A Continuous Updating Rule for Imprecise Probabilities

Marco E. G. V. Cattaneo

Department of Statistics, LMU Munich
Ludwigstraße 33, 80539 München, Germany
cattaneo@stat.uni-muenchen.de
www.statistik.lmu.de/~cattaneo

Abstract. The paper studies the continuity of rules for updating imprecise probability models when new data are observed. Discontinuities can lead to robustness issues: this is the case for the usual updating rules of the theory of imprecise probabilities. An alternative, continuous updating rule is introduced.

Keywords: coherent lower and upper previsions, natural extension, regular extension, α-cut, robustness, Hausdorff distance.

1 Introduction

Imprecise probability models must be updated when new information is gained. In particular, prior (coherent) lower previsions must be updated to posterior ones when new data are observed. Unfortunately, the usual updating rules of the theory of imprecise probabilities have some discontinuities. These can lead to robustness problems, because an arbitrarily small change in the prior lower previsions can induce a substantial change in the posterior ones.

In the next section, the discontinuity of the usual updating rules is illustrated by examples and formally studied in the framework of functional analysis. Then, in Sect. 3, an alternative, continuous updating rule is introduced and discussed. The final section gives directions for further research.

2 Discontinuous Updating Rules and Robustness Issues

Let Ω be a nonempty set of possible states of the world. A (bounded) uncertain payoff depending on the true state of the world $\omega \in \Omega$ can be represented by an element of $\mathcal{L}$, the set of all bounded real-valued functions on Ω. In the Bayesian theory, the uncertain belief or information about the true state of the world $\omega \in \Omega$ is described by a (finitely additive) probability measure P on Ω [1,2]. The expectation $P(X)$ of an uncertain payoff $X \in \mathcal{L}$ is its integral with respect to this probability measure [3, Chap. 4]. Hence, P denotes the probability as well as the expectation: $P(A) = P(I_A)$, where I_A denotes the indicator function of the event $A \subseteq \Omega$.

This is the unformatted version of: doi:10.1007/978-3-319-08852-5_44

Let $\mathcal{P}$ be the set of all expectation functionals $P : \mathcal{L} \to \mathbb{R}$ corresponding to integrals with respect to a (finitely additive) probability measure on Ω. In the theory of imprecise probabilities, the elements of $\mathcal{P}$ are called linear previsions and are a special case of coherent lower and upper previsions. A (coherent) lower prevision $\underline{P} : \mathcal{L} \to \mathbb{R}$ is the (pointwise) infimum of a nonempty set $\mathcal{M} \subseteq \mathcal{P}$ of linear previsions [4, Sect. 3.3]. Hence, a lower prevision $\underline{P}$ is determined by the set $\mathcal{M}(\underline{P}) = \{P \in \mathcal{P} : P \geq \underline{P}\}$ of all linear previsions (pointwise) dominating it.

Let $\underline{\mathcal{P}}$ be the set of all (coherent) lower previsions $\underline{P} : \mathcal{L} \to \mathbb{R}$. The upper prevision $\overline{P} : \mathcal{L} \to \mathbb{R}$ conjugate to a lower prevision $\underline{P}$ is the (pointwise) supremum of $\mathcal{M}(\underline{P})$. Hence, $\overline{P}(X) = -\underline{P}(-X)$ for all $X \in \mathcal{L}$, and $\underline{P}$ is linear (i.e., $\underline{P} \in \mathcal{P} \subseteq \underline{\mathcal{P}}$) if and only if $\overline{P} = \underline{P}$. As in the case of linear previsions, $\underline{P}$ and $\overline{P}$ denote also the corresponding lower and upper probabilities: $\underline{P}(A) = \underline{P}(I_A)$ and $\overline{P}(A) = \overline{P}(I_A) = 1 - \underline{P}(A^c)$ for all $A \subseteq \Omega$. However, contrary to the case of linear previsions, in general these lower and upper probability values do not completely determine the corresponding lower and upper previsions.

Linear previsions and lower (or upper) previsions are quantitative descriptions of uncertain belief or information. When an event $B \subseteq \Omega$ is observed, these descriptions must be updated. The updating is trivial when $B = \Omega$ or when B is a singleton. Hence, in order to avoid trivial results, it is assumed that B is a proper subset of Ω with at least two elements.

In the Bayesian theory, a linear prevision $P \in \mathcal{P}$ with $P(B) > 0$ is updated to $P(\cdot \mid B)$, the conditional linear prevision given B. That is, the integral with respect to the probability measure P conditioned on B. When $P(B) = 0$, the Bayesian updating of the linear prevision P is not defined.

In the theory of imprecise probabilities, there are two main updating rules: natural extension and regular extension [4, Appendix J]. According to both rules, a lower prevision $\underline{P} \in \underline{\mathcal{P}}$ with $\underline{P}(B) > 0$ is updated to the infimum $\underline{P}(\cdot \mid B)$ of all conditional linear previsions $P(\cdot \mid B)$ with $P \in \mathcal{M}(\underline{P})$. The natural extension updates each lower prevision $\underline{P} \in \underline{\mathcal{P}}$ such that $\underline{P}(B) = 0$ to the vacuous conditional lower prevision given B. That is, the lower prevision $\underline{V}(\cdot \mid B)$ such that $\underline{V}(X \mid B) = \inf_{\omega \in B} X(\omega)$ for all $X \in \mathcal{L}$, describing the complete ignorance about $\omega \in B$. By contrast, the regular extension updates a lower prevision $\underline{P} \in \underline{\mathcal{P}}$ such that $\overline{P}(B) > 0$ to the infimum $\underline{P}(\cdot \mid B)$ of all conditional linear previsions $P(\cdot \mid B)$ with $P \in \mathcal{M}(\underline{P})$ and $P(B) > 0$, and updates only the lower previsions $\underline{P} \in \underline{\mathcal{P}}$ such that $\overline{P}(B) = 0$ to the vacuous conditional lower prevision given B. Hence, the natural and regular extensions are always defined, and they agree for all lower previsions $\underline{P} \in \underline{\mathcal{P}}$ such that either $\underline{P}(B) > 0$ or $\overline{P}(B) = 0$.

Both natural and regular extensions generalize the Bayesian updating: if a lower prevision $\underline{P}$ is linear and $\overline{P}(B) = \underline{P}(B) > 0$, then $\underline{P}(\cdot \mid B)$ is the conditional linear prevision given B. If a lower prevision $\underline{P}$ is linear and $\overline{P}(B) = \underline{P}(B) = 0$, then the conditional linear prevision given B does not exist, and both natural and regular extensions update $\underline{P}$ to $\underline{V}(\cdot \mid B)$, which is not linear. But with lower previsions, besides the cases with $\underline{P}(B)$ and $\overline{P}(B)$ both positive or both zero, there is also the case with $\overline{P}(B) > \underline{P}(B) = 0$. The fact that there are two different updating rules for this case shows that it is more challenging than the others.

Example 1. A simple instance of discontinuity in the updating of lower and upper previsions is the following [5, Example 2]. Let X be a function on Ω with image $\{1, 2, 3\}$, and let $\underline{P}$ be the lower prevision determined by the set $\mathcal{M}(\underline{P}) = \{P \in \mathcal{P} : P(X) \geq x\}$, where $x \in [1, 3]$. That is, $\underline{P}$ is the lower prevision based on the unique assessment $\underline{P}(X) = x$.

Assume that the event $B = \{X \neq 2\}$ is observed and the lower prevision $\underline{P}$ must be updated. If $x > 2$, then $\underline{P}(B) = x - 2 > 0$, and both natural and regular extensions update $\underline{P}$ to $\underline{P}(\cdot \mid B)$, with in particular $\underline{P}(X \mid B) = x$. On the other hand, if $x \leq 2$, then $\overline{P}(B) = 1 > \underline{P}(B) = 0$, and the regular extension updates $\underline{P}$ to $\underline{P}(\cdot \mid B)$, while the natural extension updates it to $\underline{V}(\cdot \mid B)$. However, $\underline{P}(\cdot \mid B)$ and $\underline{V}(\cdot \mid B)$ are equal when $x < 2$, while they are different when $x = 2$, with in particular $\underline{P}(X \mid B) = 2$ and $\underline{V}(X \mid B) = 1$.

Therefore, according to both rules, the updated lower previsions of X are discontinuous functions of $x \in [1, 3]$ at $x = 2$. These functions are plotted as a solid line in Fig. 1 (the other functions plotted in Fig. 1 will be discussed in Sect. 3). Hence, inferences and decisions based on imprecise probability models can depend in a discontinuous way from the prior assessments, and this can lead to robustness issues [5].

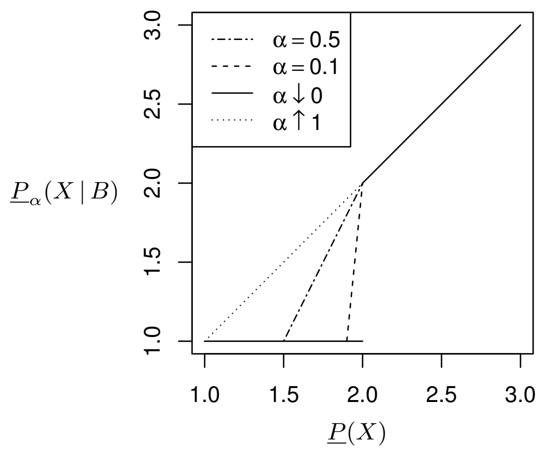

Fig. 1. Updated lower prevision $\underline{P}_\alpha(X \mid B)$ from Example 1 as a function of $\underline{P}(X) = x$, for some values of $\alpha \in (0, 1)$.

Example 2. A more important instance of discontinuity in the updating of lower and upper previsions is the following [6, Example 9]. Let $X_1, \ldots, X_{10}, Y_1, \ldots, Y_{10}$ be 20 functions on Ω with image $\{0, 1\}$, and for each $t \in (0, 1)$, let $P_t \in \mathcal{P}$ be a linear prevision such that

$$P_t(X_1 = x_1, \dots, X_{10} = x_{10}, Y_1 = y_1, \dots, Y_{10} = y_{10}) = \int_0^1 \frac{\theta^{t-1+\sum_{i=1}^{10} x_i} \, (1-\theta)^{10-t-\sum_{i=1}^{10} x_i}}{\Gamma(t)\,\Gamma(1-t)} \, d\theta \, \varepsilon^{\sum_{i=1}^{10} |x_i - y_i|} \, (1-\varepsilon)^{10-\sum_{i=1}^{10} |x_i - y_i|}$$

for all $x_1, \dots, x_{10}, y_1, \dots, y_{10} \in \{0, 1\}$, where $\varepsilon \in [0, 1/2]$. That is, P_t describes a Bayesian model in which $X_1, \dots, X_{10}$ follow the beta-Bernoulli distribution with (prior) expectation t and variance $t\,(1-t)/2$, while each Y_i is independent of the other 18 variables given $X_i = x_i$, with conditional probability ε that $Y_i \neq x_i$.

Let the lower prevision $\underline{P} \in \underline{\mathcal{P}}$ be the infimum of all linear previsions P_t with $t \in (0,1)$, and assume that the event $B = \{\sum_{i=1}^{9} Y_i = 7\}$ is observed. That is, $\underline{P}$ describes an imprecise beta-Bernoulli model for $X_1, \dots, X_{10}$ with hyperparameter $s = 1$ [7], but instead of the (latent) variables X_i, only the proxy variables Y_i can be observed, which can be wrong with a (small) probability ε. In the first 9 binary experiments, 7 successes are (possibly incorrectly) reported, and the updated lower prevision of a success in the last experiment is the quantity of interest.

If $\varepsilon > 0$, then $\underline{P}(B) \geq \varepsilon^{10} > 0$, and both natural and regular extensions update $\underline{P}$ to $\underline{P}(\,\cdot \mid B)$, with in particular $\underline{P}(X_{10} \mid B) = 0$ [6]. On the other hand, if $\varepsilon = 0$, then $\overline{P}(B) > \underline{P}(B) = 0$, and the regular extension updates $\underline{P}$ to $\underline{P}(\,\cdot \mid B)$, while the natural extension updates it to $\underline{V}(\,\cdot \mid B)$, with in particular $\underline{P}(X_{10} \mid B) = 0.7$ and $\underline{V}(X_{10} \mid B) = 0$ [7].

Hence, according to the natural extension, the updated lower prevision of X_{10} is a continuous function of $\varepsilon \in [0, 1/2]$. However, since vacuous conditional previsions are not very useful in statistical applications, the usual updating rule for the imprecise beta-Bernoulli model is regular extension, according to which the updated lower prevision of X_{10} is a discontinuous function of $\varepsilon \in [0, 1/2]$ at $\varepsilon = 0$. That is, the assumption that the experimental results can be incorrectly reported with an arbitrarily small positive probability leads to very different conclusions than the assumption that the results are always correctly reported. This is an important robustness problem of the imprecise beta-Bernoulli and other similar models [6].

In order to investigate more carefully such discontinuities in the updating of imprecise probability models, a metric on the set $\underline{\mathcal{P}}$ of all (coherent) lower previsions can be introduced. Lower previsions are functionals on the set $\mathcal{L}$ of all bounded real-valued functions on Ω. The set $\mathcal{L}$ has a natural norm, which makes it a Banach space: the supremum norm $\|\cdot\|_\infty$, defined by $\|X\|_\infty = \sup_{\omega \in \Omega} |X(\omega)|$ for all $X \in \mathcal{L}$ [8, Sect. IV.5].

The set $\mathcal{P}$ of all linear previsions is a subset of the dual space $\mathcal{L}^*$ of $\mathcal{L}$, consisting of all continuous linear functionals on $\mathcal{L}$ [3, Sect. 4.7]. The dual space of $\mathcal{L}$ is a Banach space when endowed with the dual norm $\|\cdot\|$, defined by $\|F\| = \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} |F(X)|$ for all $F \in \mathcal{L}^*$ [8, Sect. II.3]. The dual norm induces a metric d on $\mathcal{P}$ such that the distance between two linear previsions $P, P' \in \mathcal{P}$ is 2 times the total variation distance between the corresponding (finitely additive) probability measures [8, Sect. IV.5]:

$$d(P, P') = \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} |P(X) - P'(X)| = 2 \sup_{A \subseteq \Omega} |P(A) - P'(A)| .$$

A lower prevision is determined by a nonempty set of linear previsions. The Hausdorff distance between two nonempty sets $\mathcal{M}, \mathcal{M}' \subseteq \mathcal{P}$ of linear previsions is defined as

$$d_H(\mathcal{M}, \mathcal{M}') = \max \left\{ \sup_{P \in \mathcal{M}} \inf_{P' \in \mathcal{M}'} d(P, P'), \sup_{P' \in \mathcal{M}'} \inf_{P \in \mathcal{M}} d(P, P') \right\} .$$

The next theorem shows that the Hausdorff distance can be used to extend d to a metric on $\underline{\mathcal{P}}$, which can also be interpreted as a direct generalization of the metric induced by the the dual norm. However, the connection with the total variation distance is lost, because in general the lower and upper probability measures do not completely determine the corresponding lower and upper previsions.

Theorem 1. *The metric d on $\mathcal{P}$ can be extended to a metric on $\underline{\mathcal{P}}$ by defining, for all $\underline{P}, \underline{P}' \in \underline{\mathcal{P}}$,*

$$d(\underline{P}, \underline{P}') = \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} \left| \underline{P}(X) - \underline{P}'(X) \right| = d_H \left(\mathcal{M}(\underline{P}), \mathcal{M}(\underline{P}') \right) .$$

Proof. The first equality clearly defines a metric on $\underline{\mathcal{P}}$ that extends the metric d on $\mathcal{P}$. The second equality can be shown by generalizing the proof of [9, Theorem 2] to the case of a possibly infinite set Ω. Let $\underline{P}, \underline{P}' \in \underline{\mathcal{P}}$, and define $f : (P, X) \mapsto P(X) - P'(X)$ on $\mathcal{M}(\underline{P}) \times \mathcal{L}$, where $P' \in \mathcal{P}$. In the weak* topology, $\mathcal{M}(\underline{P})$ is compact and Hausdorff [4, Appendix D], while f is continuous and convex in P, and concave in X. Therefore, the minimax theorem [10, Theorem 2] implies

$$\inf_{P \in \mathcal{M}(\underline{P})} \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} |P(X) - P'(X)| = \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} \inf_{P \in \mathcal{M}(\underline{P})} (P(X) - P'(X)),$$

since the absolute value on the left-hand side has no effect, because P, P' are linear and X can be replaced by $-X$. Hence,

$$\begin{aligned} \sup_{P' \in \mathcal{M}(\underline{P}')} \inf_{P \in \mathcal{M}(\underline{P})} d(P, P') &= \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} \sup_{P' \in \mathcal{M}(\underline{P}')} \inf_{P \in \mathcal{M}(\underline{P})} (P(X) - P'(X)) \\ &= \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} \left(\underline{P}(X) - \underline{P}'(X) \right) , \end{aligned}$$

and analogously, by exchanging the roles of $\underline{P}$ and $\underline{P}'$,

$$\sup_{P \in \mathcal{M}(\underline{P})} \inf_{P' \in \mathcal{M}(\underline{P}')} d(P, P') = \sup_{X \in \mathcal{L} \,:\, \|X\|_\infty \leq 1} \left(\underline{P}'(X) - \underline{P}(X) \right) ,$$

from which the desired result follows. □

The continuity of the updating rules can now be studied with respect to the metric d on $\underline{\mathcal{P}}$. The next theorem shows that, contrary to Bayesian updating, the usual updating rules of the theory of imprecise probabilities are not continuous on the whole domain on which they are defined.

Theorem 2. *The Bayesian updating rule $P \mapsto P(\,\cdot\,|\,B)$ on $\{P \in \mathcal{P} : P(B) > 0\}$ is continuous.*

The regular extension $\underline{P} \mapsto \underline{P}(\,\cdot\,|\,B)$ is continuous on $\{\underline{P} \in \underline{\mathcal{P}} : \underline{P}(B) > 0\}$, but not on $\{\underline{P} \in \underline{\mathcal{P}} : \overline{P}(B) > 0\}$. The same holds for the natural extension.

Proof. Let $g : \mathcal{P} \to \mathcal{L}^*$ and $h : \mathcal{P} \to \mathbb{R}$ be the two functions that are defined by $g(P) = P(\,\cdot\, I_B)$ and $h(P) = P(B)$, respectively, for all $P \in \mathcal{P}$. The functions g, h are uniformly continuous with respect to the dual and Euclidean norms, respectively, since $\|P(\,\cdot\, I_B) - P'(\,\cdot\, I_B)\| \leq d(P, P')$ and $|P(B) - P'(B)| \leq {}^1\!/\!{}_2\, d(P, P')$ for all $P, P' \in \mathcal{P}$. Hence, their ratio ${}^g\!/\!{}_h : P \mapsto P(\,\cdot\,|\,B)$ is continuous on the set $\{P \in \mathcal{P} : P(B) > 0\}$, and uniformly continuous on the set $\{P \in \mathcal{P} : P(B) > \alpha\}$, for all $\alpha \in (0, 1)$.

For each $\alpha \in (0, 1)$, let $\mathcal{S}_\alpha$ be the power set of $\{P \in \mathcal{P} : P(B) > \alpha\}$. The uniform continuity of $P \mapsto P(\,\cdot\,|\,B)$ on $\{P \in \mathcal{P} : P(B) > \alpha\}$ implies the uniform continuity of $\mathcal{M} \mapsto \{P(\,\cdot\,|\,B) : P \in \mathcal{M}\}$ on $\mathcal{S}_\alpha \setminus \{\varnothing\}$ with respect to the Hausdorff pseudometric d_H, for all $\alpha \in (0, 1)$.

Therefore, in order to complete the proof of the continuity of $\underline{P} \mapsto \underline{P}(\,\cdot\,|\,B)$ on $\{\underline{P} \in \underline{\mathcal{P}} : \underline{P}(B) > 0\}$, it suffices to show that the sets $\{P(\,\cdot\,|\,B) : P \in \mathcal{M}(\underline{P})\}$ and $\mathcal{M}\left(\underline{P}(\,\cdot\,|\,B)\right)$ are equal for all $\underline{P} \in \underline{\mathcal{P}}$ such that $\underline{P}(B) > 0$. That is, it suffices to show that $\{P(\,\cdot\,|\,B) : P \in \mathcal{M}(\underline{P})\}$ is nonempty, weak*-compact, and convex, for all $\underline{P} \in \underline{\mathcal{P}}$ such that $\underline{P}(B) > 0$ [4, Sect. 3.6], but this is implied by the fact that the function $P \mapsto P(\,\cdot\,|\,B)$ on $\{P \in \mathcal{P} : P(B) > 0\}$ is weak*-continuous and maintains segments [11, Theorem 2].

Finally, let X be a function on Ω with image $\{1, 2, 3\}$ such that $B = \{X \neq 2\}$, and for each $x \in [1, 3]$, let $\underline{P}_x$ be the lower prevision based on the unique assessment $\underline{P}_x(X) = x$. Example 1 implies that both regular and natural extensions updating of $\underline{P}_x$ are discontinuous functions of $x \in [1, 3]$ at $x = 2$. Hence, in order to prove that both regular and natural extensions are not continuous on $\{\underline{P} \in \underline{\mathcal{P}} : \overline{P}(B) > 0\}$, it suffices to show that the function $x \mapsto \underline{P}_x$ on $[1, 3]$ is continuous.

Let $x, x' \in [1, 3]$ with $x < x'$. For each $P \in \mathcal{P}$ such that $x \leq P(X) < x'$, another $P' \in \mathcal{P}$ with $P'(X) = x'$ and $d(P, P') \leq 2\,(x' - x)$ can be obtained by moving some probability mass from $\{X \neq 3\}$ to $\{X = 3\}$. Therefore, $d_H\left(\mathcal{M}(\underline{P}_x), \mathcal{M}(\underline{P}_{x'})\right) \leq 2\,(x' - x)$, and the function $x \mapsto \underline{P}_x$ on $[1, 3]$ is thus continuous. □

Hence, the case with $\overline{P}(B) > \underline{P}(B) = 0$ is a source of discontinuity in the usual updating rules of the theory of imprecise probabilities. Unfortunately, this case is very common and leads to robustness issues in practical applications of the theory, such as the ones based on the imprecise beta-Bernoulli model discussed in Example 2. In the next section, an alternative updating rule avoiding these difficulties is introduced.

3 A Continuous Updating Rule

The discontinuities of the regular extension when $\overline{P}(B) > \underline{P}(B) = 0$ are due to the fact that in this case $\mathcal{M}(\underline{P})$ contains linear previsions P with arbitrarily small $P(B)$, so that their updating $P \mapsto P(\cdot \mid B)$ is arbitrarily sensitive to small changes in P. However, these linear previsions are almost refuted by the observed data B. In fact, in a hierarchical Bayesian model with a second-order probability measure on $\mathcal{M}(\underline{P})$, the linear previsions $P \in \mathcal{M}(\underline{P})$ with arbitrarily small $P(B)$ would not pose a problem for the updating, since they would be downweighted by the (second-order) likelihood function $P \mapsto P(B)$. Theorem 2 implies that the updating of such a hierarchical Bayesian model would be continuous, if possible (i.e., if the second-order probability measure is not concentrated on the linear previsions $P \in \mathcal{M}(\underline{P})$ with $P(B) = 0$).

The regular extension updates a lower prevision $\underline{P} \in \underline{\mathcal{P}}$ such that $\overline{P}(B) > 0$ by conditioning on B all linear previsions $P \in \mathcal{M}(\underline{P})$ with positive likelihood $P(B)$. A simple way to define an alternative updating rule avoiding the discontinuities of the regular extension is to discard not only the linear previsions with zero likelihood, but also the too unlikely ones. The α-cut rule updates each lower prevision $\underline{P} \in \underline{\mathcal{P}}$ such that $\overline{P}(B) > 0$ to the infimum $\underline{P}_\alpha(\cdot \mid B)$ of all conditional linear previsions $P(\cdot \mid B)$ with $P \in \mathcal{M}(\underline{P})$ and $P(B) \geq \alpha\, \overline{P}(B)$, where $\alpha \in (0, 1)$. That is, the linear previsions whose (relative) likelihood is below the threshold α are discarded from the set $\mathcal{M}(\underline{P})$, before conditioning on B all linear previsions in this set. Hence, the α-cut updating corresponds to the natural and regular extensions when $\underline{P}(B) \geq \alpha\, \overline{P}(B) > 0$, and thus in particular it generalizes the Bayesian updating.

From the standpoint of the theory of imprecise probabilities, the α-cut updating rule consists in replacing the lower prevision $\underline{P}$ by $\underline{P}_\alpha$ before updating it by regular extension (or natural extension, since they agree in this case), where the lower prevision $\underline{P}_\alpha$ is obtained from $\underline{P}$ by including the unique additional assessment $\underline{P}_\alpha(B) \geq \alpha\, \overline{P}(B)$. That is, $\mathcal{M}(\underline{P}_\alpha) = \mathcal{M}(\underline{P}) \cap \{P \in \mathcal{P} : P(B) \geq \alpha\, \overline{P}(B)\}$. Updating rules similar to the α-cut have been suggested in the literature on imprecise probabilities [12,13,14,15]. When $\overline{P}(B) > 0$, the regular extension $\underline{P}(\cdot \mid B)$ is the limit of the α-cut updated lower prevision $\underline{P}_\alpha(\cdot \mid B)$ as α tends to 0. Furthermore, when Ω is finite and $\underline{P}$ corresponds to a belief function on Ω, the limit of the α-cut updated lower prevision $\underline{P}_\alpha(\cdot \mid B)$ as α tends to 1 corresponds to the result of Dempster's rule of conditioning [16,12].

Example 1 (continued). Since $\overline{P}(B) = 1$, the α-cut updating of $\underline{P}$ is well-defined, and $\mathcal{M}(\underline{P}_\alpha) = \{P \in \mathcal{P} : P(X) \geq x \text{ and } P(B) \geq \alpha\}$. Therefore, $\underline{P}_\alpha$ and $\underline{P}$ are equal if and only if $x \geq 2 + \alpha$. However, $\underline{P}_\alpha(\cdot \mid B)$ and $\underline{P}(\cdot \mid B)$ can be equal also when $\underline{P}_\alpha$ and $\underline{P}$ are not, and in fact, $\underline{P}_\alpha(\cdot \mid B)$ and $\underline{P}(\cdot \mid B)$ differ if and only if $2 - \alpha < x < 2$. That is, the α-cut updating of $\underline{P}$ corresponds to the regular extension updating when $x \notin (2 - \alpha, 2)$ (and thus also to the natural extension updating when $x \notin (2-\alpha, 2]$), with in particular $\underline{P}_\alpha(X \mid B) = 0$ when $x \leq 2-\alpha$, and $\underline{P}_\alpha(X \mid B) = x$ when $x \geq 2$. If $2 - \alpha < x < 2$, then the α-cut updating of $\underline{P}$ differs from the natural or regular extensions updating, with in particular $\underline{P}_\alpha(X \mid B) = {}^{x}/_{\alpha} + 2 - {}^{2}/_{\alpha}$.

Hence, the updated lower prevision $\underline{P}_\alpha(X \mid B)$ is a continuous, piecewise linear function of $x \in [1,3]$, which is plotted in Fig. 1 for some values of $\alpha \in (0,1)$. Since the slope of the central line segment is ${}^{1}/_{\alpha}$, the limit as α tends to 0 leads to the discontinuity of the regular extension $\underline{P}(X \mid B)$ at $x = 2$.

Example 2 (continued). Since $\overline{P}(B) > 0$, the α-cut updating of $\underline{P}$ is well-defined. The explicit consideration of $\underline{P}_\alpha$ is not necessary for calculating $\underline{P}_\alpha(X_{10} \mid B)$. In fact, the (relative) profile likelihood function

$$\lambda : p \mapsto \sup_{P \in \mathcal{M}(\underline{P})\,:\, P(B)>0,\, P(X_{10} \mid B)=p} \frac{P(B)}{\overline{P}(B)}$$

on $[0,1]$ can be easily obtained [11, Theorem 3], and $\underline{P}_\alpha(X_{10} \mid B) = \inf\{\lambda \geq \alpha\}$. That is, $\underline{P}_\alpha(X_{10} \mid B)$ is the infimum of the α-cut of the profile likelihood λ, consisting of all $p \in [0,1]$ such that $\lambda(p) \geq \alpha$.

The (relative) profile likelihood function λ is plotted in Fig. 2 for some values of $\varepsilon \in [0, {}^{1}/_{2}]$, together with the threshold $\alpha = 0.01$. In particular, $\underline{P}_{0.01}(X_{10} \mid B)$ has approximately the values 0.701, 0.704, 0.718, and 0.727, when ε is 0, 0.01, 0.05, and 0.1, respectively. Hence, there is no hint of any discontinuity of the updated lower prevision $\underline{P}_\alpha(X_{10} \mid B)$ as a function of $\varepsilon \in [0, {}^{1}/_{2}]$.

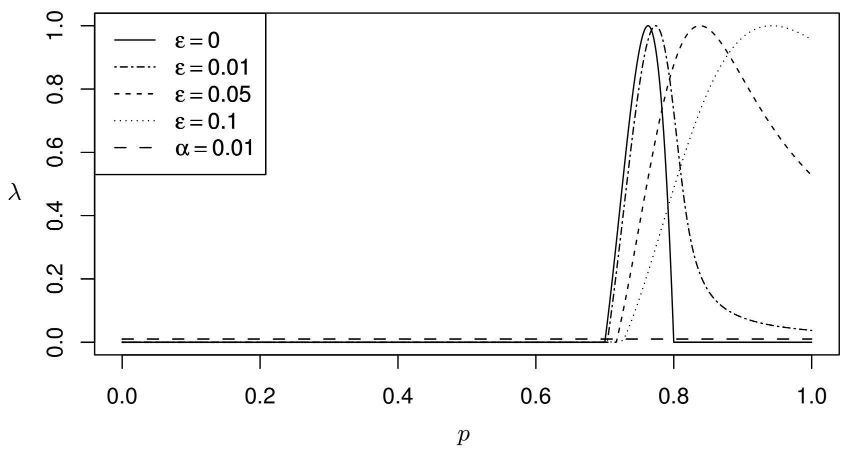

Fig. 2. Profile likelihood function of $P(X_{10} \mid B)$ from Example 2, for some values of $\varepsilon \in [0, {}^{1}/_{2}]$.

The next theorem shows that, contrary to the usual updating rules of the theory of imprecise probabilities, the α-cut updating rule is continuous on the set of all lower previsions $\underline{P}$ with $\overline{P}(B) > 0$, and thus avoids the robustness issues related to discontinuity of updating. Moreover, the α-cut rule has the huge advantage of allowing to use the vacuous prevision as prior imprecise prevision in statistical analyses, and still get non-trivial conclusions, while this is not possible with the natural or regular extensions. For these reasons, the α-cut updating rule has been successfully used in several practical applications of the theory of imprecise probabilities [17,18,19].

Theorem 3. *The α-cut updating rule $\underline{P} \mapsto \underline{P}_\alpha(\cdot \mid B)$ on $\{\underline{P} \in \underline{\mathcal{P}} : \overline{P}(B) > 0\}$ is continuous for all $\alpha \in (0,1)$.*

Proof. The α-cut updating rule is the composition of the two functions $\underline{P} \mapsto \underline{P}_\alpha$ on $\{\underline{P} \in \underline{\mathcal{P}} : \overline{P}(B) > 0\}$ and $\underline{P} \mapsto \underline{P}(\cdot \mid B)$ on $\{\underline{P} \in \underline{\mathcal{P}} : \underline{P}(B) > 0\}$. Theorem 2 implies that the latter is continuous, so only the continuity of the former remains to be proved.

Let $\delta \in \mathbb{R}_{>0}$ and let $\underline{P}, \underline{P}' \in \{\underline{P}'' \in \underline{\mathcal{P}} : \overline{P}''(B) > 0\}$ such that $d(\underline{P}, \underline{P}') < \delta$. If $P \in \mathcal{M}(\underline{P}_\alpha)$, then $P \in \mathcal{M}(\underline{P})$, and thus there is a $P' \in \mathcal{M}(\underline{P}')$ with $d(P, P') < \delta$. Moreover, there is also a $P'' \in \mathcal{M}(\underline{P}')$ such that $P''(B) = \overline{P}'(B)$ [4, Sect. 3.6]. Finally, define $P_\delta = \gamma\, P' + (1-\gamma)\, P''$, where

$$\gamma = \frac{2\,(1-\alpha)\,\overline{P}'(B)}{2\,(1-\alpha)\,\overline{P}'(B) + (\alpha+1)\,\delta}.$$

Then $P_\delta \in \mathcal{M}(\underline{P}'_\alpha)$, since $\mathcal{M}(\underline{P}')$ is convex, and $P_\delta(B) > \alpha\,\overline{P}'(B)$ is implied by $P'(B) > P(B) - \delta/2 \geq \alpha\,\overline{P}(B) - \delta/2 > \alpha\,\left(\overline{P}'(B) - \delta/2\right) - \delta/2$. That is, for each $P \in \mathcal{M}(\underline{P}_\alpha)$, there is a $P_\delta \in \mathcal{M}(\underline{P}'_\alpha)$ such that $d(P, P_\delta) < \gamma\,\delta + (1-\gamma)\,2$.

Since the roles of $\underline{P}$ and $\underline{P}'$ can be exchanged, $d(\underline{P}_\alpha, \underline{P}'_\alpha) \leq \gamma\,\delta + (1-\gamma)\,2$, and therefore the function $\underline{P} \mapsto \underline{P}_\alpha$ on $\{\underline{P} \in \underline{\mathcal{P}} : \overline{P}(B) > 0\}$ is continuous, because $\lim_{\delta \downarrow 0}(\gamma\,\delta + (1-\gamma)\,2) = 0$. □

4 Conclusion

In the present paper, the α-cut updating rule for imprecise probability models has been introduced. It is continuous and allows to use the vacuous prevision as prior imprecise prevision in statistical analyses. Therefore, it avoids the robustness problems of the usual updating rules of the theory of imprecise probabilities, and the difficulties related to the choice of near-ignorance priors.

As a consequence, the α-cut rule cannot always satisfy the property of coherence between conditional and unconditional previsions [20], since this property is incompatible with both the continuity of the updating and the successful use of vacuous priors in statistical analyses. The relative importance of these properties depends on the application field and on the exact interpretation of imprecise probabilities, as will be discussed in future work.

Anyway, the α-cut updating rule is not iteratively consistent. That is, when the updating is done in several steps, the order in which the data come in can influence the result. This iterative inconsistency can be resolved if the imprecise probability models are generalized by including also the second-order likelihood functions as part of the models [21,11]. The corresponding generalization of the α-cut updating rule remains continuous, as will also be discussed in future work.

Acknowledgments. The author wishes to thank the anonymous referees and the participants of WPMSIIP 2013 for their valuable comments and suggestions.

References

1. de Finetti, B.: Theory of Probability. 2 volumes. Wiley (1974–1975)
2. Savage, L.J.: The Foundations of Statistics. 2nd edn. Dover (1972)
3. Bhaskara Rao, K.P.S., Bhaskara Rao, M.: Theory of Charges. Academic Press (1983)
4. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall (1991)
5. Cattaneo, M.: On the robustness of imprecise probability methods. In Cozman, F., Denœux, T., Destercke, S., Seidfenfeld, T., eds.: ISIPTA '13, SIPTA (2013) 33–41
6. Piatti, A., Zaffalon, M., Trojani, F., Hutter, M.: Limits of learning about a categorical latent variable under prior near-ignorance. Int. J. Approx. Reasoning **50** (2009) 597–611
7. Walley, P.: Inferences from multinomial data: learning about a bag of marbles. J. R. Stat. Soc., Ser. B **58** (1996) 3–57
8. Dunford, N., Schwartz, J.T.: Linear Operators. 3 volumes. Wiley (1958–1971)
9. Škulj, D., Hable, R.: Coefficients of ergodicity for Markov chains with uncertain parameters. Metrika **76** (2013) 107–133
10. Fan, K.: Minimax theorems. Proc. Natl. Acad. Sci. USA **39** (1953) 42–47
11. Cattaneo, M.: A generalization of credal networks. In Augustin, T., Coolen, F.P.A., Moral, S., Troffaes, M.C.M., eds.: ISIPTA '09, SIPTA (2009) 79–88
12. Moral, S., de Campos, L.M.: Updating uncertain information. In Bouchon-Meunier, B., Yager, R.R., Zadeh, L.A., eds.: Uncertainty in Knowledge Bases, Springer (1991) 58–67
13. Moral, S.: Calculating uncertainty intervals from conditional convex sets of probabilities. In Dubois, D., Wellman, M.P., D'Ambrosio, B., Smets, P., eds.: UAI '92, Morgan Kaufmann (1992) 199–206
14. Gilboa, I., Schmeidler, D.: Updating ambiguous beliefs. J. Econ. Theory **59** (1993) 33–49
15. Moral, S., Wilson, N.: Revision rules for convex sets of probabilities. In Coletti, G., Dubois, D., Scozzafava, R., eds.: Mathematical Models for Handling Partial Knowledge in Artificial Intelligence, Springer (1995) 113–127
16. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
17. Antonucci, A., Cattaneo, M., Corani, G.: Likelihood-based robust classification with Bayesian networks. In Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R., eds.: Advances in Computational Intelligence. Volume 3. Springer (2012) 491–500
18. Cattaneo, M., Wiencierz, A.: Likelihood-based Imprecise Regression. Int. J. Approx. Reasoning **53** (2012) 1137–1154
19. Destercke, S.: A pairwise label ranking method with imprecise scores and partial predictions. In Blockeel, H., Kersting, K., Nijssen, S., Filip, Ž., eds.: Machine Learning and Knowledge Discovery in Databases, Springer (2013) 112–127
20. Miranda, E.: Updating coherent previsions on finite spaces. Fuzzy Sets Syst. **160** (2009) 1286–1307
21. Cattaneo, M.: Fuzzy probabilities based on the likelihood function. In Dubois, D., Lubiano, M.A., Prade, H., Gil, M.A., Grzegorzewski, P., Hryniewicz, O., eds.: Soft Methods for Handling Variability and Imprecision, Springer (2008) 43–50