# ON FREQUENCY DIAGRAMS WITH UNPRECISELY DEFINED CLASSES

Krzysztof PIASECKI
Department of Mathematics, Academy of Economy,
ul. Marchlewskiego 146/150, 60-967 Poznań,
Poland.

Let us look into the classical problem of investigation of distribution of characteristic in the general population $X=\{x\}$. The variation domain of characteristic $\Omega=\{\omega\}$ is interpretated as a set of elementary events and it can be divided into finite or infinite number classes $A_n$. The classes $A_n$ are mutually disjoint subsets in $\Omega$ satisfying additionally the condition

$$\Omega = \bigcup_n \{A_n\} . \tag{1}$$

On the other side, the general population is representated by the sample $\{x_k\} \subset X$ given as the finite series of primary samples. Then the observation function $f: X \rightarrow \Omega$ assigns the observation $\omega_k$ to each primary sample $x_k$ i.e. $\omega_k = f(x_k)$. For this case the empirical frequency $f_n$ of class $A_n$ is defined as follows

$$f_n = \operatorname{card}\{x_k : \omega_k \in A_n\} \tag{2}$$

for all $n \in \mathbb{N}$. For this case, as we know, we have

$$\sum_n f_n = m_x , \tag{3}$$

where symbol $m_x$ denotes a sample size.
Finally, the empirical frequences $f_n$ explicitly defines a probability

$$P(A) = \frac{\sum_n f_n}{m_x} \tag{4}$$

for each random event described by

$$A = \bigcup_{n'} \{A_{n'}\} \quad , \tag{5}$$

where $\{A_{n'}\}$ is any subsequence of $\{A_n\}$.

Very often the variation domain of characteristic is d vided into classes $A_n$ which are unprecisely defined. For example, the variation domain of price of fixed consumer goods can be divided into two parts: "low prices" and "high prices". Then each class $A_n$ is represented by such fuzzy subset $\mu_{A_n} \in \mathbb{F}(\Omega)$ that the sequence $\{\mu_{A_n}\}$ is a fuzzy complete partition of $\Omega$ (see [4]). So, the sequence $\{\mu_{A_n}\}$ has the following properties:

- the fuzzy subsets $\mu_{A_n}$ are pairwise W-separated [2], i.e. for each pair $(i,j)$ such that $i \neq j$, we have

$$\mu_{A_i} \leqslant 1 - \mu_{A_j} \tag{6}$$

- the fuzzy subset $\sup_n \{\mu_{A_n}\}$ is a W-universum [2] i.e.

$$\sup_n \{\mu_{A_n}\} \geqslant [\![ \tfrac{1}{2} ]\!]_{\Omega} \tag{7}$$

where $[\![ \frac{1}{2} ]\!]_{\Omega} : \Omega \to \{\frac{1}{2}\}$. We additionally assume about used complete partition $\{\mu_{A_n}\}$ that for each positive integer $n$ there exists such characteristic $\omega \in \Omega$ that

$$\mu_{A_n}(\omega) > \tfrac{1}{2} \tag{8}$$

From practical view-point the above assumption is not unreasonable demand because it says that each class $A_n$ cannot be represented by W-empty fuzzy subset (see [2]).

In agreement with (2), the empirical frequence $\tilde{f}_n$ of class $A_n$ is given, for this case, by the formula

$$\tilde{f}_n = \sum_{k=1}^{m_x} \mu_{A_n}(\omega_k) \tag{9}$$

Note that, in general, the empirical frequences $\tilde{f}_n$ do not satisfy the condition (3).

Let us consider now a family $\mathfrak{S}_A$ defined as follows

$$\mathfrak{S}_A = \{\mu : \mu \in \mathbb{F}(\Omega),\ \exists\, \mathbb{K} \subset \mathbb{N}_A : \sup_{k \in \mathbb{K}} \{\mu_{A_k}\} \leqslant \mu \leqslant$$
$$\leqslant 1 - \sup_{k \in \mathbb{N}_A \setminus \mathbb{K}} \{\mu_{A_k}\}$$
$$\text{and} \quad \mu \wedge (1 - \sup_{k \in \mathbb{K}} \{\mu_{A_k}\}) \leqslant [\![ \tfrac{1}{2} ]\!]_{\Omega}\}, \tag{10}$$

where the symbol $\mathbb{N}_A$ denotes the set of indexes of all classes $A_n$.

Since $\{\mu_{A_n}\}_{n \in \mathbb{N}_A} \subset \mathfrak{S}_A$, the family $\mathfrak{S}_A$ is not empty. Moreover, the family can be interpreted as family of some unprecisely defined parts of variation domain of characteristic. In general, the family $\mathfrak{S}_A$ does not contain precisely defined classes of partition. Moreover, then we have:

Lemma 1: If $\mu \in \mathfrak{S}_A$ then there exists the unique subset, $\mathbb{M}(\mu)$ say, of $\mathbb{N}_A$ such that

$$\sup_{k \in \mathbb{M}(\mu)} \{\mu_{A_k}\} \leqslant \mu \quad , \tag{11}$$

$$\mu \wedge (1 - \sup_{k \in \mathbb{M}(\mu)} \{\mu_{A_k}\}) \leqslant [\![ \tfrac{1}{2} ]\!] \quad . \tag{12}$$

Proof: Let $\mu$ be fixed fuzzy subset from $\mathfrak{S}_A$. Then there exists subset $\mathbb{M}_1$ in $\mathbb{N}_A$ such that the sequence $\{\mu_{A_k}\}_{k \in \mathbb{M}_1}$ satisfies the conditions (11) and (12) for $\mu$. Suppose that there exists such subset $\mathbb{M}_2$ in $\mathbb{N}_A$ that $M_2 \neq M_1$ and the sequence $\{\mu_{A_k}\}_{k \in \mathbb{M}_2}$ satisfies (11) and (12) for $\mu$, too. Then we can find a positive integer $l$ which belongs to $\mathbb{M}_2 \setminus \mathbb{M}_1$ or $\mathbb{M}_1 \setminus \mathbb{M}_2$. Assume now that $l \in \mathbb{M}_1 \setminus \mathbb{M}_2$. The condition (8) implies that there exists such $\omega^* \in \Omega$ that $\mu_{A_l}(\omega^*) > \frac{1}{2}$. Thus, using (11), we get

$$\mu(\omega^*) > \tfrac{1}{2} \quad . \tag{*}$$

On the other side, by (11) we obtain

$$\frac{1}{2} < \mu_{A_1}(\omega^*) \leqslant 1 - \mu_{A_k}(\omega^*)$$

for each $k \in \mathbb{M}_2$. This along with (*) implies

$$\mu(\omega^*) \wedge (1 - \sup_{k \in \mathbb{M}_2}\{\mu_{A_k}(\omega^*)\}) =$$
$$= \mu(\omega^*) \wedge \inf_{k \in \mathbb{M}_2}\{1 - \mu_{A_k}(\omega^*)\} \geqslant \mu(\omega^*) \wedge \mu_{A_1}(\omega^*) > \frac{1}{2} .$$

Contradiction! So, $\mathbb{M}_1 \subset \mathbb{M}_2$. By analogous way, we get $\mathbb{M}_2 \subset \mathbb{M}_1$. Therefore, $\mathbb{M}_1 = \mathbb{M}_2$. ■

Lemma 2: We have $\mathbb{M}(\mu \vee \nu) = \mathbb{M}(\mu) \cup \mathbb{M}(\nu)$ for any pair $(\mu, \nu) \in \sigma_A^2$.

Proof: For any fixed pair $(\mu, \nu) \in \sigma_A^2$ we get:

$$\mu \vee \nu \geqslant \sup_{k \in \mathbb{M}(\mu)}\{\mu_{A_k}\} \vee \sup_{k \in \mathbb{M}(\nu)}\{\mu_{A_k}\} =$$
$$= \sup_{k \in \mathbb{M}(\mu) \cup \mathbb{M}(\nu)}\{\mu_{A_k}\} ,$$
$$\mu \vee \nu \leqslant (1 - \sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\mu)}\{\mu_{A_k}\}) \vee (1 - \sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\nu)}\{\mu_{A_k}\}) =$$
$$= 1 - \sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\mu)}\{\mu_{A_k}\} \wedge \sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\nu)}\{\mu_{A_k}\} \leqslant$$
$$\leqslant 1 - \sup_{k \in (\mathbb{N}_A \setminus \mathbb{M}(\mu)) \cap (\mathbb{N}_A \setminus \mathbb{M}(\nu))}\{\mu_{A_k}\} =$$
$$= 1 - \sup_{k \in (\mathbb{N}_A \setminus (\mathbb{M}(\mu) \cup \mathbb{M}(\nu)))}\{\mu_{A_k}\} .$$

and

$$(\mu \vee \nu) \wedge (1 - \sup_{k \in \mathbb{M}(\mu) \cup \mathbb{M}(\nu)}\{\mu_{A_k}\}) =$$
$$= (\mu \vee \nu) \wedge (1 - \sup_{k \in \mathbb{M}(\mu)}\{\mu_{A_k}\} \vee \sup_{k \in \mathbb{M}(\nu)}\{\mu_{A_k}\}) =$$
$$= (\mu \vee \nu) \wedge (1 - \sup_{k \in \mathbb{M}(\mu)}\{\mu_{A_k}\}) \wedge (1 - \sup_{k \in \mathbb{M}(\nu)}\{\mu_{A_k}\}) \leqslant$$
$$\leqslant (\mu \wedge (1 - \sup_{k \in \mathbb{M}(\mu)}\{\mu_{A_k}\})) \vee (\nu \wedge (1 - \sup_{k \in \mathbb{M}(\nu)}\{\mu_{A_k}\})) \leqslant$$
$$\leqslant [\![ \tfrac{1}{2} ]\!]_\Omega .$$

The thesis is proved. ■

Lemma 3: We have $\mathbb{M}(1-\mu) = \mathbb{N}_A \setminus \mathbb{M}(\mu)$ for each $\mu \in \mathfrak{S}_A$.

Proof: If $1-\mu(\omega) > \frac{1}{2}$, for fixed $\omega \in \Omega$, then we have

$$\sup_{k \in \mathbb{M}(\mu)} \{\mu_{A_k}(\omega)\} \leqslant \mu(\omega) < \frac{1}{2}.$$

Thus, by (7), we get

$$\sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\mu)} \{\mu_{A_k}(\omega)\} \geqslant \frac{1}{2}.$$

This fact proves that

$$(1-\mu) \wedge (1 - \sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\mu)} \{\mu_{A_k}\}) \leqslant [\![ \tfrac{1}{2} ]\!]_\Omega .$$

The condition

$$\sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\mu)} \{\mu_{A_k}\} \leqslant 1-\mu \leqslant 1 - \sup_{k \in \mathbb{M}(\mu)} \{\mu_{A_k}\}$$

follows from

$$\sup_{k \in \mathbb{M}(\mu)} \{\mu_{A_k}\} \leqslant \mu \leqslant 1 - \sup_{k \in \mathbb{N}_A \setminus \mathbb{M}(\mu)} \{\mu_{A_k}\} . \blacksquare$$

Lemma 4: We have $\mathbb{M}(\mu \vee (1-\mu)) = \mathbb{N}_A$ for each $\mu \in \mathfrak{S}_A$.

Proof: Immediatelly from the Lemmas 2 and 3. ■

Lemma 5: We have $\mathbb{M}(\mu) \cap \mathbb{M}(\nu) = 0$ for any pair of W-separated fuzzy subsets $\mu$ and $\nu$ from $\mathfrak{S}_A$.

Proof: Let $\mu$ and $\nu$ be W-separated fuzzy subsets from $\mathfrak{S}_A$. Assume that $1 \in \mathbb{M}(\mu) \cap \mathbb{M}(\nu)$. Since (8), there exists such $\omega^* \in \Omega$ that $\mu_{A_1}(\omega^*) > \frac{1}{2}$. Therefore, by (11) we obtain $\mu(\omega^*) > \frac{1}{2}$ and $\nu(\omega^*) > \frac{1}{2}$. Then the W-separatity between $\mu$ and $\nu$ implies

$$\frac{1}{2} < \mu(\omega^*) \leqslant 1 - \nu(\omega^*) < \frac{1}{2}$$

Contradiction! The proof is ended. ■

Finally, we can show:

Theorem 1: The family $\sigma_A$ is a soft fuzzy algebra (see [3]).
Proof: Since $\emptyset \subset \mathbb{N}_A$, $0_\Omega \in \sigma_A$. The Lemmas 2 and 3 show that $\sigma_A$ is closed under union and complementation.

Assume now that $[\![\frac{1}{2}]\!]_\Omega \in \sigma_A$. Then there exists such $\mathbb{M} \subset \mathbb{N}_A$ that

$$\sup_{k \in \mathbb{M}} \{\mu_{A_k}\} \leqslant [\![\tfrac{1}{2}]\!]_\Omega \qquad (*)$$

and

$$[\![\tfrac{1}{2}]\!]_\Omega \leqslant 1 - \sup_{k \in \mathbb{N}_A \setminus \mathbb{M}} \{\mu_{A_k}\} \qquad (**)$$

Using (8) and (*) we obtain $\mathbb{M}=\emptyset$. Furthermore, the conditions (8) and (**) imply $\mathbb{M}=\mathbb{N}_A$. Contradiction! So, $[\![\frac{1}{2}]\!]_\Omega \notin \sigma_A$. ∎

Consider now any fuzzy P-measure $p: \sigma_A \to [0,1]$ on $\sigma_A$ (see [3]). In accordance with the definition of fuzzy P-measure we have:

- for any $\mu \in \sigma_A$, $p(\mu \vee (1-\mu))=1$ ; (13)
- if $\{\mu_n\}$ is finite or infinite sequence of pairwise W-separated fuzzy subsets from $\sigma_A$ then

$$p(\sup_n \{\mu_n\}) = \textstyle\sum_n p(\mu_n) \quad . \qquad (14)$$

This measure will be interpreted as empirical distribution of characteristic. Since the characteristics are unprecisely grouped, this distribution is a distribution of probability of fuzzy events. Ph.Smets [7] says that probability of fuzzy event $\mu$ can be related to the cordinality of fuzzy subset $\mu$. Therefore, we additionally assume that there exists such positive real number $\alpha$ that

$$p(\mu_{A_n}) = \alpha \cdot \tilde{f}_n \qquad (15)$$

for each class $A_n$. Replacement cardinality by empirical frequence follows from (2) and (9).

Theorem 2: The mapping $p\colon \sigma_A \to [0,1]$, defined by the identity

$$p(\mu) = 0 + \frac{\sum_{k\in \mathbb{M}(\mu)} \tilde{f}_k}{\sum_{k\in \mathbb{N}_A} \tilde{f}_k} \tag{16}$$

for each $\mu \in \sigma_A$, is the unique fuzzy P-measure on $\sigma_A$ which satisfies (15).

Proof: The Lemma 1 proves that the mapping $p$, given by (16), is explicitly defined. Using the Lemmas 4 and 5 we can to show that the mapping $p$ fulfils the properties (13) and (14). So, the identity (16) describes a well-defined fuzzy P-measure on $\sigma_A$. Let $\tilde{p}\colon \sigma_A \to [0,1]$ be any fuzzy P-measure on $\sigma_A$ satisfying the condition (15). Taking into account (6), (7), (13) and (14) we obtain

$$1 = \tilde{p}(\sup_{k\in \mathbb{N}_A}\{\mu_{A_k}\} \vee (1 - \sup_{k\in \mathbb{N}_A}\{\mu_{A_k}\})) = \tilde{p}(\sup_{k\in \mathbb{N}_A}\{\mu_{A_k}\}) =$$

$$= \sum_{k\in \mathbb{N}_A} p(\mu_{A_k}) = \alpha \sum_{k\in \mathbb{N}_A} \tilde{f}_k \ .$$

Thus

$$p(\mu_{A_n}) = \frac{\tilde{f}_n}{\sum_{k\in \mathbb{N}_A} \tilde{f}_k} = p(\mu_{A_n}) \ .$$

Let $\mu$ be any fuzzy subset in $\sigma_A$. Note that $\{\mu_{A_k}\}_{k\in \mathbb{M}(\mu)}$ is a fuzzy partition of $\mu$ ( see [4]). Therefore, we have

$$\tilde{p}(\mu) = \sum_{k\in \mathbb{M}(\mu)} \tilde{p}(\mu_{A_k}) = \sum_{k\in \mathbb{M}(\mu)} p(\mu_{A_k}) = p(\mu) \ .$$

The uniqueness of $p$ is proved. ∎

We have seen that the empirical frequences explicitly define a distribution of characteristic which is a fuzzy P-measure. E.P.Klement et.al. [1] have defined a fuzzy probability measure. Their definition is more general than the definition of fuzzy P-measure

It is very easy to check that the mapping $m: \mathfrak{S}_A \to [0,1]$, given by

$$m(\mu) = 0 + \frac{\sum_{l=1}^{m_x} \sup_{k \in \mathbb{M}(\mu)} \{\mu_{A_k}(\omega_l)\}}{\sum_{l=1}^{m_x} \sup_{k \in \mathbb{N}_A} \{\mu_{A_k}(\omega_l)\}} \tag{17}$$

for each $\mu \in \mathfrak{S}_A$, is a fuzzy probability measure on $\mathfrak{S}_A$ which satisfies the condition (15). Moreover, in general we have $p(\mu) \neq m(\mu)$. So, if we assume that distribution of characteristic is a fuzzy probability measure then it cannot be explicitly defined by the empirical frequences. On the other side, the fuzzy P-measures are the unique fuzzy probability measures satisfying the Bayes Formula (see [5]). These facts are justifying for qualification a distribution of characteristic as fuzzy P-measure.

Finally, some extension of distribution of characteristic will be presented. We shall use the following mappings defined by

$$K(\mu) = \{\omega : \omega \in \Omega, \ \mu(\omega) > \tfrac{1}{2}\},$$

$$K^*(\mu) = \{\omega : \omega \in \Omega, \ \mu(\omega) = \tfrac{1}{2}\},$$

$$L(\mu) = K(\mu) \cup K^*(\mu)$$

for each $\mu \in \mathbb{F}(\Omega)$. Using the above mappings, we define the next families:

$$\mathbb{K}(\mathfrak{S}_A) = \{B : B \in 2^{\Omega}, \ \exists \mu \in \mathfrak{S}_A : K(\mu) \subset B \subset L(\mu)\},$$

$$\mathbb{K}^*(\mathfrak{S}_A) = \{B : B \in 2^{\Omega}, \ \exists \mu \in \mathfrak{S}_A : B = K^*(\mu)\},$$

$$\mathbb{E}(\mathfrak{S}_A) = \{\mu : \mu \in \mathbb{F}(\Omega), \ \exists (B,C) \in \mathbb{K}^2(\mathfrak{S}_A) : B \subset C \ \& \ B = K(\mu) \ \& \ C = L(\mu)\}.$$

Furthermore, let as define a such family $\mathbb{E}^*(\mathfrak{S}_A)$ as the subfamily of $\mathbb{E}(\mathfrak{S}_A)$ that for each $\{\mu_n\} \subset \mathbb{E}^*(\mathfrak{S}_A)$ there exists $A \in \mathbb{K}^*(\mathfrak{S}_A)$ containing $\sup_n \{\mu_n\}$. The mapping $P^*: \mathbb{K}(\mathfrak{S}_A) \to [0,1]$, given by the implication:

$$\text{" if } K(\mu) \subset B \subset L(\mu) \text{ then } P^*(B) = p(\mu) \text{ "} \tag{19}$$

for each $B \in \mathbb{K}(\sigma_A)$ , is explicitly given usual probability measure [6] . Using all above notions we can to present the following thesies:

Theorem 3: The mapping $\bar{p}: \mathbb{E}^*(\sigma_A) \to [0,1]$ , defined by (19) and

$$\bar{p}(\mu) = P^*(K(\mu)) \tag{20}$$

for each $\mu \in \mathbb{E}^*(\sigma_A)$ , is the unique extension of $p$ on $\sigma_A$ to $\mathbb{E}^*(\sigma_A)$ which is a fuzzy P-measure [6] .

Theorem 4: The mapping $\hat{p}: \mathbb{E}^*(\sigma_A) \to [0,1]$ , defined by (19) and

$$\hat{p}(\mu) = P^*(L(\mu)) \tag{21}$$

for each $\mu \in \mathbb{E}^*(\sigma_A)$ , is the unique extension of $p$ on $\sigma_A$ to $\mathbb{E}^*(\sigma_A)$ which is a fuzzy P-measure [6] .

Theorem 5: The mapping $\bar{p}: \mathbb{E}(\sigma_A) \to [0,1]$ , defined by (19) and (20) for each $\mu \in \mathbb{E}(\sigma_A)$ is a extension of $p$ on $\sigma_A$ to $\mathbb{E}(\sigma_A)$ which is a fuzzy probability measure [6] .

Theorem 6: The mapping $\hat{p}: \mathbb{E}(\sigma_A) \to [0,1]$ , defined (19) and (21) for each $\mu \in \mathbb{E}(\sigma_A)$ , is a extension of $p$ on $\sigma_A$ to $\mathbb{E}(\sigma_A)$ which is a fuzzy probability measure [6] .

Note that mappings $\bar{p}$ and $\hat{p}$ describe a different fuzzy probability measures on $\mathbb{E}(\sigma_A)$ . Moreover, then we have:

Theorem 7: If the fuzzy probability measure $m: \mathbb{E}(\sigma_A) \to [0,1]$ satisfies $m(\mu) = \bar{p}(\mu) = \hat{p}(\mu)$ for each $\mu \in \mathbb{E}^*(\sigma_A)$ then it fulfils $\overline{p}(\mu) \leq m(\mu) \leq \overline{\overline{p}}(\mu)$ for any $\mu \in \mathbb{E}(\sigma_A)$ [6].

We observe that $\mathcal{A} \subset \mathbb{E}(\sigma_A)$, where $\mathcal{A}$ is the family of all subsets in $\Omega$ defined by (5), but in general the family $\mathbb{E}^*(\sigma_A)$ does not contain the family $\mathcal{A}$. As we know, the family $\mathcal{A}$ is a family of precisely defined classes dividing the variation domain of characteristic. In agreement with the Theorems 5,6,7, the distribution of characteristic on $\mathcal{A}$ in unprecisely defined. It is immediate consequence of primary unprecise qualification of classes. Also the Bayes method of inference cannot be confined to $\mathcal{A}$. I suppose that investigation of distribution of characteristic can be limited to $\mathbb{E}^*(\sigma_A)$.

References:


[1] E.P.Klement, R.Lowen, W.Schwihla, Fuzzy Probability Measures, Fuzzy Sets and Systems 5 (1981), 21-30.

[2] K.Piasecki, New Concept of Separated Fuzzy Subsets, Proc. the Polish Symposium on Interval and Fuzzy Mathematics (1985) 193-195.

[3] K.Piasecki, Probability of Fuzzy Events Defined as Denumerable Additivity Measure, Fuzzy Sets and Systems 17 (1985), 271-284.

[4] K.Piasecki, Fuzzy Partition of Sets, BUSEFAL 25 (1986), 52-60.

[5] K.Piasecki, On the Bayes Formula for Fuzzy probability Measures, Short Communication, Fuzzy Sets and Systems 18 (1986), 183-185.

[6] K.Piasecki, Extension of Fuzzy P-measure Generated by Usual Measure, preparing paper.

[7] Ph.Smets, Probability of Fuzzy Events: an Axiomatic Approach, Fuzzy Sets and Systems 7 (1982), 153-164