Continuous updating rules for imprecise probabilities

Marco Cattaneo
Department of Statistics, LMU Munich

WPMSIIP 2013, Lugano, Switzerland
6 September 2013

example

- $X \in \{1, 2, 3\}$

example

- $X \in \{1, 2, 3\}$
- prior assessment: $\underline{P}(X) = \ldots$

example

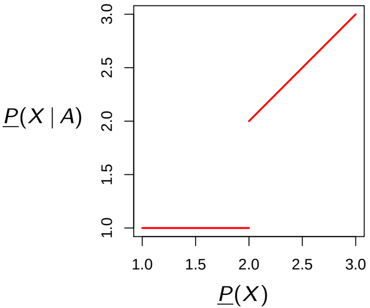
- $X \in \{1, 2, 3\}$
- prior assessment: $\underline{P}(X) = \ldots$
- observation: $A = \{X \neq 2\}$

example

- $X \in \{1, 2, 3\}$
- prior assessment: $\underline{P}(X) = \dots$
- observation: $A = \{X \neq 2\}$
- updated prevision: $\underline{P}(X \mid A) = \dots$

example

- $X \in \{1, 2, 3\}$
- prior assessment: $\underline{P}(X) = \ldots$
- observation: $A = \{X \neq 2\}$
- updated prevision: $\underline{P}(X \mid A) = \ldots$
- natural/regular extension:



updating

- likelihood of linear previsions: $lik(P) \propto P(A)$

updating

- likelihood of linear previsions: $\mathit{lik}(P) \propto P(A)$
- **Bayesian updating**: $P \mapsto P(\,\cdot \mid A)$ when $\mathit{lik}(P) \neq 0$

updating

- likelihood of linear previsions: $lik(P) \propto P(A)$
- **Bayesian updating**: $P \mapsto P(\cdot \mid A)$ when $lik(P) \neq 0$
- **regular updating**: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P}) \,:\, P(A) > 0} P(\cdot \mid A)$ when $lik \not\equiv 0$ on $\mathcal{M}(\underline{P})$

updating

- likelihood of linear previsions: $lik(P) \propto P(A)$
- **Bayesian updating**: $P \mapsto P(\cdot \mid A)$ when $lik(P) \neq 0$
- **regular updating**: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(A)>0} P(\cdot \mid A)$ when $lik \not\equiv 0$ on $\mathcal{M}(\underline{P})$
- the discontinuities are “caused” by linear previsions $P \in \mathcal{M}(\underline{P})$ with arbitrarily small likelihood (i.e., “almost contradicted” by the data)

updating

- likelihood of linear previsions: $\mathit{lik}(P) \propto P(A)$
- **Bayesian updating**: $P \mapsto P(\,\cdot\,|\,A)$ when $\mathit{lik}(P) \neq 0$
- **regular updating**: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(A)>0} P(\,\cdot\,|\,A)$ when $\mathit{lik} \not\equiv 0$ on $\mathcal{M}(\underline{P})$
- the discontinuities are "caused" by linear previsions $P \in \mathcal{M}(\underline{P})$ with arbitrarily small likelihood (i.e., "almost contradicted" by the data)
- $\boldsymbol{\beta}$**-cut updating**, with $\beta \in (0,1)$: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(A)>\beta\,\overline{P}(A)} P(\,\cdot\,|\,A)$ when $\mathit{lik} \not\equiv 0$ on $\mathcal{M}(\underline{P})$

updating

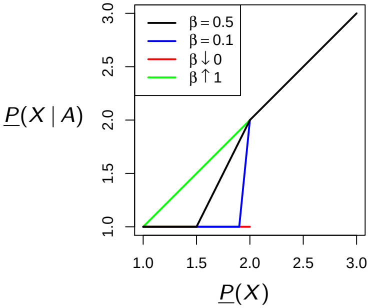
- likelihood of linear previsions: $\mathit{lik}(P) \propto P(A)$
- **Bayesian updating**: $P \mapsto P(\,\cdot\,|\,A)$ when $\mathit{lik}(P) \neq 0$
- **regular updating**: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(A)>0} P(\,\cdot\,|\,A)$ when $\mathit{lik} \not\equiv 0$ on $\mathcal{M}(\underline{P})$
- the discontinuities are "caused" by linear previsions $P \in \mathcal{M}(\underline{P})$ with arbitrarily small likelihood (i.e., "almost contradicted" by the data)
- **β-cut updating**, with $\beta \in (0,1)$: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(A)>\beta\,\overline{P}(A)} P(\,\cdot\,|\,A)$ when $\mathit{lik} \not\equiv 0$ on $\mathcal{M}(\underline{P})$
- regular updating is the limit of β-cut updating as $\beta \downarrow 0$, while the limit as $\beta \uparrow 1$ is a generalization of Dempster's rule of conditioning

updating

- likelihood of linear previsions: $lik(P) \propto P(A)$
- **Bayesian updating**: $P \mapsto P(\cdot \mid A)$ when $lik(P) \neq 0$
- **regular updating**: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(A)>0} P(\cdot \mid A)$ when $lik \not\equiv 0$ on $\mathcal{M}(\underline{P})$
- the discontinuities are "caused" by linear previsions $P \in \mathcal{M}(\underline{P})$ with arbitrarily small likelihood (i.e., "almost contradicted" by the data)
- **β-cut updating**, with $\beta \in (0,1)$: $\underline{P} \mapsto \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(A)>\beta\,\overline{P}(A)} P(\cdot \mid A)$ when $lik \not\equiv 0$ on $\mathcal{M}(\underline{P})$
- regular updating is the limit of β-cut updating as $\beta \downarrow 0$, while the limit as $\beta \uparrow 1$ is a generalization of Dempster's rule of conditioning
- β-cut updating was used, e.g., in Antonucci et al. (2012); Cattaneo and Wiencierz (2012); Destercke (2013), and similar updating rules for imprecise probabilities were suggested, e.g., in Moral and de Campos (1991); Moral (1992); Cano and Moral (1996)

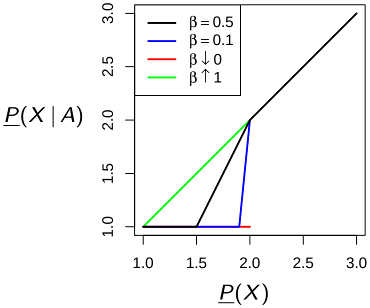
example

- β-cut updating:



example

- β-cut updating:



- the slope of the central line segment is $^1/_\beta$

a more interesting example

- Piatti et al. (2005) studied the behavior of the IDM model when the realization of each binary experiment $X_1, X_2, \ldots$ can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)

a more interesting example

- Piatti et al. (2005) studied the behavior of the IDM model when the realization of each binary experiment $X_1, X_2, \ldots$ can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)
- observation: $X_1 + \cdots + X_{10} = 7$ (i.e., 7 successes in 10 experiments)

a more interesting example

- Piatti et al. (2005) studied the behavior of the IDM model when the realization of each binary experiment $X_1, X_2, \dots$ can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \dots$)
- observation: $X_1 + \cdots + X_{10} = 7$ (i.e., 7 successes in 10 experiments)
- lower probability of success in the next experiment (with $s = 2$): $\underline{P}(X_{11} = 1 \mid X_1 + \cdots + X_{10} = 7) = \dots$

a more interesting example

- Piatti et al. (2005) studied the behavior of the IDM model when the realization of each binary experiment $X_1, X_2, \ldots$ can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)
- observation: $X_1 + \cdots + X_{10} = 7$ (i.e., 7 successes in 10 experiments)
- lower probability of success in the next experiment (with $s = 2$): $\underline{P}(X_{11} = 1 \mid X_1 + \cdots + X_{10} = 7) = \ldots$
- regular updating:

$$\underline{P}(X_{11} = 1 \mid X_1 + \cdots + X_{10} = 7) \approx \begin{cases} 0.583 & \text{if } \varepsilon = 0 \\ 0 & \text{if } \varepsilon > 0 \end{cases}$$

a more interesting example

- Piatti et al. (2005) studied the behavior of the IDM model when the realization of each binary experiment $X_1, X_2, \ldots$ can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)
- observation: $X_1 + \cdots + X_{10} = 7$ (i.e., 7 successes in 10 experiments)
- lower probability of success in the next experiment (with $s = 2$): $\underline{P}(X_{11} = 1 \,|\, X_1 + \cdots + X_{10} = 7) = \ldots$
- regular updating:
$$\underline{P}(X_{11} = 1 \,|\, X_1 + \cdots + X_{10} = 7) \approx \begin{cases} 0.583 & \text{if } \varepsilon = 0 \\ 0 & \text{if } \varepsilon > 0 \end{cases}$$
- β-cut updating with $\beta = 0.01$:
$$\underline{P}(X_{11} = 1 \,|\, X_1 + \cdots + X_{10} = 7) \approx \begin{cases} 0.585 & \text{if } \varepsilon = 0 \\ 0.584 & \text{if } \varepsilon = 0.01 \\ 0.581 & \text{if } \varepsilon = 0.05 \\ 0.567 & \text{if } \varepsilon = 0.1 \end{cases}$$

distances

- distance between two **linear previsions** P, P' on Ω (dual/operator norm = total variation distance):

$$d(P, P') = \sup_{X : \Omega \to [-1, 1]} |P(X) - P'(X)| \\ = 2 \sup_{A \subseteq \Omega} |P(A) - P'(A)|$$

distances

- distance between two **linear previsions** P, P' on Ω (dual/operator norm = total variation distance):

$$d(P, P') = \sup_{X\,:\,\Omega \to [-1,1]} |P(X) - P'(X)| = 2 \sup_{A \subseteq \Omega} |P(A) - P'(A)|$$

- distance between two (coherent) **lower previsions** $\underline{P}, \underline{P}'$ on Ω (Hausdorff distance = "dual/operator norm"):

$$d(\underline{P}, \underline{P}') = \max \left\{ \sup_{P \in \mathcal{M}(\underline{P})} \inf_{P' \in \mathcal{M}(\underline{P}')} d(P, P'),\ \sup_{P' \in \mathcal{M}(\underline{P}')} \inf_{P \in \mathcal{M}(\underline{P})} d(P, P') \right\} = \sup_{X\,:\,\Omega \to [-1,1]} \left| \underline{P}(X) - \underline{P}'(X) \right|$$

continuity

- Bayesian updating is **continuous** on all linear previsions P with $P(A) > 0$

continuity

- Bayesian updating is **continuous** on all linear previsions P with $P(A) > 0$

- regular updating is **continuous** on all lower previsions $\underline{P}$ with $\underline{P}(A) > 0$

continuity

- Bayesian updating is **continuous** on all linear previsions P with $P(A) > 0$
- regular updating is **continuous** on all lower previsions $\underline{P}$ with $\underline{P}(A) > 0$
- regular updating is **not** continuous on all lower previsions $\underline{P}$ with $\overline{P}(A) > 0$, apart if $A = \Omega$ or $|A| = 1$ (and the same holds for natural extension)

continuity

- Bayesian updating is **continuous** on all linear previsions P with $P(A) > 0$
- regular updating is **continuous** on all lower previsions $\underline{P}$ with $\underline{P}(A) > 0$
- regular updating is **not** continuous on all lower previsions $\underline{P}$ with $\overline{P}(A) > 0$, apart if $A = \Omega$ or $|A| = 1$ (and the same holds for natural extension)
- β-cut updating is **continuous** on all lower previsions $\underline{P}$ with $\overline{P}(A) > 0$, for each $\beta \in (0, 1)$

hierarchical model

- a likelihood function *lik′* on the set $\mathcal{P}$ of all linear previsions on Ω is described by its normal hypograph

$$\{(P, \beta) \in \mathcal{P} \times [0,1] : \beta \sup \mathit{lik}' \leq \mathit{lik}'(P)\}$$

hierarchical model

- a likelihood function lik' on the set $\mathcal{P}$ of all linear previsions on Ω is described by its normal hypograph

$$\{(P, \beta) \in \mathcal{P} \times [0, 1] : \beta \sup lik' \leq lik'(P)\}$$

- **hierarchical updating**: $lik' \mapsto lik''(P') \propto \sup_{P \in \mathcal{P} \,:\, P(\,\cdot\,|\,A) = P'} lik'(P)\, lik(P)$ when $lik'\, lik \not\equiv 0$

hierarchical model

- a likelihood function lik' on the set $\mathcal{P}$ of all linear previsions on Ω is described by its normal hypograph

$$\{(P, \beta) \in \mathcal{P} \times [0,1] : \beta \sup lik' \leq lik'(P)\}$$

- **hierarchical updating**: $lik' \mapsto lik''(P') \propto \sup_{P \in \mathcal{P} \,:\, P(\,\cdot\,|\,A) = P'} lik'(P)\, lik(P)$ when $lik'\, lik \not\equiv 0$

- distance between two **likelihood functions** lik', lik'' on $\mathcal{P}$: Hausdorff distance between their normal hypographs, with respect to a product metric on $\mathcal{P} \times [0,1]$

hierarchical model

- a likelihood function lik' on the set $\mathcal{P}$ of all linear previsions on Ω is described by its normal hypograph

$$\{(P, \beta) \in \mathcal{P} \times [0,1] : \beta \sup lik' \leq lik'(P)\}$$

- **hierarchical updating**: $lik' \mapsto lik''(P') \propto \sup_{P \in \mathcal{P} \,:\, P(\,\cdot\,|\,A) = P'} lik'(P)\, lik(P)$ when $lik'\, lik \not\equiv 0$

- distance between two **likelihood functions** lik', lik'' on $\mathcal{P}$: Hausdorff distance between their normal hypographs, with respect to a product metric on $\mathcal{P} \times [0,1]$

- hierarchical updating is **continuous** on all likelihood functions lik' with $lik'\, lik \not\equiv 0$

references

- Antonucci, A., Cattaneo, M., and Corani, G. (2012). Likelihood-based robust classification with Bayesian networks. In *Advances in Computational Intelligence, Part 3*. Springer, 491–500.
- Cano, A., and Moral, S. (1996). A genetic algorithm to approximate convex sets of probabilities. In *IPMU '96*. Vol. 2. Proyecto Sur, 859–864.
- Cattaneo, M., and Wiencierz, A. (2012). Likelihood-based Imprecise Regression. *Int. J. Approx. Reasoning* 53, 1137–1154.
- Destercke, S. (2013). A pairwise label ranking method with imprecise scores and partial predictions. In *Machine Learning and Knowledge Discovery in Databases*. Springer, 112–127.
- Moral, S. (1992). Calculating uncertainty intervals from conditional convex sets of probabilities. In *UAI '92*. Morgan Kaufmann, 199–206.
- Moral, S., and de Campos, L. M. (1991). Updating uncertain information. In *Uncertainty in Knowledge Bases*. Springer, 58–67.
- Piatti, A., Zaffalon, M., and Trojani, F. (2005). Limits of learning from imperfect observations under prior ignorance: the case of the imprecise Dirichlet model. In *ISIPTA '05*. SIPTA, 276–286.