Types of Probabilities Associated with Rough Membership Functions

A. Mani
Department of Pure Mathematics
University of Calcutta
9/1B, Jatin Bagchi Road
Kolkata(Calcutta)-700029, India
Email: a.mani.cms@gmail.com
Homepage: http://www.logicamani.in

Abstract—The intent of this research is to critically analyze probabilistic rough sets (PSTs) and methods and propose new approaches. Towards this objective connections between meta theoretical approaches to probability and rough membership functions are reviewed and variants are proposed in this research. The contamination problem has been proposed and studied by the present author across a number of papers. The scope of the problem within PSTs is also explained in some depth by her. A new definition of artificial intelligence applicable in rough perspectives is also proposed on the basis of recent advances in algebraic semantics related to rough membership functions. All this is intended to eventually bridge the gap between theory and practice in these fields.

Keywords: Rough Sets, Probability, Membership Functions, Axiomatic Granules, Contamination Problem, Epistemology.

I. INTRODUCTION

The literature on rough membership functions (RMF) (for example [1], [2]) is vast due to its practical applications. Studies in Bayesian and non-Bayesian probabilistic perspectives of rough sets include [1], [3], [4]. At the practical level, the contamination problem introduced and explored by the present author in [5], [6], is a less intrusive approach for rough modeling of data and vagueness. From this perspective, approaches to probabilistic interpretations of rough sets are reviewed and new problems and methodologies are posed by the present author. Recent progress on ontology of probability as in [7]–[9] are taken into account for the purpose.

In [10], a semantics of rough membership functions has been recently advanced. Technical omissions in the paper are rectified and the interpretation of the results are fixed from the perspective of the contamination problem. The omission of the induced topological operators is also questioned. A fall out of the reinterpretation is a very technical definition of artificial intelligence in the context of rough semantics of information systems.

Three-way decision making has been integrated with PST in [11]. It is a unique approach as three-way decision making is not well known in other areas of soft computing/machine learning. Related ontology is critically reviewed in the present paper as the approach is essentially related to rough membership functions and variants.

A. Some Background

An Information System \( I \), is a system of the form

\[
I = (S, At, \{V_a : a \in At\}, \{f_a : a \in At\})
\]

with \( S, At \) and \( V_a \) being respectively sets of Objects, Attributes and Values respectively. The approximation spaces and generalizations thereof of RST arise from relational or relator spaces derived from these. For example, \( x, y \in S \) and \( B \subseteq At \), let \((x, y) \in R \) if and only if \( (\forall a \in B) v(a, x) = v(a, y) \), then \( R \) is an equivalence relation on \( S \). \((S, R)\) is said to be an approximation space.

In classical RST, on the power set \( p(S) \), lower and upper approximations of a subset \( A \in p(S) \) operators, apart from the usual Boolean operations, are defined as per:

\[
A^l = \bigcup \{x : A^u = \bigcup \{x | x \cap A \neq \emptyset\},
\]

with \( [x] \) being the equivalence class generated by \( x \in S \). If \( A, B \in p(S) \), then \( A \) is said to be roughly included in \( B \) (\( A \subseteq B \)) if and only if \( A^l \subseteq B^l \) and \( A^u \subseteq B^u \). \( A \) is roughly equal to \( B \) (\( A \approx B \)) if and only if \( A \subseteq B \) and \( B \subseteq A \). The positive, negative and boundary region determined by a subset \( A \) are respectively \( A^l, A^u \) and \( A^u \setminus A^l \) respectively.

Given a fixed \( A \in p(S) \), a Rough membership function is a map \( f_A : S \mapsto [0, 1] \) that are defined (#(\)) being the cardinality function) via

\[
(\forall x) f_A(x) = \frac{\#(x \cap A)}{\#(x)}.
\]

Rough membership functions can be generalized to other general rough structures but lose many of the better properties valid in the classical context.

A pre-rough algebra \( PRA \) is an algebraic system of the form \( S = \langle S, \cap, \cup, \Rightarrow, L, \neg, 0, 1 \rangle \) of type \( (2, 2, 2, 1, 1, 0, 0) \) that satisfies:

- \( \langle S, \cap, \cup, \Rightarrow \rangle \) is a De Morgan lattice.
- \( \neg \neg a = a \); \( L(a) \cap a = L(a) \); \( L(a \cup b) = L(a) \cup L(b) \); \( \neg L \neg L(a) = L(a) \)
- \( LL(a) = L(a) \); \( L(1) = 1 \); \( L(a \cap b) = L(a) \cap L(b) \); \( \neg L(a) \cup L(a) = L(a) \).
\[ (L(a) \cap L(b) = L(a), -L(\neg(a \cap b))) = -L(\neg a) \rightarrow a \cap b = a \]
\[ a \Rightarrow b = (\neg L(a) \cup L(b)) \cup (L(\neg a) \cup \neg L(b)). \]

A completely distributive PRA is called a rough algebra. These algebras are examples of algebraic models of classical rough sets. The modality \( \Box \) is definable on PRA via \( \Box x = \neg L(\neg x) \) for each element \( x \).

For classical rough sets, the power set of an approximation space with operators forms an unsatisfactory semantics as rough and crisp objects are mixed up. This scenario holds for general approximation spaces (where \( R \) can be any binary relation). A more general structure would be a set with a parthood relation and approximation operators defined on it [5]. Such structures can be associated with special categories called classical semantic domains and are defined by restrictions on objects, predicates, functions, constants and low level operations in a language. The rough semantic domain is the semantic domain related to sets of roughly equivalent or rough objects. Apart from these two primary domains a number of hybrid semantic domains have been used in the literature (see for example [12]).

The basic problem of contamination is that of mix up of semantic notions during modeling the dynamics within a specific semantic domain. At the practical level it can be in using semantic aspects of the classical semantic domain in modeling interaction in the rough semantic domain - if somebody is reasoning about aggregating two rough objects, then they are not likely to know about the classical ontology associated with their awareness (\( \Box \) is an example of such an operation) and operations like union and intersection. But at the theoretical level many models assume as much. In other words in classical semantic domain operations and predicates used to describe semantics of specific rough semantic domains may not exist in the rough domain in the first place. Details can be found in [6], [13].

II. TYPES OF PROBABILITY THEORIES

It is necessary to review the various interpretations of probability [7]–[9] as interpretations comparable to probability and the frequentist interpretation have been extended to RSTs by various authors [4], [14], [15] in many ways. In all of the approaches, the use of rational-valued numeric measures invariably results in simplification at the cost of information loss and the connections with probability are not transparent. The question of where these generalizations stand is considered in the next section.

Subjective and Bayesian approach to probability involve epistemic assumptions while frequentist and variants try to be objective by ignoring additional information that are often available in the context. Relative such a view probability theory can be classified into (similar views can be found in [7]–[9]):

- **Epistemic Probability Theories**
  - 1) Frequentism Based including Classical Probability
  - 2) Propensity Based
  - 3) Subjective Non-Bayesian
  - 4) Inter-subjective

- **Grounded Probability Theories**

A. Epistemic Probability

Logical epistemic probability (or evidentialism) applies to contexts in which probabilities arise from rational beliefs that can be formalized and the associated logics concern partial entailment relations. An example of a rational belief is the principle of insufficient reason: Given no prior information and the fact that an experiment can potentially have \( N \) outcomes, then the probability of each outcome must necessarily be \( \frac{1}{N} \).

The principle of insufficient reason is one among the many symmetry principles used by evidentialists. Strict frequentists in contrast to evidentialists would not assign probabilities in the context of principle of insufficient reason without further empirical trials.

Applicability of symmetry principles are known to be of unclear justification especially when it is about quantities derived from available information: Let the possible outcomes be of \( k \) types with the \( i \)th type having attribute set \( A_i \) and let the parameter of interest be \( f(A_i) \) for any \( i \), \( f \) being a function). Then the probability of an outcome with parameters \( f(A) \) need not have a simple direct interpretation.

In subjective Bayesianism SB, the probability of a proposition is the degree of belief one has that the proposition is true. Coherence with Kolmogorov’s axioms of probability is a desirable feature, but is not always satisfied. When they do satisfy them, then SB’s difference with evidentialism becomes less clear -unless they believe the probabilities are some approximate numbers. For evidentialists there exist definite probabilities that correspond to their credences. The fundamental theorem of the subjective view is the Ramsey-De Finetti theorem:

**Theorem 1.** When the process is viewed from a gambling perspective, a collection of betting quotients is coherent iff the collection satisfies the axioms of probability.

Results of mathematical statistics like the Von Mises theorem of the form Under a lot of statistical assumptions, the posterior probabilities of any two agents would eventually become the same in a topological sense can also be used to justify the subjectivist position.

Of the many subjective non-Bayesian approaches, axiomatic approaches based on probabilisticic dependence as opposed to independence constitute an important approach. Abstract axiomatic versions of these have been considered for comparison with ideas of rough dependence in a contamination-free way by the present author in [6], [16] and it has been shown that the concepts are not comparable. The contamination-free aspect means that no membership functions were used in all this.

Statistics books can be rewritten in the dependence approach of [17]. The starting point of this is this definition: A dependence function in a probability space \( (X, S, p) \) over the set \( X \) (\( S \) is a \( \sigma \)-algebra over \( X \) and \( p \) is a probability function) is
a function $\delta : S^2 \rightarrow \mathcal{R}$ defined by

$$\delta(x, y) = p(x \cap y) - p(x) \cdot p(y).$$

The inter-subjective or consensus view is based on common beliefs of social groups and is intended to be a consensus between the logical and the subjective way. The consensus value of probability is said to be the inter-subjective probability. Clearly it can differ from the probability assigned by particular individuals. The basic idea behind it is to stick to consensus views in the matter of degrees of beliefs and this in turn is based on the following result:

**Theorem 2.** If a girl A is betting against $n$ other girls $H = \{H_i\}_{i=1}^n$ on an event $E$ and if $H_i$ selects betting quotient $b_i$, then $\hat{A}$ will be able to select stakes so that she gains money from $H$ in all cases except when $b_i = b_k$ for all $i, k$.

### B. Grounded Probability

Classical probability assigns probabilities based on physical or hypothetical idealized symmetry and is based on finite combinatorics and the principle of indifference. There are many versions of frequentism. The simplest view being that relative frequencies obtained empirically correspond to probabilities. So if five tosses of a coin produces five heads, then the probability of getting a head in a toss is 1. In hypothetical frequentism, the probability of an event is the limit of relative frequencies (when it exists) of an infinite sequence of trials. The probability of an outcome $A$ relative a collective $B$ is simply the limiting relative frequency of $A$ in $S$. Though intuitive, it fails $\sigma$-adaptivity axioms. Many statisticians however identify classical symmetry based probability with frequentism and not with the non-frequentist approaches.

Frequentism is well known to suffer from:

- Reference Class Problem: One needs a reference and when all events belong to multiple reference classes, the probabilities of non events may be definable.
- Problem of a Single Trial

One definition of probability defined by propensity is the following: If a repeated trial favors an outcome of a type with limiting relative frequency $a$, then the probability $a$ of the outcome of the type is the propensity of the experiment. This definition suffers from drawbacks similar to that of the limiting frequency approach. Other definitions violate other basic principles and so no clear workable definitions exist of the concept. In the present author’s opinion the basic problem of the propensity approach should be to find recursive or functional definitions of a predicate $\Pi(F, P, S)$ that can yield a set of relative frequencies $F$ from a set of probabilities $P$ that satisfy the probability axioms $S$ and vice versa. This would be a way of removing the vagueness inherent in the proposed definitions in the literature.

In the present author’s opinion, hypothetical frequentism has many aspects and in finite contexts, relative frequencies possess an ontology that includes the principle of indifference and includes classical symmetry based probability.

### III. Membership Function Perspective

The results of [10] pretty much support the contamination perspective because the results essentially mean that all the membership functions put together do not make a rough semantics. In the paper, the rough membership functions again are viewed as fuzzy sets with a hypothetical frequentist evolution as in [18]. Further as noted in [18], the semantics of fuzzy sets within RSTs is quite restricted and form a special class. The lower and upper approximations are shown to coincide with the core and support of a fuzzy set in the same paper. The negative aspect in the former paper is not so big a thing for the contamination perspective - the more important thing that needs to be understood both formally and informally is the nature of relative failure. It is also shown that the finiteness assumption in [10] can be relaxed. Rough membership functions can be interpreted in probabilistic perspectives in many ways and is not tied to a single view - in the subsections below an improved reinterpretation of the state of affairs is proposed by the present author.

The essential result in [10] is the following. Let $S = \langle S, R \rangle$ be a well-founded approximation space with $S$ being a set and $R \in EQ(S)$ (well foundedness is not explicitly required in the paper but finiteness of equivalence classes is required). Further $R$ can be any kind of equivalence - it need not be a well equivalence. Given a fixed $C \in p(S)$, a Rough membership function $\text{RMF}$ is a function $f_S : S \rightarrow [0, 1]$ defined by

$$\forall x f_S(x) = \frac{|\{x \mid x \cap C\}|}{|\{x\}|}.$$  

The Rough Membership Function Algebra $\text{RMFA}$ for a well founded approximation space $S$ is an algebra (defined on the set of all distinct $\text{RMF}$s $F$) of the form $F = \langle F, \wedge, \vee, \sim, f_0, f_S \rangle$ with the operations defined as per algorithms specified in Thm 2.8 and Thm 2.9 in the paper. $\sim f_B = 1 - f_B = f_{\sim B}$ and the partial order on $F$ is defined as per $f_P \leq f_Q \iff (\forall x) f_P(x) \leq f_Q(x)$.

It is shown that the above algebra is a Quasi Boolean Algebra and using the lower and upper approximation operators of RSTs, modal operators are defined via $\Box f_A = f_{A^\sim}$ and $\Diamond f_A = f_{A^\sim}$. It is shown that the resulting enhanced algebra fails to be a PRA because the following property does not hold:

$$\Box f_A \leq \Box f_B \& \Diamond f_A \leq \Diamond f_B \rightarrow f_A \leq f_B$$

and also that a logic which satisfies modal axioms (K, T, S4, B) is an appropriate logic for the algebra. In the present author’s view this logic mixes up three different semantic domains.

From the contamination perspective, $\text{RMFA}$ without the additional modal operators is of greater interest. The semantic assumptions are clearly in the classical frequentist domain. The introduction of modal operators mixes up reasoning from the attribute-value perspective. In the development of a semantics of membership functions in [10], the following topological aspect is not considered at all and neither is the relation to the relatively artificial modal operators mentioned.

**Proposition 1.** The following hold in any $\text{RMFA} F$:

- $F$ is a suposet of $\mathcal{F}(S, [0, 1])$ with the induced order.
• F has two nonequivalent topologies induced on it - one by the Alexandrov topology and the other by closure operator ♦.

Proof: The topology on $F(S, [0, 1]) = H$ (set of all fuzzy sets on $S$) is given by

$$\tau_S = \{ A : A \subseteq H \& (\forall x,y \in H)(x \in A, x \leq y \rightarrow y \in A) \}.$$ 

The closure of a subset $Q \subseteq H$ coincides with the $o$-ideal generated by $Q$ in $H$: $cl(Q) = \{x, (\exists y \in Q) x \leq y\}$. The other topology on $F$ is defined by the topological closure operator ♦.

IV. A DEFINITION OF ARTIFICIAL INTELLIGENCE IN ROUGH DOMAINS

In the present author’s perspective, the pre-rough semantics arises as a description of the possible operations on rough objects that in turn are defined by attributes and their associated values. In contrast in many rough semantics the membership functions are highly suspect measures when they are based on classical probabilisticic or frequentist perspective that apparently do not care about properties or attributes. The nature of the gap in the semantics of rough algebras and the RMFA are missing a theory. All this points to the real gap between the following definition is a reasonable one because it relates knowledge from very different perspectives that operate on the same information.

**Theorem 3.** Pre-rough algebras are the classical rough semantics of well approximation spaces

The construction of PRAs shows that it is not concerned about the size or fine structure of classes. Neither is a rough algebra concerned with those aspects and comparisons in [10] are missing a theory. All this points to the real gap between the frequentist perspective (of RMFA) and the property based one (of rough algebras). The rough algebra semantics is not perfect as the operations in it are not contamination free - this means a semantic domain between the classical semantic domain and rough semantic domain should be definable. Any program that can carry reasoning from the semantic domain of RMFA to the intermediate domain must be artificially intelligent (modulo the usual caveats of soft computing).

Definitions of artificial intelligence can only be relative and the following definition is a reasonable one because it relates knowledge from very different perspectives that operate on the same information.

**Definition 1.** A learning algorithm $L$ will be said to be artificially intelligent over classical rough domain if and only if its relative action on a RMFA (without the modal operators) $F_{rmf}(S)$ yields the rough algebra $S(S)$ on the well founded approximation space $S$. That is

$$L : F_{rmf}(S) \rightarrow S(S).$$

This is illustrated in the schematic diagram below with $IK$ being an abbreviation for Initial Knowledge and $EK$ for Enhanced Knowledge.

![Fig. 1: An AI By Roughness](image)

More aspects of this definition will be studied in a forthcoming paper.

V. TYPES OF PROBABILITY AND ROUGH MEMBERSHIP FUNCTIONS

Attempts at probabilistic interpretations of aspects of rough reasoning go back to the origins of rough sets. In [3], [11], some effort has been made towards specifying generalities over different types of probability interpretations for RSTs and probabilisticic rough sets (PST). Connections with subjective Bayesian reasoning have apparently been the dominant view in the literature on PSTs. Some papers in which such a viewpoint is used are [4], [11], [15]. Connections with Dempster-Shafer theory (and related generalized probabilities) have been explored in [1], [4] for example.

The reason rough membership functions being read as probability functions is frequentist in many cases as opposed to being a common notion shared by classical, frequentist and epistemic interpretations of probability are the following:

• More information about the attributes is actually available in every interpretation and the contextual application of the principle of indifference is not justified.
• The use of cardinalities of subsets actually distorts the impact of attributes and their valuations in the information system or decision table. The essential nature of order structure present in the rough data is lost as a consequence.
• The method of derivation of rough membership functions is not epistemological.
• The method of derivation of rough membership functions is objectivist and
• though the empirical aspect of frequentism is missing, computations rely on hypothetical frequentism.

These have been partly explained in [5], [6] by the present author. Rough membership functions and general versions thereof can also be arrived through other kinds of RSTs.
Relative membership of an object in \( C \) and the three way decision, there is a set \( \Omega = \{ C, C^* \} \) two states (corresponding to the object being in the set \( C \) or otherwise respectively). Note that \( C \) is used for the set and state. Each state can again have actions from the action set \( \mathcal{A} = \{ P, B, N \} \) associated, with \( P, B, N \) corresponding respectively to accepting \( e \in \text{Pos}(C) \), \( x \in \text{Bnd}(C) \) and \( x \in \text{Neg}(C) \) respectively. The associated loss function is intended as an expert view of the potential risk of cost of actions in different states, and forms a \( 3 \times 2 \) matrix:

<table>
<thead>
<tr>
<th>( C(P) )</th>
<th>( C(N) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_{PP} )</td>
<td>( L_{PN} )</td>
</tr>
<tr>
<td>( L_{BP} )</td>
<td>( L_{BN} )</td>
</tr>
<tr>
<td>( L_{NP} )</td>
<td>( L_{NN} )</td>
</tr>
</tbody>
</table>

\( \text{TABLE I: Labeled Loss Matrix} \)

\( L_{\text{PP}} \) and \( L_{\text{NN}} \) respectively are the losses involved for taking action \( i \) when the object belongs to \( C \) and does not belong to \( C \) respectively. For example, \( L_{PP} \) may be the risk involved in prescribing a medicine to a patient with a medical condition - in general, domain knowledge is necessary for arriving at reasonable values of the loss matrix.

From the above, the expected losses are estimable (under the assumptions) by the system

\[
\begin{align*}
R(P|e) &= L_{PP}p(C|e) + L_{PN}p(C^*|e), \\
R(B|e) &= L_{BP}p(C|e) + L_{BN}p(C^*|e), \\
R(N|e) &= L_{NP}p(C|e) + L_{NN}p(C^*|e).
\end{align*}
\]

The Bayesian rules of solving this correspond to

- If \( R(P|e) \leq R(N|e) \) and \( R(P|e) \leq R(B|e) \) (P) then \( e \in \text{Pos}(C) \).
- If \( R(B|e) \leq R(P|e) \) and \( R(B|e) \leq R(N|e) \) (B) then \( e \in \text{Bnd}(C) \).
- If \( R(N|e) \leq R(P|e) \) and \( R(N|e) \leq R(B|e) \) (N) then \( e \in \text{Neg}(C) \).

Under natural conditions like

\[ L_{PP} \leq L_{BP} \leq L_{NP} \& L_{NN} \leq L_{BN} \leq L_{PN}, \]

the decision rules can be simplified into (\[ 11 \])

- If \( \alpha \leq p(C|e) \) then \( e \in \text{Pos}(C) \). \( (P1) \)
- If \( \beta < p(C|e) < \alpha \) then \( e \in \text{Bnd}(C) \). \( (B1) \)
- If \( p(C|e) \leq \beta \) then \( e \in \text{Neg}(C) \). \( (N1) \)

This determines \( (\alpha, \beta) \)-positive, negative and boundary regions of its own as follows:

\[
\begin{align*}
\text{Pos}_{\alpha, \beta}(C) &= \{ e : e \in S \& \alpha \leq p(C|e) \}, \\
\text{Bnd}_{\alpha, \beta}(C) &= \{ e : e \in S \& \beta < p(C|e) < \alpha \}, \\
\text{Neg}_{\alpha, \beta}(C) &= \{ e : e \in S \& p(C|e) \leq \beta \}.
\end{align*}
\]

Related \( (\alpha, \beta) \)-probabilistic approximations are defined via

\[
\begin{align*}
C^{1, \alpha, \beta} &= \{ e : e \in S \& \alpha \leq p(C|e) \}, \\
C^{\alpha, 0, \beta} &= \{ e : e \in S \& \beta < p(C|e) \}.
\end{align*}
\]
This is just one of the many ways of moving from decision theoretic data expressed in terms of loss matrix into PSTs. Though in [11], it is claimed that loss functions provide a more intuitive way of understanding the error tolerance, the complications involved in the derivations of $\alpha, \beta$ from loss matrices suggest that a higher order statistical theory valid for at least some problem subclasses is necessary for theoretical ratification of such choice. At the empirical level, it is a matter of the context appearing to fit the computational semantics.

In short, the suggested theory is far from complete in a mathematical statistical perspective. In the present author’s view, the following problems are closely associated with the approach and needs to be investigated further:

- **Converse Problem**: Given a PST, find sufficient/minimal conditions in the form of rules for ensuring a unique risk based decision process - this is essential for consolidating the argument for claiming that some values of $\alpha$, $\beta$ are better than others.

- **Polemical Steps**: Apart from the steps relating to expert knowledge and their representation, the steps involving intrusive assumptions are in the steps relating to computation of $\alpha$, $\beta$, $\gamma$ from the loss matrix.

- **Hidden Rule Problems**: The suggested ontology means admissible sets of rules are definable over the decision/information system. From a AI perspective, modulo few assumptions, this would provide means of tracking the reasoning of experts that they have obviously not written down in black and white. This leads to two kinds of problems:
  - the problem of finding better logics using second order aspects of the information system,
  - and the problem of better quality risk measurement scales - this is natural in psychology where scales are routinely updated, but not through formal methods.

- **Membership Functions**: Preferences over general membership functions applicable to PSTs possible in a context arise as the result of the technique.

- **Type of Ontology**: The methodology fits the Subjective Bayesian one, but is also involved in providing a method of constructing better ‘classical relative probabilities’. The latter aspect suggests the possibility of direct estimation through extensions of counting procedures to rough contexts by the present author in [5].

An algebraic semantics, developed by the present author in [20], to general RSTs is essentially a semantics constructed from one such counting procedure.

VI. Remarks

In this research connections between rough membership functions and probability theories are made more clear and a number of new problems have been posed towards clearer understanding of contamination reduction and making rough analysis even less intrusive by the present author. Problems of applied nature relating to three way decision making have also been formulated by her. Recent results on algebraic semantics from rough membership functions have also been improved and a new definition of artificial intelligence, applicable to rough contexts derived from information systems or decision tables or more generally from information falling under object-property-value perspective have also been defined.

REFERENCES


