $$V(l_d, lik) \leq V(l_{d'}, lik) \quad \Leftrightarrow \quad V(l_d + c, lik) \leq V(l_{d'} + c, lik).$$

convex risk measures

The likelihood-based decision criteria satisfying both the sure-thing principle and the location invariance (such as the MLD criterion) present important discontinuities.

convex risk measures

The likelihood-based decision criteria satisfying both the sure-thing principle and the location invariance (such as the MLD criterion) present important discontinuities.

If we substitute the location invariance for the scale invariance in the definition of likelihood-based decision criterion, then the (lower and upper) evaluations of $g : \mathcal{P} \to \mathbb{R}$ corresponding to the evaluations $V_{\alpha\text{-MPL}}(l_d, lik)$ are

$$\inf_{P \in \mathcal{P}} [g(P) - \alpha \log lik(P)] \quad \text{and} \quad \sup_{P \in \mathcal{P}} [g(P) + \alpha \log lik(P)].$$

convex risk measures

The likelihood-based decision criteria satisfying both the sure-thing principle and the location invariance (such as the MLD criterion) present important discontinuities.

If we substitute the location invariance for the scale invariance in the definition of likelihood-based decision criterion, then the (lower and upper) evaluations of $g : \mathcal{P} \to \mathbb{R}$ corresponding to the evaluations $V_{\alpha\text{-MPL}}(l_d, lik)$ are

$$\inf_{P \in \mathcal{P}} [g(P) - \alpha \log lik(P)] \quad \text{and} \quad \sup_{P \in \mathcal{P}} [g(P) + \alpha \log lik(P)].$$

In particular, if $g : P \mapsto E_P(X)$, then these evaluations are the lower and upper (centered) convex previsions (Pelessoni and Vicig, 2003)

$$\inf_{P \in \mathcal{P}} [E_P(X) - \alpha \log lik(P)] \quad \text{and} \quad \sup_{P \in \mathcal{P}} [E_P(X) + \alpha \log lik(P)],$$

which correspond to a convex risk measure (Föllmer and Schied, 2002).