A Continuous Updating Rule for Imprecise Probabilities

Marco Cattaneo
Department of Statistics, LMU Munich

IPMU 2014, Montpellier, France
18 July 2014

previsions

- let Ω be a nonempty set (of possible states of the world) and $\mathcal{L}$ be the set of all bounded functions $X : \Omega \to \mathbb{R}$ (describing an uncertain quantity)

previsions

- let Ω be a nonempty set (of possible states of the world) and $\mathcal{L}$ be the set of all bounded functions $X : \Omega \to \mathbb{R}$ (describing an uncertain quantity)
- let $\mathcal{P}$ be the set of all **linear previsions** $P : \mathcal{L} \to \mathbb{R}$, i.e., all expectation functionals $P : X \to \int X \, \mathrm{d}\pi$ corresponding to some finitely additive probability measure π on Ω (Bhaskara Rao and Bhaskara Rao, 1983)

previsions

- let Ω be a nonempty set (of possible states of the world) and $\mathcal{L}$ be the set of all bounded functions $X : \Omega \to \mathbb{R}$ (describing an uncertain quantity)
- let $\mathcal{P}$ be the set of all **linear previsions** $P : \mathcal{L} \to \mathbb{R}$, i.e., all expectation functionals $P : X \to \int X \, \mathrm{d}\pi$ corresponding to some finitely additive probability measure π on Ω (Bhaskara Rao and Bhaskara Rao, 1983)
- let $\underline{\mathcal{P}}$ be the set of all **coherent lower previsions** $\underline{P} : \mathcal{L} \to \mathbb{R}$, i.e., all functionals of the form $\underline{P} = \min \mathcal{M}(\underline{P})$ with $\mathcal{M}(\underline{P}) = \{P \in \mathcal{P} : P \geq \underline{P}\}$ (Walley, 1991)

previsions

- let Ω be a nonempty set (of possible states of the world) and $\mathcal{L}$ be the set of all bounded functions $X : \Omega \to \mathbb{R}$ (describing an uncertain quantity)
- let $\mathcal{P}$ be the set of all **linear previsions** $P : \mathcal{L} \to \mathbb{R}$, i.e., all expectation functionals $P : X \to \int X \, \mathrm{d}\pi$ corresponding to some finitely additive probability measure π on Ω (Bhaskara Rao and Bhaskara Rao, 1983)
- let $\underline{\mathcal{P}}$ be the set of all **coherent lower previsions** $\underline{P} : \mathcal{L} \to \mathbb{R}$, i.e., all functionals of the form $\underline{P} = \min \mathcal{M}(\underline{P})$ with $\mathcal{M}(\underline{P}) = \{P \in \mathcal{P} : P \geq \underline{P}\}$ (Walley, 1991)
- the upper prevision conjugate to $\underline{P} \in \underline{\mathcal{P}}$ is the functional $\overline{P} = \max \mathcal{M}(\underline{P})$

previsions

- let Ω be a nonempty set (of possible states of the world) and $\mathcal{L}$ be the set of all bounded functions $X : \Omega \to \mathbb{R}$ (describing an uncertain quantity)

- let $\mathcal{P}$ be the set of all **linear previsions** $P : \mathcal{L} \to \mathbb{R}$, i.e., all expectation functionals $P : X \to \int X \, \mathrm{d}\pi$ corresponding to some finitely additive probability measure π on Ω (Bhaskara Rao and Bhaskara Rao, 1983)

- let $\underline{\mathcal{P}}$ be the set of all **coherent lower previsions** $\underline{P} : \mathcal{L} \to \mathbb{R}$, i.e., all functionals of the form $\underline{P} = \min \mathcal{M}(\underline{P})$ with $\mathcal{M}(\underline{P}) = \{P \in \mathcal{P} : P \geq \underline{P}\}$ (Walley, 1991)

- the upper prevision conjugate to $\underline{P} \in \underline{\mathcal{P}}$ is the functional $\overline{P} = \max \mathcal{M}(\underline{P})$

- $P, \underline{P}, \overline{P}$ denote probabilities as well as previsions: e.g., $P(A) = P(I_A)$

updating

- when an event $B \subset \Omega$ is observed, the previsions must be updated (to avoid trivial results, $|B| > 1$ is assumed)

updating

- when an event $B \subset \Omega$ is observed, the previsions must be updated (to avoid trivial results, $|B| > 1$ is assumed)

- **Bayesian updating**: each $P \in \mathcal{P}$ with $P(B) > 0$ is updated to

$$P(\cdot \mid B) : X \mapsto P(X \mid B) = \frac{P(X\, I_B)}{P(B)}$$

updating

- when an event $B \subset \Omega$ is observed, the previsions must be updated (to avoid trivial results, $|B| > 1$ is assumed)

- **Bayesian updating**: each $P \in \mathcal{P}$ with $P(B) > 0$ is updated to

$$P(\,\cdot \mid B) : X \mapsto P(X \mid B) = \frac{P(X\, I_B)}{P(B)}$$

- **regular extension**: each $\underline{P} \in \underline{\mathcal{P}}$ with $\overline{P}(B) > 0$ is updated to

$$\underline{P}(\,\cdot \mid B) : X \mapsto \underline{P}(X \mid B) = \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(B)>0} P(X \mid B)$$

example

- Piatti et al. (2009) studied exchangeable sequences of binary experiments $X_1, X_2, \ldots \in \{0, 1\}$ such that the realization of each X_i can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)

example

- Piatti et al. (2009) studied exchangeable sequences of binary experiments $X_1, X_2, \ldots \in \{0, 1\}$ such that the realization of each X_i can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)

- event B: (possibly incorrect) observation of 7 times "1" and 2 times "0" in the first 9 experiments

example

- Piatti et al. (2009) studied exchangeable sequences of binary experiments $X_1, X_2, \ldots \in \{0, 1\}$ such that the realization of each X_i can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)

- event B: (possibly incorrect) observation of 7 times "1" and 2 times "0" in the first 9 experiments

- Bayesian conjugate model P with Jeffreys' prior, and IDM model $\underline{P}$ with hyperparameter $s = 1$:

	$\varepsilon = 0$	$\varepsilon = 10^{-6}$
$P(X_1) = 0.5$	$P(X_{10} \mid B) = 0.75$	$P(X_{10} \mid B) \approx 0.750$
$\overline{P}(X_1) = 1$	$\overline{P}(X_{10} \mid B) = 0.8$	$\overline{P}(X_{10} \mid B) = 1$
$\underline{P}(X_1) = 0$	$\underline{P}(X_{10} \mid B) = 0.7$	$\underline{P}(X_{10} \mid B) = 0$

continuity

- metric on $\mathcal{P}$ (dual norm $= 2\times$ total variation distance):

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

continuity

- metric on $\mathcal{P}$ (dual norm $= 2\times$ total variation distance):

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

- metric on $\underline{\mathcal{P}}$ (Hausdorff distance = "dual norm"):

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

continuity

- metric on $\mathcal{P}$ (dual norm = $2\times$ total variation distance):

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

- metric on $\underline{\mathcal{P}}$ (Hausdorff distance = "dual norm"):

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

- the Bayesian updating $\{P \in \mathcal{P} : P(B) > 0\} \to \{P \in \mathcal{P} : P(B) = 1\}$ is **continuous**

continuity

- metric on $\mathcal{P}$ (dual norm $= 2\times$ total variation distance):

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

- metric on $\underline{\mathcal{P}}$ (Hausdorff distance = "dual norm"):

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

- the Bayesian updating $\{P \in \mathcal{P} : P(B) > 0\} \to \{P \in \mathcal{P} : P(B) = 1\}$ is **continuous**

- the regular extension $\left\{\underline{P} \in \underline{\mathcal{P}} : \overline{P}(B) > 0\right\} \to \{\underline{P} \in \underline{\mathcal{P}} : \underline{P}(B) = 1\}$ is **not** continuous: it has discontinuities at points $\underline{P} \in \underline{\mathcal{P}}$ with $\overline{P}(B) > \underline{P}(B) = 0$

updating

- when an event $B \subset \Omega$ is observed, the previsions must be updated (to avoid trivial results, $|B| > 1$ is assumed)

- **Bayesian updating**: each $P \in \mathcal{P}$ with $P(B) > 0$ is updated to

$$P(\cdot \mid B) : X \mapsto P(X \mid B) = \frac{P(X\, I_B)}{P(B)}$$

- **regular extension**: each $\underline{P} \in \underline{\mathcal{P}}$ with $\overline{P}(B) > 0$ is updated to

$$\underline{P}(\cdot \mid B) : X \mapsto \underline{P}(X \mid B) = \inf_{P \in \mathcal{M}(\underline{P})\,:\,P(B)>0} P(X \mid B)$$

- ***α*-cut rule**, where $\alpha \in (0,1)$: each $\underline{P} \in \underline{\mathcal{P}}$ with $\overline{P}(B) > 0$ is updated to

$$\underline{P}_\alpha(\cdot \mid B) : X \mapsto \underline{P}_\alpha(X \mid B) = \min_{P \in \mathcal{M}(\underline{P})\,:\,P(B) \geq \alpha\, \overline{P}(B)} P(X \mid B)$$

example

- Piatti et al. (2009) studied exchangeable sequences of binary experiments $X_1, X_2, \ldots \in \{0, 1\}$ such that the realization of each X_i can be observed incorrectly with a known probability ε (the errors of observation are independent, conditional on the realizations of $X_1, X_2, \ldots$)

- event B: (possibly incorrect) observation of 7 times "1" and 2 times "0" in the first 9 experiments

- Bayesian conjugate model P with Jeffreys' prior, and IDM model $\underline{P}$ with hyperparameter $s = 1$ (and $\alpha = 0.001$):

	$\varepsilon = 0$	$\varepsilon = 10^{-6}$
$P(X_1) = 0.5$	$P(X_{10} \mid B) = 0.75$	$P(X_{10} \mid B) \approx 0.750$
$\overline{P}(X_1) = 1$	$\overline{P}(X_{10} \mid B) = 0.8$	$\overline{P}(X_{10} \mid B) = 1$
$\underline{P}(X_1) = 0$	$\underline{P}(X_{10} \mid B) = 0.7$	$\underline{P}(X_{10} \mid B) = 0$
$\overline{P}(X_1) = 1$	$\overline{P}_\alpha(X_{10} \mid B) \approx 0.800$	$\overline{P}_\alpha(X_{10} \mid B) \approx 0.800$
$\underline{P}(X_1) = 0$	$\underline{P}_\alpha(X_{10} \mid B) \approx 0.700$	$\underline{P}_\alpha(X_{10} \mid B) \approx 0.700$

continuity

- metric on $\mathcal{P}$ (dual norm $=$ $2\times$ total variation distance):

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

- metric on $\underline{\mathcal{P}}$ (Hausdorff distance $=$ "dual norm"):

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

- the Bayesian updating $\{P\in\mathcal{P} : P(B)>0\} \to \{P\in\mathcal{P} : P(B)=1\}$ is **continuous**

- the regular extension $\left\{\underline{P}\in\underline{\mathcal{P}} : \overline{P}(B)>0\right\} \to \{\underline{P}\in\underline{\mathcal{P}} : \underline{P}(B)=1\}$ is **not** continuous: it has discontinuities at points $\underline{P}\in\underline{\mathcal{P}}$ with $\overline{P}(B)>\underline{P}(B)=0$

- the α-cut rule $\left\{\underline{P}\in\underline{\mathcal{P}} : \overline{P}(B)>0\right\} \to \{\underline{P}\in\underline{\mathcal{P}} : \underline{P}(B)=1\}$ is **continuous**, for all $\alpha\in(0,1)$

conclusion

- the Bayesian updating of linear previsions is continuous

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues
- the α-cut rule:

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues
- the α-cut rule:
  - generalizes Bayesian updating, and corresponds to regular/natural extension (and any other coherent updating rule) when $\underline{P}(B) \geq \alpha \, \overline{P}(B) > 0$

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues
- the α-cut rule:
  - generalizes Bayesian updating, and corresponds to regular/natural extension (and any other coherent updating rule) when $\underline{P}(B) \geq \alpha\,\overline{P}(B) > 0$
  - is **continuous**

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues
- the α-cut rule:
  - generalizes Bayesian updating, and corresponds to regular/natural extension (and any other coherent updating rule) when $\underline{P}(B) \geq \alpha\,\overline{P}(B) > 0$
  - is **continuous**
  - allows **vacuous priors** in statistical analyses, and has been successfully applied (Antonucci et al., 2012; Cattaneo and Wiencierz, 2012; Destercke, 2013)

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues
- the α-cut rule:
  - generalizes Bayesian updating, and corresponds to regular/natural extension (and any other coherent updating rule) when $\underline{P}(B) \geq \alpha\,\overline{P}(B) > 0$
  - is **continuous**
  - allows **vacuous priors** in statistical analyses, and has been successfully applied (Antonucci et al., 2012; Cattaneo and Wiencierz, 2012; Destercke, 2013)
  - lies between regular extension (limit $\alpha \to 0$) and a generalization of Dempster's rule of conditioning (limit $\alpha \to 1$), and is strictly related to other proposals (Moral, 1992; Gilboa and Schmeidler, 1993; Cano and Moral, 1996)

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues
- the α-cut rule:
  - generalizes Bayesian updating, and corresponds to regular/natural extension (and any other coherent updating rule) when $\underline{P}(B) \geq \alpha\,\overline{P}(B) > 0$
  - is **continuous**
  - allows **vacuous priors** in statistical analyses, and has been successfully applied (Antonucci et al., 2012; Cattaneo and Wiencierz, 2012; Destercke, 2013)
  - lies between regular extension (limit $\alpha \to 0$) and a generalization of Dempster's rule of conditioning (limit $\alpha \to 1$), and is strictly related to other proposals (Moral, 1992; Gilboa and Schmeidler, 1993; Cano and Moral, 1996)
  - is not iteratively consistent (i.e., updates do not commute), but this can be remedied by recording the whole likelihood function (Cattaneo, 2008, 2009)

conclusion

- the Bayesian updating of linear previsions is continuous
- the regular extension, as well as the natural extension and any other updating rule that is **coherent** in the sense of Walley (1991):
  - generalizes Bayesian updating to lower previsions
  - is **discontinuous** when $\underline{P}(B) = 0$, and this leads to robustness issues
- the α-cut rule:
  - generalizes Bayesian updating, and corresponds to regular/natural extension (and any other coherent updating rule) when $\underline{P}(B) \geq \alpha\, \overline{P}(B) > 0$
  - is **continuous**
  - allows **vacuous priors** in statistical analyses, and has been successfully applied (Antonucci et al., 2012; Cattaneo and Wiencierz, 2012; Destercke, 2013)
  - lies between regular extension (limit $\alpha \to 0$) and a generalization of Dempster's rule of conditioning (limit $\alpha \to 1$), and is strictly related to other proposals (Moral, 1992; Gilboa and Schmeidler, 1993; Cano and Moral, 1996)
  - is not iteratively consistent (i.e., updates do not commute), but this can be remedied by recording the whole likelihood function (Cattaneo, 2008, 2009)
- the relative importance of properties like coherence, continuity, and ability of using vacuous priors depends on the application field and on the exact interpretation of imprecise probabilities

references

Antonucci, A., Cattaneo, M., and Corani, G. (2012). Likelihood-based robust classification with Bayesian networks. In *IPMU '12*. Vol. 3. Springer, 491–500.

Bhaskara Rao, K. P. S., and Bhaskara Rao, M. (1983). *Theory of Charges*. Academic Press.

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. (2008). Fuzzy probabilities based on the likelihood function. In *SMPS '08*. Springer, 43–50.

Cattaneo, M. (2009). A generalization of credal networks. In *ISIPTA '09*. SIPTA, 79–88.

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 *ECMLPKDD '13*. Vol. 2. Springer, 112–127.

Gilboa, I., and Schmeidler, D. (1993). Updating ambiguous beliefs. *J. Econ. Theory* 59, 33–49.

Moral, S. (1992). Calculating uncertainty intervals from conditional convex sets of probabilities. In *UAI '92*. Morgan Kaufmann, 199–206.

Piatti, A., Zaffalon, M., Trojani, F., and Hutter, M. (2009). Limits of learning about a categorical latent variable under prior near-ignorance. *Int. J. Approx. Reasoning* 50, 597–611.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.