STATISTICAL CLASSES AND POSSIBILISTIC MODELS OF CLASSIFYING PROBABILITY DISTRIBUTIONS

Leonid S. Berstein, Andrei G. Bronevitch
Alexander N. Karkishchenko and Vladislav G. Zakharevitch

University of Radio-Engineering, Nekrasovskij bystr. 44, Taganrog, Russia

1. Introduction

It is shown by the experience of intelligent, control and decision support systems working out that the classical methods of a multidimensional statistical analysis are often useless because of their computational complexity and restrictions of the classical probability scheme [1]. In these cases it is expedient to use non-traditional methods of data processing that are in most cases based on the theory of fuzzy sets, particularly, on fuzzy measures of Sugeno [2] and the possibility theory of Zadeh [3].

However, there is a deep connection between the probability theory and the theory of fuzzy sets. Thus, any fuzzy set can be identified with a random set [4], a membership function of the fuzzy set coinciding with a cover function of a random set. The probabilistic interpretation can also be given to upper and lower belief measures in a set-possibilistic model. Also, we point out the works [5,6,7] where Dubois and Prade, developing the Dempster-Shafer theory of evidence, offer methods for constructing fuzzy sets upon statistical data.

A set-theoretic method of statistical data investigation proposed in this paper requires partial ordering of the set of probabilistic distributions. This algebraic structure is formalized with the help of an idea of a statistical class, introduced below, which can be considered as an empirical or a posteriori image of a certain probability space. On the set of the statistical classes inclusion and equality relations as well as inclusion and equality measures are defined. It enables us to solve the main problem of the given paper - to classify the statistical classes.

2. Basic definitions and problem formulating

Let X be a measurable space of elementary events together with a σ-algebra of events $\mathfrak{A}$. We will assume that there is a volume measure V upon $\mathfrak{A}$ satisfying the usual requirements of non-negativity and additivity. For any $A \in \mathfrak{A}$ the meaning of $V(A)$ can be interpreted as a power of the set A or as a closure of the event A to an elementary event.

We are given a probability measure P. The triple $F = (X, \mathfrak{A}, P)$ will be called as *a statistical class* on X. Thus, any statistical class on X is completely defined by assigning the probability measure P. We subsequently denote by $\mathfrak{F} = \{F_i \mid i = 1,2,...\}$ a finite or infinite aggregate of the statistical classes. Let us assume that there is a certain subset $S = \{S_1, S_2, ..., S_r\}$ of *standard* classes in $\mathfrak{F}$. Then the problem of classification the statistical classes on the standard ones consists in constructing the inclusion measure $\psi(F_1 \subseteq F_2) \in [0,1]$ upon $\mathfrak{F}$ with the help of which for any $F \in \mathfrak{F}$ there can be found a classifying vector

$$\left(\psi(F \subseteq S_1),\ \psi(F \subseteq S_2), \ldots,\ \psi(F \subseteq S_r)\right).$$

Consider a set $\mathcal{A}(p) = \{A \in \mathfrak{A} \mid P(A) = p\}$ of p-probable events. An event $E \in \mathcal{A}(p)$ is called *the minimal event (m.e.)* for a class F if the condition $V(E) = \inf\limits_{A \in \mathcal{A}(p)} V(A)$ holds. The set of all m.e. for a class F will be denoted as $\mathfrak{M}$. It should be emphasized that m.e. $E \in \mathfrak{M}$ give us in a certain sense the most "accurate" description of the statistical class.

We will subsequently assume that probability measure P is absolutely continuous regarding a volume measure V. It means that for any class $F \in \mathfrak{F}$ there is the probability density function $h(x)$, $x \in X$, connecting the volume and probability measures, i.e.

$$P(A) = \int\limits_A h(x)dV(x).$$

Also, we will assume that the volume measure V is continuous in regard to its values, that is if $A \in \mathfrak{A}$ and $V(A) = a$ then for an arbitrary $b \in [0,a]$ there exists $B \in \mathfrak{A}$ such as $B \subseteq A$ and $V(B) = b$. The continuous property of a probability measure regarding a volume measure gives us a possibility to introduce definitions of the inclusion and the equality of events in a volume measure. Namely, $A \subseteq B$ in measure V for any $A, B \in \mathfrak{A}$ if and only if $V(A \setminus B) = 0$. Then $A = B$ in measure V when both $A \subseteq B$ and $B \subseteq A$ in measure V.

3. Description of the minimal events set

We first establish a lemma that gives us a characteristic property of m.e.

Lemma_1. *Let $E \in \mathfrak{M}$, $V(E) = a$ and $\mathcal{B}(a)$ is the set of events of the volume a, i.e. $\mathcal{B}(a) = \{A \in \mathfrak{A} \mid V(A) = a\}$. Then $P(E) = \sup\limits_{A \in \mathcal{B}(a)} P(A)$.*

Thus, the result of the lemma shows that m.e. are the most probable among the events of the same volume. The following lemma establishes an important class of m.e.

Lemma_2. *For any $\alpha > 0$ the event $E = \{x \in X \mid h(x) \geq \alpha\}$ is minimal.*

The fundamental set of m.e. is the set of events $E(\alpha) = \{x \in X \mid h(x) \geq \alpha\}$. It is not hard to see that the events $E(\alpha)$ are linearly ordered regarding to the inclusion operation.

The concept of the fundamental set of m.e. causes a natural question: does the fundamental set of m.e. coincide with the set of all m.e. or not ? That is whether it is possible to represent an arbitrary m.e. in the form $\{x \in X \mid h(x) \geq \alpha\}$ with a certain α. The following results answer this question.

Consider an event $C \in \mathfrak{A}$. *The least upper bound* Lub(C) is defined as Lub(C) = $= \sup\{\alpha \mid C \subseteq E(\alpha)\}$. Note that in this definition $C \subseteq E(\alpha)$ means the inclusion C into $E(\alpha)$ in measure V, i.e. $V(C \setminus E(\alpha)) = 0$. We also introduce the following notations:

$\text{int}\,E(\alpha)=\{x\in X\mid h(x)>\alpha\}$; $\text{bd}\,E(\alpha)=\{x\in X\mid h(x)=\alpha\}$. Thus, the equality $E(\alpha)=\text{int}\,E(\alpha)\cup\text{bd}\,E(\alpha)$ holds.

Lemma_3. *Let for m.e. C $\text{Lub(C)}=\alpha$. Then $\text{int}\,E(\alpha)\subseteq C$ in measure V.*

The following theorem gives a description of the structure of an arbitrary m.e.

Theorem_1. *Let $C\in\mathfrak{M}$ and $\text{Lub(C)}=\alpha$. Then*

1) if $\alpha>0$ then $C=\text{int}\,E(\alpha)\cup A$, where $A\subseteq\text{bd}\,E(\alpha)$; conversely, for any $\alpha>0$ the event $C=\text{int}\,E(\alpha)\cup A$ is m.e.;

2) if $\alpha=0$ then $C=\text{int}\,E(0)$; conversely, $\text{int}\,E(0)$ is m.e.

Now we establish a feature of equiprobable m.e. that we state as a lemma.

Lemma_4. *Let $C_1,C_2\in\mathfrak{M}$ and $P(C_1)=P(C_2)$, then there is an equality $\text{Lub}(C_1)=\text{Lub}(C_2)$.*

Let $p\in[0,1]$, then as the probability measure is continuous regarding its values the set $\mathcal{A}(p)$ of the equiprobable events is not empty. Consequently, the set of the equiprobable m.e. is also not empty. The following theorem establishes conditions for the p-probable m.e. to be unique in measure V for any $p\in[0,1]$.

Theorem_2. *Any m.e. is determined by its probability uniquely if and only if for any α the equality $P(\text{bd}\,E(\alpha))=0$ holds.*

It is obvious, that when satisfying the conditions of the theorem the set of all m.e. $\mathfrak{M}$ coincides with the fundamental set. A statistical class $F\in\mathfrak{F}$ is called *regular* if each m.e. in it is defined by its probability uniquely. Note that the set of m.e. $\mathfrak{M}$ of the regular class F coincides with the fundamental set.

Theorem_3. *The fundamental set of m.e. determines each statistical class uniquely.*

4. Set-theoretic operations on statistical classes. Inclusion and equality measures

We shall first consider regular statistical classes. Let F_1 and F_2 be regular statistical classes from $\mathfrak{F}$; $A_1(p)$ and $A_2(p)$ - equiprobable with the probability p m.e. that correspond to the classes F_1 and F_2. Informally the probability p gives us representativity estimation of the m.e. $A_1(p)$ and $A_2(p)$ when describing the classes F_1 and F_2. These events possess an important extremal property since they are the least events with a given probability. That is why when defining set-theoretic operations and relations on statistical classes we take as a basis the corresponding operations on m.e.

We define $F_1\subseteq F_2$ if and only if for any $p\in[0,1]$ there is the inclusion $A_1(p)\subseteq A_2(p)$ in measure V; $F_1=F_2$ if $A_1(p)=A_2(p)$ in measure V for any $p\in[0,1]$. *The union* of classes F_1 and F_2 is a class $F_3=F_1\cup F_2$ such that for any $p\in[0,1]$ there is $A_3(p)=A_1(p)\cup A_2(p)$. *The intersection* of classes F_1 and F_2 is a class $F_3=F_1\cap F_2$ such that the equality $A_3(p)=A_1(p)\cap A_2(p)$ holds for any $p\in[0,1]$.

To classify the statistical classes we construct an appropriate inclusion measure ψ. Let $F_1,F_2\in\mathfrak{F}$ be regular statistical classes. Given $p\in[0,1]$ we define *p-local inclusion measure* of F_1 into F_2 as $\psi_p(F_1 \subseteq F_2) = P_1[A_2(p)|A_1(p)]$, that is the conditional probability of the event $A_2(p)$ occurrence provided that the event $A_1(p)$ has taken place in measure P_1. Note that the p-local measure possesses all required properties of a measure.

Meanwhile, it is more useful to introduce *an integral inclusion measure* (or simply an inclusion measure) $\psi(F_1 \subseteq F_2)$, that can be defined on the basis of the p-local measure:

$$\psi(F_1 \subseteq F_2) = \int_0^1 \psi_p(F_1 \subseteq F_2)(2p)dp = 2\int_0^1 P_1[A_2(p)|A_1(p)]pdp\,,$$

where $2p$ is a normalization factor. *The equality measure* we define such that

$$\psi(F_1 = F_2) = \min\{\psi(F_1 \subseteq F_2),\ \psi(F_2 \subseteq F_1)\}\,.$$

Theorem_4. *For arbitrary regular statistical classes $F_1, F_2 \in \mathfrak{F}$ we have:*
1) $\psi(F_1 \subseteq F_2) = 1$ if and only if $F_1 \subseteq F_2$; 2) $\psi(F_1 = F_2) = 1$ if and only if $F_1 = F_2$.

The formula of the inclusion measure can be transformed to the form that is more convenient for practical calculations. Let $\chi(z)$ be the function that equals 1, if $z \geq 0$, and equals 0, otherwise. We introduce a value

$$\mathcal{P}_i(x) = P_i[E_i(h_i(x))] = \int_X h_i(y)\chi[h_i(y) - h_i(x)]dV(y), \quad i = 1,2,$$

that has a sense of the probability of m.e. from $\mathfrak{M}$, whose density is not less than $h_i(x)$, $i = 1,2$. Then we can show that

$$\psi(F_1 \subseteq F_2) = 2\int_X h_1(x)\min\{\mu_1(x),\ \mu_2(x)\}dV(x),$$

where $\mu_1(x) = 1 - \mathcal{P}_1(x)$, $\mu_2(x) = 1 - \mathcal{P}_2(x)$. The function $\mu(x)$ has a profound sense when representing statistical class F by a fuzzy set. This will be discussed below.

5. Fuzzy representation of the regular statistical classes

Consider the function $\mu(x) = 1 - \mathcal{P}(x)$ of a regular statistical class $F \in \mathfrak{F}$.

Lemma_5. *For any $p > 0$ $\ A(1-p) = \{x \in X \mid \mu(x) \geq p\}$.*

It is obvious that the function $\mu(x)$ determines each statistical class uniquely. A class $F \in \mathfrak{F}$ can be considered as a fuzzy subset of the space X with a membership function $\mu(x)$. Let $F_1, F_2 \in \mathfrak{F}$. Then, if $F_3 = F_1 \cup F_2$ then $\mu_3(x) = \max\{\mu_1(x), \mu_2(x)\}$, and if $F_3 = F_1 \cap F_2$, then $\mu_3(x) = \min\{\mu_1(x), \mu_2(x)\}$. Thus, stated above set-theoretic operations on statistical classes coincide with the traditional operations in the theory of fuzzy sets [8].

It is known [9], that we can consider probabilities of fuzzy events in the measurable space X with the probability measure P. For a fuzzy event F with a membership function

$\mu(x)$: $X \to [0,1]$ the probability $P(F)$ is defined by the formula $P(F) = \int\limits_X \mu(x)dP(x)$.

Let $F_1, F_2 \in \mathfrak{F}$ be regular statistical classes. Then $P_1(F_1) = 0.5$ and

$$\psi(F_1 \subseteq F_2) = 2\int\limits_X \min\{\mu_1(x), \mu_2(x)\}dP_1(x) = 2P_1(F_1 \cap F_2) = \frac{P_1(F_1 \cap F_2)}{P_1(F_1)} = P_1(F_2|F_1).$$

Thus, the inclusion measure has a sense of the conditional probability of the fuzzy event F_2 occurrence provided that the fuzzy event F_1 has taken place in measure P_1.

6. Inclusion relation and inclusion measure for irregular statistical classes

In the section 4 we defined set-theoretic operations and relations for regular statistical classes. However, real probability distributions often provide statistical classes that do not possess the mentioned property. Such classes we will call *irregular classes.* As examples we can mention discrete, discrete-continuous and even purely continuous distributions of random values, whose probability density function has constancy domains of a non-zero measure. The irregularity of a class becomes apparent when its m.e. are determined not uniquely, or do not exist for certain probabilities[1]. We shall consider a general method of defining set-theoretic operations and relations on the whole set of statistical classes $\mathfrak{F}$ (including irregular classes).

Let us extend the set of m.e. $\mathfrak{M}$ with *fuzzy events,* i.e. with such events E with a membership function $\mu_E(x)$: $X \to [0,1]$, whose probability $P(E)$ and volume $V(E)$ are defined by the expressions $P(E) = \int\limits_X \mu_E(x)dP(x)$, $V(E) = \int\limits_X \mu_E(x)dV(x)$, and with it $V(E) = \inf\limits_{\substack{\text{over all } p\text{-probable} \\ \text{fuzzy events } A}} V(A)$. The following lemma describes membership functions of fuzzy m.e.

Lemma_6. *The event* $E(\alpha,q)$ *with the membership function*

$$\mu_E(x) = \begin{cases} 1, & x \in \operatorname{int} E(\alpha), \\ q, & x \in \operatorname{bd} E(\alpha), \\ 0, & x \notin E(\alpha), \end{cases} \qquad \alpha > 0, \qquad q \in [0,1],$$

is the fuzzy minimal event for a certain q.

It is not hard to see that if $\operatorname{bd} E(\alpha)$ equals zero in measure V, then the proved lemma gives us a description of ordinary (not fuzzy) m.e. We consider by analogy the m.e. $E(\alpha,q)$ described in lemma_6 to belong to the fundamental set of fuzzy m.e. These events can be taken as a basis to define set-theoretic operations on arbitrary statistical class.

Lemma_7. *The fuzzy m.e. with probability* $p \in [0,1]$ *from the fundamental set of fuzzy*

---

[1] Here and below it is not assumed that the volume measure is obligatory continuous in regard to its vaues (*authors' remark*).

m.e. is determined uniquely.

Because of proved uniqueness of constructing fuzzy m.e. from the fundamental set each such an event is completely determined by its probability p. Therefore, it can be denoted as $A(p)$. This allows us to introduce an inclusion relation and inclusion measure of statistical classes in just the same way as in the section 4, but however taking into account that m.e. might be *fuzzy*. We will use the classical **min**- and **max**-operations on fuzzy sets.

Let $\Phi_1 = \{A_1(p)\}$ and $\Phi_2 = \{A_2(p)\}$ be the fundamental set of fuzzy m.e. corresponding to statistical classes F_1 and F_2 respectively. We shall take by definition that $F_1 \subseteq F_2$ if and only if for any $p \in [0,1]$ the inclusion $A_1(p) \subseteq A_2(p)$ takes place in measure V. The inclusion measure of statistical classes is defined by the formula

$$\psi(F_1 \subseteq F_2) = 2\int_0^1 P_1[A_2(p) | A_1(p)]p dp = 2\int_0^1 P_1[A_1(p) \cap A_2(p)] dp.$$

7. Relation and measure of a possibilistic inclusion

Let a fuzzy set F be normal. Then in the possibilistic model the function $\mu(x)$ is interpreted as a function of a possibility distribution. With the help of this function for any $A \in \mathfrak{A}$ we can express the necessity measure $NESS(A) = \inf_{x \notin A}(1 - \mu(x))$ and the possibility measure $POSS(A) = \sup_{x \in A} \mu(x)$.

Taking into consideration that the necessity and the possibility measures have a sense of the lower and the upper probabilities [10,11] we draw the conclusion that the function $\mu(x)$ determines a family $\mathcal{P} = \{P_i\}$ of probability measures P_i such that

$$NESS(A) \le P_i(A) \le POSS(A) \qquad \text{for any } A \in \mathfrak{A}. \qquad (1)$$

In connection with this conclusion we introduce the possibilistic inclusion "$\overset{pos}{\subseteq}$" of a statistical class F_i into statistical class F: $F_i \overset{pos}{\subseteq} F$ if the condition (1) holds.

The condition (1) is not convenient for practical application. The following theorem enables us to express this condition by m.e. $\{A(p)\}$, $A(p) = \{x \in X \mid 1 - \mu(x) < p\}$.

Theorem_5. *The inclusion* $F_i \overset{pos}{\subseteq} F$ *is valid if and only if for any* $p \in [0,1]$ *the inequality* $P_i[A(p)] \ge p$ *holds.*

Corollary_1. *The relation of the set-theoretic inclusion "$\subseteq$" implicates the relation of possibilistic inclusion "$\overset{pos}{\subseteq}$", i.e.* $F_i \overset{pos}{\subseteq} F$ *follows* $F_i \subseteq F$, *but the contrary statement is false in general. In other words the inclusion "$\subseteq$" is stronger than "$\overset{pos}{\subseteq}$".*

Corollary_2. *The relations "$\subseteq$" and "$\overset{pos}{\subseteq}$" are equivalent if m.e. of statistical classes*

$F=(X,\mathfrak{A},P)$ *and* $F_i=(X,\mathfrak{A},P_i)$ *coincide, i.e.* $F_i \subseteq F \Leftrightarrow F_i \overset{pos}{\subseteq} F$, *if* $\Phi=\Phi_i$.

Corollary_2 enables us to project an approach to constructing the possibilistic inclusion measure $\psi(F_i \overset{pos}{\subseteq} F)$. Let the fundamental sets of m.e. of statistical classes F_i and F coincide. Then on account of corollary_2 it is naturally to require the meanings of set-theoretic and possibilistic measures to be equal, i.e. $\psi(F_i \overset{pos}{\subseteq} F)=\psi(F_i \subseteq F)$. As the following lemma shows in this case we can get a simpler expression for $\psi(F_i \subseteq F)$.

Lemma_8. *Let the fundamental sets* $\Phi_i=\{A_i(p)\}$ *and* $\Phi=\{A(p)\}$ *of m.e. of statistical classes* F_i *and* F *coincide, i.e. for any* $p\in[0,1]$ *there is such a* $p_i\in[0,1]$ *that* $A(p)=A_i(p_i)$. *Then*

$$\psi(F_i \subseteq F)=2\int_0^1 \min\left[p,P_i[A(p)]\right]dp. \tag{2}$$

The lemma having been proved enables us to define a measure of possibilistic inclusion on the whole set of statistical classes. Consider the statistical class F, fixed by the membership function $\mu(x)$, and the class $F_i=\{X,\mathfrak{A},P_i\}$, fixed by a probability measure P_i. One can construct the fundamental set of m.e. $\Phi=\{A(p)\}$, $A(p)=\{x\in X \mid 1-\mu(x)<p\}$, of the statistical class F by the function $\mu(x)$.

We choose conditionally the set of m.e. Φ_i for the statistical class F_i in such a way that $\Phi=\Phi_i$. In this case to define the inclusion measure $\psi(F_i \subseteq F)$ we could use the representation (2). We will consider that this formula is also valid for possibilistic inclusion measure (even if the sets of m.e. of statistical classes F and F_i are not coincide), i.e.

$$\psi(F_i \overset{pos}{\subseteq} F)=2\int_0^1 \min\left[p,P_i[A(p)]\right]dp.$$

Theorem_6. $\psi(F_i \overset{pos}{\subseteq} F)=1$ *if and only if* $F_i \overset{pos}{\subseteq} F$.

The upper and lower membership functions of the statistical class F_i are found from the following expressions

$$\overline{\mu}_i(x)=1-P_i\{y\in X \mid \mu(y)>\mu(x)\},$$
$$\underline{\mu}_i(x)=1-P_i\{y\in X \mid \mu(y)\ge\mu(x)\}.$$

8. Conclusion

The set-theoretic representation of probability distributions with the purpose of classifying them, proposed in the article, is the idea that has some merits. It enables us to establish in a natural way the relation of a partial order on the set of probability distributions and to define algebraic operations that can be interpreted as finding the least upper (the greatest lower) bounds in a partially ordered set. With this it should be mentioned that the method of probability distributions classification essentially differs from the analogous methods in the statistical theory of pattern recognition. The main restriction of these methods is that, as a rule, a rigorous classification is considered. That is observations of a certain random process or emergence are required to be strictly associated with one of some classes. Such an approach does not take into account really existing "intersection" of the classes that could considerably distort the picture of being observed emergency. In this sense the idea of statistical classes gives us a possibility to avoid the obstacles mentioned above.

In contrast to the classical Bayesian methods the classes in the proposed classification scheme might have fuzzy boundaries, the requirement of their disjointness and completeness does not appear to be obligatory. It must be especially emphasized that similar mathematical constructions are come out from quite another consideration in [5] when approximating the probability distribution by the possibility one. The analysis executed in the present paper showed that it could be found the formal ground of applying the theory of fuzzy sets and possibility theory for classification and identification of statistical data.

References

1. *Fine, T.L.* Theories of Probability: An Examination of Foundations. N.Y.: Academic Press, 1973.
2. *Sugeno, M.* Fuzzy measure and fuzzy integral. - Trans. SICE, 1972, v.8, N 2, p.95-102.
3. *Zadeh, L.A.* Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1978, N 1, p. 3-28.
4. *Nguyen, H.T.* On random sets and belief functions. J. Math. Anal. & Applic., 1978, N 65, p. 531-542.
5. *Dubois, D., Prade, H.* Unfair coins and necessity measures: Towards a possibilistic interpretation of histograms. Fuzzy Sets and Systems, 1983, v.10, N 1, p. 15-20.
6. *Dubois, D., Prade, H.* Fuzzy sets and statistical data. European J. Operational Research, 1986, N 25, p. 345-356.
7. *Dubois, D., Prade, H.* A set-theoretic view of belief functions. Logical operations and approximations by fuzzy sets. Inter. J. of General Systems, 1986, v.12, N 3, p.193-226.
8. *Kaufmann, A.* Introduction to the theory of fuzzy subsets, v.1. -N.Y.: Academic Press, 1975.
9. *Zadeh, L.A.* Probability measures of fuzzy events. J. Math. Anal. and Appl., 1968, N 23, p. 421-427.
10. *Dempster, A.P.* Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Statis., 1967, N 38, p. 325-339.
11. *Shafer, G.* A Mathematical Theory of Evidence. Princeton University Press, Princeton N.J., 1976.