Rough Approach to Fuzzification and Defuzzification in Probability Theory

G. Cattaneo and D. Ciucci
Dipartimento di Informatica, Sistemistica e Comunicazione Università di Milano-Bicocca, Via Bicocca degli Arcimboldi 8, I–20126 Milano (Italy) {cattang, ciucci}@disco.unimib.it

Abstract

An interesting procedure of fuzzification of a probability measure is given, producing another probability measure, whose rough approximation from the bottom, by a sharp inner measure, and the top, by an outer measure, is discussed.

Keywords: Rough Approximation, Topological Space, Probability Theory, Fuzzification.

1 Introduction

In the usual approach to fuzzy sets from a universe \(X\), crisp sets are described by characteristic functionals (also indicator functions) \(\chi_A : X \rightarrow \{0, 1\}\) of subsets of the universe. These are Boolean valued functions which assume value 0 (resp., 1) in every point \(x \in A\) (resp., \(x \notin A\)). Fuzzy sets are then described by generic \([0, 1]\)–valued functionals \(f : X \rightarrow [0, 1]\), where for every \(x \in X\) the number \(f(x) \in [0, 1]\) represents the membership degree of the point \(x\) to the fuzzy set \(f\). In general \(f\) is not considered as a possible fuzzy representation of some crisp set.

The notion of rough approximation space is defined as a set whose points represent vague, uncertain elements which can be approximated by the bottom and the top from crisp elements. In the particular case of fuzzy set theory, if one considers as approximable elements just fuzzy sets and as definable elements just characteristic functionals, then it is possible to assign to any fuzzy set \(f\) the two subsets \(E_1(f) := \{x \in X : f(x) = 1\}\) (the necessity domain of \(f\)) and \(E_p(f) := \{x \in X : f(x) \neq 0\}\) (the possibility domain of \(f\)). \(\chi_{E_1}(f)\) (resp., \(\chi_{E_p}(f)\)) is the best approximation from the bottom (resp., top) by crisp sets. In this rough context one can say that \(f\) is the fuzzy representative of any subset \(A \subseteq X\) such that \(E_1(f) \subseteq A \subseteq E_p(f)\).

In this paper, the “principle” of considering a fuzzy entity as representative of any crisp one by a suitable pair of inner and outer approximations is applied to probability theory based on a measurable space. For any event \(E\) (measurable subset of the whole space \(X\)), its probability of occurrence can be expressed as the integral

\[
p(E) = \int_X \chi_E \, dp \in [0, 1] \quad (1)
\]

where \(p\) is the probability measure. A “fuzzification” of this procedure can be obtained substituting to this characteristic function some fuzzy set \(\omega_E : X \rightarrow [0, 1]\), for any possible event \(E\), obtaining in this way a new occurrence probability

\[
p^*(E) = \int_X \omega_E \, dp \in [0, 1] \quad (2)
\]
Suitable conditions about $\omega_E$ discussed in section 3 (in particular the fact that $\omega_E$ must be a fuzzy representative of $E$ according to the above introduced rough approximations), lead to a new “fuzzy” probability measure. Thus, and differently from some widely considered approaches to fuzzy probability measure in which some of the standard Kolmogorov axioms are weakened (see for instance [DP80, chap. 5]), we have a new and genuine probability measure obtained by the fuzzification of $\chi_E$ in (1). This different approach to fuzzy probability seems to us very promising, especially from the statistical point of view as will be discussed in a forthcoming paper.

2 Abstract Rough Approximation Spaces

According to the abstract approach to roughness introduced in [Cat98], let us define the notion of approximation space for rough theories according to the following.

A rough approximation space is a structure

$$\mathcal{R} := (\Sigma, \mathcal{I}(\Sigma), \mathcal{O}(\Sigma), \leq, 0, 1)$$

where:

(1) $(\Sigma, \wedge, \vee, 0, 1)$ is a complete lattice with respect to the partial order relation $\leq$, bounded by the least element 0 $(\forall x \in \Sigma, 0 \leq x)$ and the greatest element 1 $(\forall x \in \Sigma, x \leq 1)$. Elements from $\Sigma$ are interpreted as concepts, data, etc., and are said to be the approximable elements;

(2) $\mathcal{I}(\Sigma)$ and $\mathcal{O}(\Sigma)$ are sublattices of $\Sigma$ whose elements are called inner and outer definable respectively.

The structure satisfies the following conditions.

(Ax1) For any approximable element $x \in \Sigma$, there exists (at least) one element $i(x)$ such that: (In1) $i(x) \in \mathcal{I}(\Sigma)$; (In2) $i(x) \leq x$; (In3) $\forall \alpha \in \mathcal{I}(\Sigma)$, $(\alpha \leq x \Rightarrow \alpha \leq i(x))$.

(Ax2) For any approximable element $x \in \Sigma$, there exists (at least) one element $o(x)$ such that: (Ou1) $o(x) \in \mathcal{O}(\Sigma)$; (Ou2) $x \leq o(x)$; (Ou3) $\forall \gamma \in \mathcal{O}(\Sigma)$, $(x \leq \gamma \Rightarrow o(x) \leq \gamma)$.

Therefore, $i(x)$ [resp., $o(x)$] is the best approximation of the “vague”, “imprecise”, “uncertain” element $x$ from the bottom [resp., top] by inner [resp., outer] definable elements.

For any approximable element $x \in \Sigma$, the inner and the outer definable elements $i(x) \in \mathcal{I}(\Sigma)$ and $o(x) \in \mathcal{O}(\Sigma)$, whose existence is assured by (Ax1) and (Ax2), are unique. Thus it is possible to introduce in an equivalent way a rough approximation space based on a triple $(\Sigma, \mathcal{I}(\Sigma), \mathcal{O}(\Sigma))$, consisting of a complete lattice and two sublattices of its, under the assumption of the existence of an inner approximation mapping $i : \Sigma \mapsto \mathcal{I}(\Sigma)$ and an outer approximation mapping $o : \Sigma \mapsto \mathcal{O}(\Sigma)$, given for an arbitrary $x \in \Sigma$ respectively by the laws

$$i(x) := \max\{\alpha \in \mathcal{I}(\Sigma) : \alpha \leq x\} \quad (3)$$
$$o(x) := \min\{\gamma \in \mathcal{O}(\Sigma) : x \leq \gamma\}. \quad (4)$$

The rough approximation of any approximable element $x \in \Sigma$ is then the ordered inner–outer pair

$$r(x) := (i(x), o(x)) \in \mathcal{I}(\Sigma) \times \mathcal{O}(\Sigma) \quad (5)$$

with $i(x) \leq x \leq o(x)$, which is the image of the element $x$ under the rough approximation mapping $r : \Sigma \mapsto \mathcal{I}(\Sigma) \times \mathcal{O}(\Sigma)$ pictured by the following diagram:
Following [Cat98], an element \( e \) of \( X \) is said to be crisp (also exact, sharp) if and only if its inner and outer approximations coincide: \( i(e) = o(e) \), equivalently, iff its rough approximation is the trivial one \( r(e) = (e,e) \). Owing to (In1) and (Ou1) this happens iff \( e \) is simultaneously an inner and an outer definable element; therefore, \( \mathcal{I}(\Sigma) \cap \mathcal{O}(\Sigma) \) is the collection of all crisp elements, which is not empty since \( 0,1 \in \mathcal{I}(\Sigma) \cap \mathcal{O}(\Sigma) \). In the sequel the set of all crisp elements will be denoted by \( \mathcal{I} \mathcal{O}(\Sigma) \).

### 3 Topological Rough Approximation Spaces

A first concrete example of rough approximation space is based on the notion of topological space [Kel55], in this context called topological rough approximation space. To this purpose, let us consider a topological space defined as a pair \( (X, \mathcal{O}(X)) \) consisting of a nonempty set \( X \) equipped with a family of open subsets \( \mathcal{O}(X) \). As well known, a subset of \( X \) is said to be closed iff it is the set theoretic complement of an open set. In this framework one can consider the structure

\[
\mathcal{R}_T = \langle \mathcal{P}(X), \mathcal{O}(X), \mathcal{C}(X), i, o \rangle
\]

where \( \mathcal{P}(X) \) is the power set of \( X \), collection of all its subsets, which is a complete lattice \( \langle \mathcal{P}(X), \cap, \cup, c, \emptyset, X \rangle \) with respect to set theoretic intersection \( \cap \), union \( \cup \), and the set theoretic complementation \( c \); this lattice is bounded by the least element \( \emptyset \) and the greatest element \( X \). Trivially, for any subset \( A \) of \( X \) it is possible to introduce the following definitions:

\[
i(A) : = \bigcup \{ O \in \mathcal{O}(X) : O \subseteq A \} \in \mathcal{O}(X)
\]

\[
o(A) : = \bigcap \{ C \in \mathcal{C}(X) : A \subseteq C \} \in \mathcal{C}(X)
\]

In other words, \( i(A) \) (resp., \( o(A) \)) is the topological interior (resp, closure), usually denoted by \( A^o \) (resp., \( A^* \)), of the set \( A \). In particular, \( A^o \) (resp., \( A^* \)) is the open (resp., closed) set which furnishes the best approximation of the approximable subset \( A \) of \( X \) by open (resp., closed) subsets from the bottom (resp., the top), i.e., it is the rough inner (resp., outer) approximation of \( A \).

The topological rough approximation mapping \( r_T : \mathcal{P}(X) \mapsto \mathcal{O}(X) \times \mathcal{C}(X) \) is the mapping which assigns to any subset \( A \) of the topological space \( X \) the pair \( r_T(A) = (A^o, A^*) \), with \( A^o \subseteq A \subseteq A^* \), consisting of its interior (open subset \( A^o \in \mathcal{O}(X) \)) and its closure (closed subset \( A^* \in \mathcal{C}(X) \)). Trivially, a subset \( E \) of a topological space \( X \) is crisp (exact, sharp) iff it is clopen (in particular the empty set and the whole space are clopen, and so exact).

#### 3.1 The Pawlak Approach to Rough Set Theory

The usual approach to rough set theory as introduced by Pawlak [Paw82, Paw92] is formally based on a pair \( (X, \pi(X)) \) consisting of a nonempty set \( X \), the universe [with corresponding power set \( \mathcal{P}(X) \), the collection of approximable sets], and a partition \( \pi(X) := \{ M_i \in \mathcal{P}(X) : i \in I \} \) of \( X \) whose elements are the elementary sets. The partition \( \pi(X) \) can be characterized by the induced equivalence relation \( \mathcal{R} \subseteq X \times X \), defined as \( (x,y) \in \mathcal{R} \) iff \( \exists M_j \in \pi(X) : x,y \in M_j \); in this case \( x,y \) are said to be indistinguishable with respect to \( \mathcal{R} \) and the equivalence relation \( \mathcal{R} \) is called an indistinguishability relation.

A definable set (or simple proposition) is any subset of \( X \) obtained as the set the-
oretic union of elementary subsets: \( M_j = \bigcup \{ M_j \in \pi(X) : j \in J \subseteq I \} \); the collection of all such definable sets plus the empty set (also called the trivial definable set) will be denoted by \( \Pi(X) \) and it turns out to be a Boolean algebra \( \langle \Pi(X), \cap, \cup, \setminus, \emptyset, X \rangle \) with respect to set theoretic intersection, union, and complementation. This Boolean algebra is atomic whose atoms are just the elementary sets (i.e., subsets from the partition \( \pi(X) \)).

From the topological point of view \( \Pi(X) \) contains both the empty set and the whole space, moreover it is closed with respect to arbitrary union and intersection, i.e., it is a family of clopen subsets for a topology on \( X \): \( \Pi(X) = \mathcal{O}(X) = \mathcal{C}(X) \). In this way we can construct the concrete approximation space \( \langle \mathcal{P}(X), \Pi(X), \Pi(X), i, o \rangle \), simply written as \( \langle X, \Pi(X), i, o \rangle \), consisting of:

1. the Boolean (complete) lattice \( \mathcal{P}(X) \) of all approximable subsets of the universe \( X \) whose atoms are the singletons;

2. the Boolean (complete) lattice \( \Pi(X) \) of all definable subsets of \( X \);

3. the inner approximation map \( i : \mathcal{P}(X) \mapsto \Pi(X) \) associating with any approximable set \( H \) its inner approximation \( i(H) \) defined by the (clopen)

\[
i(H) := \bigcup \{ M_j \in \Pi(X) : M_j \subseteq H \}
\]

4. the outer approximation map \( o : \mathcal{P}(X) \mapsto \Pi(X) \) associating with any approximable set \( H \) its outer approximation \( o(H) \) defined by the (clopen)

\[
o(H) := \bigcap \{ M_j \in \Pi(X) : H \subseteq M_j \} = \bigcup \{ M_j \in \Pi(X) : H \cap M_j \neq \emptyset \}
\]

The rough approximation of an approximable set \( H \) is the clopen pair

\[
r(H) := (i(H), o(H))
\]

with \( i(H) \subseteq H \subseteq o(H) \).

4 Fuzzification in Probability Theories

Let us consider a measurable space \( (X, \mathcal{M}(X)) \) consisting of a nonempty set \( X \) and a collection \( \mathcal{M}(X) \) of its subsets, the measurable subsets, which:

- (M1) contains the whole space \( X \);
- (M2) is closed with respect to countable union;
- (M3) is closed with respect to set theoretic intersection.

So \( \mathcal{M}(X) \) contains also the empty set and is closed with respect to the countable intersection and is said to be a \( \sigma \)-algebra of subsets of \( X \). The power set \( \mathcal{P}(X) \) satisfies the above conditions (M1)–(M3) and so it is the trivial \( \sigma \)-algebra, in which any subset of \( X \) is a measurable set. The pair \( (X, \mathcal{P}(X)) \) is called the trivial measurable space.

If one considers any family of measurable spaces \( \{ (X_j, \mathcal{M}_j(X)) \} \) and selects the collection of all subsets of \( X \) which are common to every \( \mathcal{M}_j(X) \) (i.e., the collection \( \bigcap_j \mathcal{M}_j(X) \), which contains both \( \emptyset \) and \( X \), then this collection satisfies all the conditions (M1)–(M3) and so the pair \( (X, \bigcap_j \mathcal{M}_j(X)) \) is a measurable space. If one has a generic family \( \mathcal{A}(X) \) of subsets of \( X \) and considers \( \bigcap \{ \mathcal{M}_j(X) : \mathcal{A}(X) \subseteq \mathcal{M}_j(X) \} \), this is the minimal, also the Borel, \( \sigma \)-algebra generated by \( \mathcal{A}(X) \). In the particular case in which \( \mathcal{A}(X) \) is the collection \( \mathcal{O}(X) \) of all open subsets of a topological space, the Borel \( \sigma \)-algebra generated by \( \mathcal{O}(X) \), also denoted by \( \mathcal{O}_B(X) \), contains all the countable intersections of open sets (which in general are not open) and all the countable union of closed sets (which in general are not closed). The pair \( (X, \mathcal{O}_B(X)) \) is the measurable space generated by the topological rough approximation environment.

A probability space is a triple \( (X, \mathcal{M}(X), p) \) consisting of a measurable space \( (X, \mathcal{M}(X)) \)
and a probability measure \( p : \mathcal{M}(X) \to [0, 1] \) which satisfies the conditions

(P1) \( p(X) = 1 \);

(P2) \( p(\cup E_n) = \sum p(E_n) \) for any countable family \( \{ E_n \in \mathcal{M}(X) : n \in \mathbb{N} \} \) of mutually disjoint \( (E_i \cap E_j = \emptyset \text{ for } i \neq j) \) measurable subsets of \( X \).

In this context, measurable subsets \( E \in \mathcal{M}(X) \) are called events and the quantity \( p(E) \in [0, 1] \) is the probability of occurrence of the event \( E \). Let us recall that the characteristic function \( \chi_A : X \to \{0, 1\} \) of a generic subset \( A \) of \( X \) is said to be measurable iff for any measurable Boolean subset \( \Delta \subseteq \{0, 1\} \), the subset \( \chi^{-1}_A(\Delta) := \{ x \in X : \chi_A(x) \in \Delta \} \subseteq X \) is measurable. Then, it is simple to prove that \( \chi_A \) is a measurable function iff the subset \( A \) is measurable (an element of \( \mathcal{M}(X) \)).

Differently from a characteristic function in which \( E \) is fixed, if one also considers \( E \) as running on \( \mathcal{M}(X) \) it is possible to introduce the mapping

\[
\chi : X \times \mathcal{M}(X) \to \{0, 1\} \tag{6}
\]

defined as

\[
\chi(x, E) = \begin{cases} 
1 & \text{iff } x \in E \\
0 & \text{otherwise}
\end{cases}
\]

This mapping is a crisp Markov kernel since:

(C1) for any fixed event \( E \in \mathcal{M}(X) \) the mapping \( \chi_E(x) := \chi(x, E) \), defined for \( x \) running in \( X \), is the measurable characteristic function of \( E \);

(C2) for any fixed point \( x \in X \) the mapping \( \chi_x(E) := \chi(x, E) \), defined for \( E \) running on \( \mathcal{M}(X) \), is a probability measure, the so called Dirac measure centered in the point \( x \).

A fuzzification of the mapping (6) is a Markov kernel, i.e., a mapping

\[
\omega : X \times \mathcal{M}(X) \to [0, 1] \tag{7}
\]

satisfying the following conditions:

(CF1) for any fixed event \( E \in \mathcal{M}(X) \) the mapping \( \forall x \in X, \omega_E(x) := \omega(x, E) \) is a measurable function on \( X \).

(CF2) For any fixed point \( x \in X \) the mapping \( \forall E \in \mathcal{M}(X), \omega_x(E) := \omega(x, E) \) is a probability measure on \( \mathcal{M}(X) \), i.e.,

\[
\text{(PM1) } \omega_x(X) = \omega(x, X) = 1, \quad \text{(PM2) } \omega_x(\cup_n E_n) = \omega(x, \cup E_n) = \sum_n \omega(x, E_n) \quad \text{for any countable collection of disjoint events } \{ E_n \}.
\]

(CF3) For any event \( E \) of the measurable space \( \mathcal{M}(X) \) it is

\[
\{ x \in X : \omega(x, E) = 1 \} \subseteq E
\]

and

\[
E \subseteq \{ y \in X : \omega(y, E) \neq 0 \}
\]

For instance, a standard way to obtain a Markov kernel on \( \mathbb{R} \) equipped with the Lebesgue measure consists in considering distribution functions \( f_x : \mathbb{R} \to \mathbb{R}, y \to f_x(y) \), centered in the points \( x \in \mathbb{R} \) (i.e., such that \( \forall y \in \mathbb{R}, f_x(y) \geq 0 \) and \( \int_\mathbb{R} f_x(y) dy = 1 \) and defining \( \omega(x, E) := \int_E f_x(y) dy \).

The event \( E_1(\omega) := \{ x \in X : \omega(x, E) = 1 \} = \omega^{-1}_E(\{1\}) \) is the necessity domain of the fuzzy representation \( \omega_E : \mathbb{R} \to [0, 1] \) of the event \( E \) generated by \( \omega \), whereas the event \( E_p(\omega) := \{ y \in X : \omega(y, E) \neq 0 \} = \omega^{-1}_E(0, 1] \) is the possibility domain of the fuzzy representation \( \omega_E : \mathbb{R} \to [0, 1] \) of the event \( E \) generated by \( \omega \). The condition (C3) expresses the regularity condition that the fuzzification \( \omega_E \) of the event \( E \) must be such that \( E_1(\omega) \subseteq E \subseteq E_p(\omega) \), with both \( E_1(\omega) \) and \( E_p(\omega) \) events.

Whatever be the event \( E \) one has the trivial chain of inequalities:

\[
\chi_{E_1(\omega)}(x) \leq \omega_{E}(x) \leq \chi_{E_p(\omega)}(x) \tag{8}
\]

The “fuzzy event” \( \omega_E : X \to [0, 1] \) can be considered the unsharp realization of the crisp event \( \chi_E \) (i.e., of the event \( E \)) by \( \omega \),
with $\omega_E(x)$ describing the degree of membership of the point $x$ to the event $E$. Let us denote by $F(\omega) := \{\omega_E \in [0,1]^X : E \in \mathcal{M}(X)\}$, the collection of all measurable fuzzy events generated by $\omega$ according to (CF1), then one can introduce the two mappings $i : F(\omega) \mapsto \{0,1\}^X$ defined by the law $i(\omega_E) := \chi_{E_i(\omega)}$ and $o : F(\omega) \mapsto \{0,1\}^X$ defined by the law $o(\omega_E) := \chi_{E_o(\omega)}$. In this way we have constructed the rough approximation space $(\mathcal{F}(\omega), \{0,1\}^X, \{0,1\}^X, i, o)$, simply written as $(\mathcal{F}(\omega), \{0,1\}^X, i, o)$, consisting of

1. the collection of all approximable fuzzy events $\mathcal{F}(\omega)(X)$;
2. the collection of all definable crisp events $\{0,1\}^X$;
3. the inner approximation mapping $i$ and the outer approximation mapping $o$.

By the Markov kernel $\omega$ the “crisp” probability (1)

$$p^\omega(E) = \int_X \omega_E \, dp \in [0,1]$$

can be extended in the fuzzy way according to (2)

$$p(E) = \int_X \chi_E \, dp \in [0,1]$$

which turns out to be a probability measure on $(X, \mathcal{M}(X))$ since $p^\omega(\emptyset) = \int_X \omega_0 \, dp = 0$ and $p^\omega(\bigcup_i E_i) = \int_X \omega_{\bigcup_i E_i} \, dp = \sum_n \int_X \omega_{E_n} \, dp = \sum_n p^\omega(E_n)$ for any countable family $\{E_n\}$ of disjoint measurable subsets of $X$.

Therefore, the now discussed procedure gives rise to a new probability measure $p^\omega : \mathcal{M}(X) \mapsto [0,1]$ generated by $\omega$ on the basis of the “crisp” probability measure $p$ and acting to “unsharp” events $\omega_E$ from the measurable space $\mathcal{M}(X)$. This is a genuine probability measure which corresponds to the fuzzification procedure (2) of the original probability measure (1), this latter expressed by crisp representation of sets using characteristic functions. As pointed out in the introduction, this is a different way to obtain fuzzy measures as discussed in [DP80, chap. 5] in the more general formulation given by Sugeno [Sug77].

## 5 Rough Defuzzification on Probability theories

Now, making use of the possibility domain $E_p(\omega) = \omega^{-1}(0,1]$, we can introduce

$$\overline{p}^\omega(E) = \int_X \chi_{E_{\omega_p^{-1}(0,1]}} \, dp = (1) = p(\omega_E^{-1}(0,1])$$

which is an outer probability measure, i.e.,

(OM1) $\overline{p}^\omega(X) = 1$;

(OM2) $E_1 \subseteq E_2$ implies $\overline{p}^\omega(E_1) \leq \overline{p}^\omega(E_2)$;

(OM3) $\overline{p}^\omega(\bigcup_i E_i) \leq \sum_i \overline{p}^\omega(E_i)$ for any family of disjoint measurable sets.

Similarly, making use of the necessity domain $E_n(\omega) = \omega^{-1}(1)$, it is possible to introduce

$$\underline{p}^\omega(E) = \int_X \chi_{E_{\omega_n^{-1}(1)}} \, dp = (2) = p(\omega_E^{-1}(1])$$

which is an inner probability measure, i.e.,

(IPM1) $\underline{p}^\omega(X) = 1$;

(IPM2) $E_1 \subseteq E_2$ implies $\underline{p}^\omega(E_1) \leq \underline{p}^\omega(E_2)$;

(IPM3) $\underline{p}^\omega(\bigcup_i E_i) \geq \sum_i \underline{p}^\omega(E_i)$ for any family of disjoint measurable sets.

Trivially, from (8) we have the following “rough” approximation:

$$\overline{p}^\omega(E) \leq p^\omega(E) \leq \overline{p}^\omega(E)$$

of the unsharp probability occurrence $p^\omega(E)$ of the event $E$ by the “inner” probability from the bottom $\underline{p}^\omega(E)$ and the “outer” probability from the top $\overline{p}^\omega(E)$.

In other words, the “fuzzy probability” $p^\omega(E)$ of the “crisp event” $E$ is lower (resp.,
upper) approximated by the “crisp probability” $p$ of the “fuzzy event” $E_1(\omega)$ (resp., $E_p(\omega)$) which has a global behavior of “inner” (resp., outer) probability measure.

References


