Ann. Inst. Statist. Math.
Vol. 40, No. 4, 627-639 (1988)

ON THE LOSS OF INFORMATION DUE TO FUZZINESS IN EXPERIMENTAL OBSERVATIONS

MARÍA ANGELES GIL

Departamento de Matemáticas, Universidad de Oviedo, 33071 Oviedo, Spain

(Received April 11, 1987; revised January 28, 1988)

Abstract. The absence of exactness in the observation of the outcomes of a random experiment always entails a loss of information about the experimental distribution. This intuitive assertion will be formally proved in this paper by using a mathematical model involving the notions of fuzzy information and fuzzy information system (as intended by Tanaka, Okuda and Asai) and Zadeh's probabilistic definition. On the basis of this model we are first going to consider a family of measures of information enclosing some well-known measures, such as those defined by Kagan, Kullback-Leibler and Matusita, and then to establish methods for removing the loss of information due to fuzziness by increasing suitably the number of experimental observations.

Key words and phrases: Fuzzy information, fuzzy information system, non-parametric measures of directed divergence, probability of a fuzzy event, random experiment.

1. Introduction

An *experiment* is the process by which an observation is made. When a *random experiment* is performed it results in one outcome that cannot be previously predicted. In this way, in a random experiment one must distinguish two fundamental elements: the sample space (set consisting of all possible experimental outcomes) and the ability to observe the outcomes.

Given a random experiment and a sample from it, the aim of the Statistical Information Theory is to quantify the information contained in the sample and to use this information in making inferences about the experiment. When one tries to state such a quantification it is usually assumed that the ability to observe allows the statistician to identify each *observable event* with a subset of the sample space.

In this paper, we will suppose that the person responsible for observation cannot always crisply perceive the results of the experimental performance, but each observable event may only be identified with a fuzzy subset of the sample space (Zadeh (1965)).

More precisely, we will assimilate each observable event with fuzzy information, which is defined (Okuda *et al.* (1978), Tanaka *et al.* (1979) and Zadeh (1978)) as follows:

Let $\boldsymbol{X} = (X, \beta_X, P)$, $P \in \boldsymbol{P}$, be an experiment, where (X, β_X) is a measurable space and P belongs to a specified family of probability measures $\boldsymbol{P}$ on (X, β_X). We hereafter assume that the sample space X is a set in a Euclidean space (usually $\boldsymbol{R}$) and β_X is the smallest Borel σ-field on X.

DEFINITION 1.1. A fuzzy event $\tilde{x}$ on X, characterized by a Borel-measurable membership function $\mu_{\tilde{x}}$ from X to $[0, 1]$, where $\mu_{\tilde{x}}(x)$ represents the "grade of membership" of x to $\tilde{x}$ is called *fuzzy information associated with the experiment* $\boldsymbol{X}$.

The scheme in Fig. 1 explains the mechanism to obtain fuzzy information (Tanaka *et al.* (1979)).

Obviously, the assimilation of each observable event with fuzzy information permits us to describe the non-crisp observations with greater closeness to real-life than with the rounding or grouping of experimental data. In addition, it provides us with "less error" than the rounding and with "more precision" than the grouping of data.

In practice, the grade of membership $\mu_{\tilde{x}}(x)$ is often regarded as a kind of "probability with which the observer gets the fuzzy information $\tilde{x}$ when he really has obtained the exact outcome x". This interpretation justifies the assumption of orthogonality for the set of all observable events from $\boldsymbol{X}$, that is, to consider that this set is a fuzzy information system associated with $\boldsymbol{X}$, where (see Tanaka *et al.* (1979))

DEFINITION 1.2. A *fuzzy information system* (f.i.s.) $\tilde{\boldsymbol{X}}$ *associated with the experiment* $\boldsymbol{X}$ is a fuzzy partition with fuzzy events on X, that is, a finite set of fuzzy events on X satisfying the *orthogonality condition*

$$\sum_{\tilde{x} \in \tilde{\boldsymbol{X}}} \mu_{\tilde{x}}(x) = 1 \qquad \text{for all} \qquad x \in X .$$

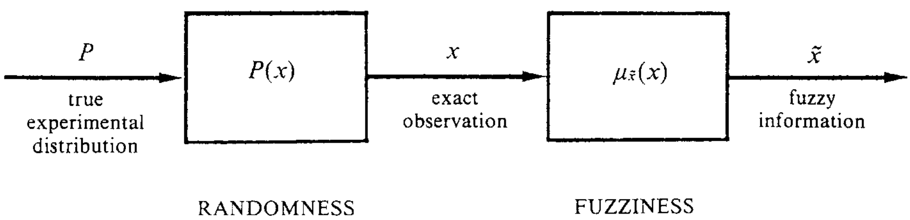

Fig. 1. Process leading to fuzzy information associated with a random experiment.

Assume that a simple random sampling of size n from the experiment $\boldsymbol{X} = (X, \beta_X, P)$, $P \in \boldsymbol{P}$, is going to be considered in order to make posterior inferences about that experiment. If the ability to observe does not permit one to perceive exactly the experimental outcomes, the following notions (Gil *et al.* (1984, 1985*a*, 1985*b*), Casals *et al.* (1986) and Gil (1988)) supply an operative model to express the available sample observations with fuzzy imprecision:

Let $\boldsymbol{X}^{(n)} = (X^n, \beta_{X^n}, P)$, $P \in \boldsymbol{P}$, be a simple random sample of size n from $\boldsymbol{X}$, and let $\tilde{\boldsymbol{X}}$ be an f.i.s. associated with $\boldsymbol{X}$.

DEFINITION 1.3. An n-tuple of elements in $\tilde{\boldsymbol{X}}$, $(\tilde{x}_1, \ldots, \tilde{x}_n)$, representing the algebraic product of $\tilde{x}_1, \ldots, \tilde{x}_n$ is called *sample fuzzy information of size n from* $\tilde{\boldsymbol{X}}$ ($(\tilde{x}, \tilde{x}')$ = algebraic product of $\tilde{x}$ and $\tilde{x}'$, with $\mu_{(\tilde{x}, \tilde{x}')}(x, x') = \mu_{\tilde{x}}(x)\mu_{\tilde{x}'}(x')$).

DEFINITION 1.4. A *fuzzy random sample of size n from* $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{X}}^{(n)}$, (*associated with the random sample* $\boldsymbol{X}^{(n)}$) is the set consisting of all algebraic products of n elements in $\tilde{\boldsymbol{X}}$.

Remark 1.1. It should be emphasized that we could use a more general definition for the concepts in Definitions 1.3 and 1.4, so that the membership function of each n-tuple $(\tilde{x}_1, \ldots, \tilde{x}_n)$ would be given by the expression $\mu_{(\tilde{x}_1, \ldots, \tilde{x}_n)}(x_1, \ldots, x_n) = f(\mu_{\tilde{x}_1}(x_1), \ldots, \mu_{\tilde{x}_n}(x_n), \tilde{x}_1, \ldots, \tilde{x}_n)$, f being a function taking on the values in the unit interval $[0, 1]$ and satisfying some natural conditions. But, in practice, when we consider examples involving probabilities one of the most operative and suitable functions f is the product of the first n components. This suitability is confirmed by the fact that the probabilistic independence of the experimental performances implies that (in Zadeh's sense (1968)) of the fuzzy observations from them, whenever f is the product as above.

The first purpose in this paper is to quantify the amount of information contained in a fuzzy random sample. This purpose is going to be achieved by extending the family of the non-parametric measures of (non-additive) divergence of order α, and their limit as $\alpha \to 1$.

In terms of such a family we will formalize the intuitive statement "the presence of fuzziness entails a loss of information". The second objective of this paper is to take advantage of some properties of the considered measures in order to remove the loss of information caused by the fuzziness by adequately increasing the sample size.

2. A non-parametric family of divergence measures for fuzzy information systems; loss of information and sample size

The mathematical model for a random experiment containing fuzzy observations may be completed by the introduction of the probability of a fuzzy event (Zadeh (1968)).

Let $\boldsymbol{X} = (X, \beta_X, P)$, $P \in \boldsymbol{P}$, be a random experiment and let $\tilde{\boldsymbol{X}}$ be an f.i.s. associated with it. Each probability measure P on (X, β_X) induces a probability distribution $\tilde{P}$ on $\tilde{\boldsymbol{X}}$ defined by:

DEFINITION 2.1. The *probability distribution on $\tilde{\boldsymbol{X}}$ induced by P* is the mapping $\tilde{P}$ from $\tilde{\boldsymbol{X}}$ to $[0, 1]$ given by

$$\tilde{P}(\tilde{x}) = \int_X \mu_{\tilde{x}}(x)\, dP(x), \qquad \text{for all} \qquad \tilde{x} \in \tilde{\boldsymbol{X}}$$

(the integral being the Lebesgue-Stieltjes integral).

It is worth pointing out that Definition 2.1 becomes a generalization of the Total Probability Rule when the grade of membership $\mu_{\tilde{x}}(x)$ is interpreted as the probability with which the observer gets the fuzzy information $\tilde{x}$ when he really has obtained the exact outcome x.

With the concepts we have defined we can establish an operative model for a random experiment with previous probabilistic uncertainty (randomness in the experimental outcomes) and actual fuzzy imprecision (fuzziness in the observation). Thus, although the probabilistic framework is not enough by itself to provide us with a suitable model characterizing such a random experiment, the Theory of Fuzzy Sets complements the Probability Theory and supplies concepts permitting us finally to construct that model in the probabilistic setting. More precisely, the approach based on the assimilation of each imprecise observable event with fuzzy information, and involving the notion of fuzzy information system and Zadeh's probabilistic definition, will allow us to pass from the original probability space (X, β_X, P) to a new probability space $(\tilde{\boldsymbol{X}}, \alpha_{\tilde{X}}, \tilde{P})$, where $\alpha_{\tilde{X}}$ is a σ-field on the (nonfuzzy) set $\tilde{\boldsymbol{X}}$ (e.g., parts of $\tilde{\boldsymbol{X}}$) (see Fig. 2).

The main advantage of this approach is that many statistical problems with imprecise data can be mathematically handled as statistical problems with a finite number of exact data (although the first problem is essentially an extension of the second one). On the basis of this argument several measures, principles and procedures have been extended in previous papers (Gil *et al.* (1984, 1985*a*, 1985*b*), Casals *et al.* (1986) and Gil (1988)) from the nonfuzzy case to the fuzzy one. In the same way, we are now going to extend a family of information measures.

Let P_1 and P_2 be two probability measures in the specified family $\boldsymbol{P}$ in the experiment $\boldsymbol{X}$.

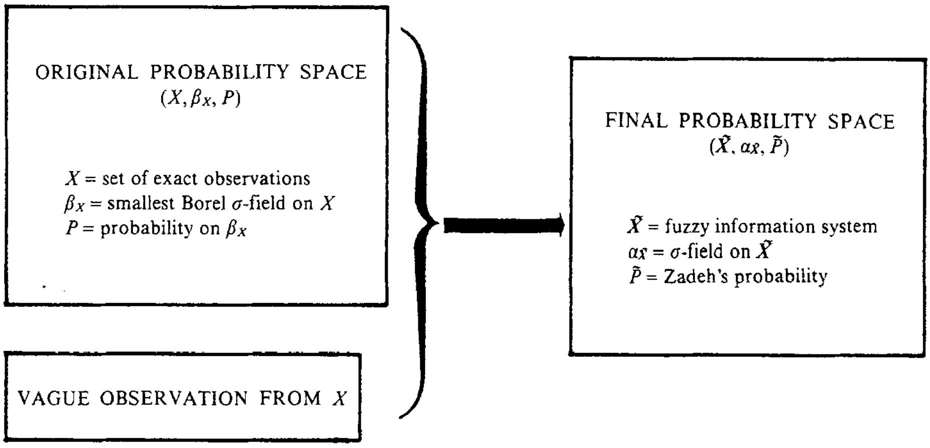

Fig. 2. Modelization within the probabilistic framework of the random experiments containing vague observations.

DEFINITION 2.2. The *(non-additive) directed divergence of order* α $(\alpha \neq 1)$ *in favor of* P_1 *against* P_2 *per observation from* P_2 *in* $\tilde{\mathcal{X}}$ is the value

$$I^*_\alpha(P_1 | P_2; \tilde{\mathcal{X}}) = (2^{\alpha-1} - 1)^{-1} \left\{ \sum_{\substack{\tilde{x} \in \tilde{\mathcal{X}} \\ \tilde{P}_2(\tilde{x})>0}} [\tilde{P}_1(\tilde{x})]^{\alpha} [\tilde{P}_2(\tilde{x})]^{1-\alpha} - 1 \right\},$$

where $\tilde{P}_1$ and $\tilde{P}_2$ are the induced probability distributions from P_1 and P_2, respectively.

Obviously, for $\alpha = 2$ the measure I^*_α reduces to the extension of Kagan's divergence (1963), for $\alpha = 1/2$ it reduces (unless a positive constant) to the extension of Matusita's measure of squared distance (1967) and the Kullback-Leibler directed information (1951) is the limit of I^*_α as $\alpha \to 1$.

Interpretations, properties, axiomatic characterizations and applications of the nonextended measures can be found in the literature (see Rathie (1973), Mathai and Rathie (1975), and references in them). Similar interpretations and properties could be immediately derived for the measures in Definition 2.2.

The absence of exactness in the observations from the performance of a random experiment (which determines herein the presence of fuzziness) entails a loss of information. This intuitive assertion is now formalized through the following theorem suggesting that the divergence in an experiment is at least equal to the corresponding divergence in an f.i.s. associated with it, whatever the pair of probability measures on the experiment may be.

THEOREM 2.1. *Let $\boldsymbol{X} = (X, \beta_X, P)$, $P \in \boldsymbol{P}$, be a random experiment and assume that the divergence of order α between P_1 and P_2 in $\boldsymbol{X}$ exists and is finite. Let $\tilde{\boldsymbol{X}}$ be an f.i.s. associated with $\boldsymbol{X}$. Then,*

$$I_\alpha^*(P_1 | P_2; \boldsymbol{X}) \geq I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}) ,$$

with equality if and only if $f_1(x)/f_2(x) = \tilde{P}_1(\tilde{x})/\tilde{P}_2(\tilde{x})$ for almost all $x \in \operatorname{supp} \tilde{x}$ and for all $\tilde{x} \in \tilde{\boldsymbol{X}}$ such that $\tilde{P}_1(\tilde{x}) > 0$, $\tilde{P}_2(\tilde{x}) > 0$ (where $f_i(\cdot)$ represents the density associated with P_i with respect to a σ-finite measure on (X, β_X), and $\operatorname{supp} \tilde{x} = \{x \in X, \mu_{\tilde{x}}(x) > 0\}$).

PROOF. Indeed, if we denote $g_i(x; \tilde{x}) = \mu_{\tilde{x}}(x) f_i(x)/\tilde{P}_i(\tilde{x})$ for all $x \in \operatorname{supp} \tilde{x}$ and $\tilde{P}_i(\tilde{x}) > 0$, and γ denotes the σ-finite measure on (X, β_X) determining densities f_i for P_i, we have

$$\int_{\operatorname{supp} \tilde{x}} g_i(x; \tilde{x}) d\gamma(x) = 1, \qquad i = 1, 2 .$$

Then, on the basis of the Jensen inequality, we obtain for $\alpha > 1$,

$$\begin{aligned} 1 &= \left\{ \int_{\operatorname{supp} \tilde{x}} g_i(x; \tilde{x}) [g_j(x; \tilde{x})/g_i(x; \tilde{x})] d\gamma(x) \right\}^{1-\alpha} \\ &\leq \int_{\operatorname{supp} \tilde{x}} g_i(x; \tilde{x}) [g_j(x; \tilde{x})/g_i(x; \tilde{x})]^{1-\alpha} d\gamma(x) \\ &= [\tilde{P}_i(\tilde{x})/\tilde{P}_j(\tilde{x})]^{1-\alpha} [1/\tilde{P}_i(\tilde{x})] . \end{aligned}$$

$\int_{\{f_j(x)>0\}} \mu_{\tilde{x}}(x) f_i(x) [f_j(x)/f_i(x)]^{1-\alpha} d\gamma(x)$, $i, j \in \{1, 2\}$, $i \neq j$, that is,

$$\begin{aligned} I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}) &= (2^{\alpha-1} - 1)^{-1} \left\{ \sum_{\substack{\tilde{x} \in \tilde{\boldsymbol{X}} \\ \tilde{P}_2(\tilde{x}) > 0}} [\tilde{P}_1(\tilde{x})][\tilde{P}_2(\tilde{x})/\tilde{P}_1(\tilde{x})]^{1-\alpha} - 1 \right\} \\ &\leq (2^{\alpha-1} - 1)^{-1} \left\{ \int_{\{f_2(x)>0\}} f_1(x) [f_2(x)/f_1(x)]^{1-\alpha} d\gamma(x) - 1 \right\} \\ &= I_\alpha^*(P_1 | P_2; \boldsymbol{X}). \end{aligned}$$

In the same way, we can verify that $I_\alpha^*(P_2 | P_1; \boldsymbol{X}) \geq I_\alpha^*(P_2 | P_1; \tilde{\boldsymbol{X}})$.

For $0 < \alpha < 1$ we can follow similar arguments to the preceding ones by taking into account the change in the sense of the concavity and the change in the sign of the constant $(2^{\alpha-1} - 1)^{-1}$.

In virtue of the Jensen inequality, the equality is clearly obtained when $f_1(x)/f_2(x) = \tilde{P}_1(\tilde{x})/\tilde{P}_2(\tilde{x})$, a.e. $[\gamma]$ (that is, a.s. $[P_1]$ and $[P_2]$) in $\operatorname{supp} \tilde{x}$. □

Assume that a simple random sample of size m from an experiment $\boldsymbol{X} = (X, \beta_X, P)$, $P \in \boldsymbol{P}$, is considered as informative enough to draw posterior inferences about the experimental distribution. However, if the observer cannot exactly perceive the experimental outcomes a loss of information (in the sense of the directed divergence of order α) arises.

We are now going to verify that this loss of information can frequently be removed by increasing appropriately the sample size from m to n, so that if $\tilde{\boldsymbol{X}}$ is the resulting f.i.s. $I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}^{(n)}) \geq I_\alpha^*(P_1 | P_2; \boldsymbol{X}^{(m)})$ for all $P_1, P_2 \in \boldsymbol{P}$.

In order to determine such a value n we must previously compute the information in the random sample $\boldsymbol{X}^{(m)}$ and the information in the fuzzy random sample $\tilde{\boldsymbol{X}}^{(n)}$. This last computation is not easy to accomplish with any measure of information (for instance, it is in fact unmanageable to find the Shannon amount of information contained in a fuzzy random sample because of the lack of exact relations connecting the amount in $\tilde{\boldsymbol{X}}^{(n)}$ with that in the f.i.s. $\tilde{\boldsymbol{X}}$ and the sample size). Nevertheless, the measures in Definition 2.2 satisfy the following property that makes easy to calculate the information in a fuzzy random sample.

THEOREM 2.2. *Let $\tilde{\boldsymbol{X}}^{(n)}$ be a fuzzy random sample of size n from an f.i.s. $\tilde{\boldsymbol{X}}$ and associated with a simple random sample $\boldsymbol{X}^{(n)}$. Then,*

$$I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}^{(n)}) = (2^{\alpha-1} - 1)^{-1}\{[(2^{\alpha-1} - 1) I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}) + 1]^n - 1\},$$

for all $P_1, P_2 \in \boldsymbol{P}$.

PROOF. Indeed, under the assumed hypotheses

$$\tilde{P}_i(\tilde{x}_1, \ldots, \tilde{x}_n) = \tilde{P}_i(\tilde{x}_1) \cdots \tilde{P}_i(\tilde{x}_n), \qquad i = 1, 2,$$

hence,

$$I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}^{(n)}) = (2^{\alpha-1} - 1)^{-1} \left\{ \left[\sum_{\substack{\tilde{x} \in \tilde{\boldsymbol{X}} \\ \tilde{P}_2(\tilde{x}) > 0}} [\tilde{P}_1(\tilde{x})]^\alpha [\tilde{P}_2(\tilde{x})]^{1-\alpha} \right]^n - 1 \right\}. \qquad \square$$

On the basis of Theorem 2.2, we can try to look for the minimum integer $n(\alpha)$, if it exists, with $n(\alpha) \geq m \log [(2^{\alpha-1} - 1) I_\alpha^*(P_1 | P_2; \boldsymbol{X}) + 1] / \log [(2^{\alpha-1} - 1) I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}) + 1]$ for all $P_1, P_2 \in \boldsymbol{P}$. If such an integer exists, then $I_\alpha^*(P_1 | P_2; \tilde{\boldsymbol{X}}^{(n(\alpha))}) \geq I_\alpha^*(P_1 | P_2; \boldsymbol{X}^{(m)})$, that is, a fuzzy random sample $\tilde{\boldsymbol{X}}^{(n(\alpha))}$ resulting from $n(\alpha)$ independent performances of $\boldsymbol{X}$ can be regarded as informative at least (in the sense of the directed divergence of order α) as a simple random sample of size m from $\boldsymbol{X}$. Therefore, if $\boldsymbol{X}^{(m)}$ were considered as informative enough to draw posterior conclusions about $\boldsymbol{X}$, $\tilde{\boldsymbol{X}}^{(n(\alpha))}$ may also be considered as informative enough for the same purposes.

We now examine two examples illustrating the preceding ideas:

Example 2.1. An investigator is interested in the control of a certain microorganism. He intends to prepare slides after treatment by a chemical and then count the organisms per square centimeter (random experiment X). Nevertheless, the treatment does not permit him to identify with sharpness the presence of a microorganism and, consequently, he cannot establish the exact number of microorganisms per square centimeter, but rather he can only perceive one of the following observations: $\tilde{x}_1$ = "a very small number of microorganisms are found", $\tilde{x}_2$ = "a moderate number of microorganisms are found" and $\tilde{x}_3$ = "a great number of microorganisms are found", that the investigator describes by means of the membership functions in Fig. 3 (Obviously, we may construct an f.i.s. $\tilde{\boldsymbol{X}} = \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4\}$, on X = {nonnegative integers}, where $\mu_{\tilde{x}_4}(x) = 1$ if $x = 21, \ldots, = 0$ otherwise).

Assume that the number of microorganisms per square centimeter has a Poisson distribution with mean 5 or 10.

Then, if P_1 represents the Poisson distribution with mean 5 and P_2 represents the Poisson distribution with mean 10, we have

$$P_1(x) = e^{-5} 5^x / x!, \qquad P_2(x) = e^{-10} 10^x / x! \ ,$$

hence, the induced probability distributions on $\tilde{\boldsymbol{X}}$ are given by

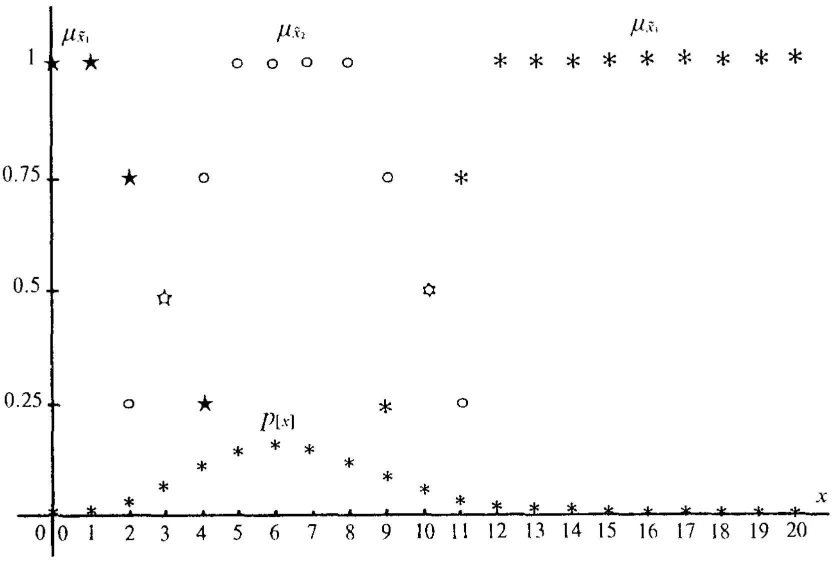

Fig. 3. Membership functions of "a very small number of mic. are found" ($\tilde{x}_1, \bigstar$), "a moderate number of mic. are found" ($\tilde{x}_2, \circ$), "a great number of mic. are found" ($\tilde{x}_3, \ast$), and Poisson probability function ($p, *$).

$$\tilde{P}_1(\tilde{x}_1) = 0.2176483, \qquad \tilde{P}_1(\tilde{x}_2) = 0.7525832\,,$$

$$\tilde{P}_1(\tilde{x}_3) = 0.0297682, \qquad \tilde{P}_1(\tilde{x}_4) = 0.0000003\,,$$

$$\tilde{P}_2(\tilde{x}_1) = 0.0107142, \qquad \tilde{P}_2(\tilde{x}_2) = 0.5069262\,,$$

$$\tilde{P}_2(\tilde{x}_3) = 0.4807695, \qquad \tilde{P}_2(\tilde{x}_4) = 0.0015901\,.$$

Therefore, the following are the directed divergences of order $\alpha = 2$ in $\tilde{X}$:

$$I_2^*(P_1|P_2;\tilde{X}) = 4.5404294, \qquad I_2^*(P_2|P_1;\tilde{X}) = 15.4399540\,,$$

whereas the directed divergences of order $\alpha = 2$ in X (Kagan's divergences) are given by

$$I_2^*(P_1|P_2;X) = 11.1824940, \qquad I_2^*(P_2|P_1;X) = 147.4131591\,.$$

Consequently, if a sample of size $m = 30$ from X were adequate enough to make posterior inferences about the true probability distribution of X, then the loss of information caused by the fuzziness in the available information could be removed by taking a fuzzy random sample of size $n(\alpha) = 54$, since 54 is the minimum integer higher than 30[log 148.4131591]/[log 16.4399540].

Example 2.2. A geologist is interested in analyzing the length of the largest axis of boulders in the upper reaches of a particular river (experiment X). The literature dealing with this subject asserts that for a half of the rivers in the country this length follows a normal distribution with mean 25 inches and standard deviation 10 inches, whereas for the other half this length follows a normal distribution with mean 30 inches and standard deviation 10 inches.

Assume that the geologist has not a mechanism of measurement sufficiently precise to determine exactly the length of the largest axis of boulders in the particular river. More precisely, suppose that the lack of roundness of these boulders only allows him to approximate the length of their largest axes by means of the following fuzzy observations: $\tilde{l}_1$ = "approximately lower than 10 inches", $\tilde{l}_2$ = "approximately 15 to 20 inches", $\tilde{l}_3$ = "approximately 25 inches", $\tilde{l}_4$ = "approximately 30 inches", $\tilde{l}_5$ = "approximately 35 to 40 inches" and $\tilde{l}_6$ = "approximately higher than 45 inches", which are characterized by the membership functions in Fig. 4 (Clearly, an f.i.s. $\tilde{X} = \{\tilde{l}_1, \ldots, \tilde{l}_7\}$ can be immediately constructed by defining $\mu_{\tilde{l}_7} = 1 - \Sigma_i \mu_{\tilde{l}_i}$, $i = 1, \ldots, 6$).

Then, if P_1 denotes the Normal distribution with mean 25 and

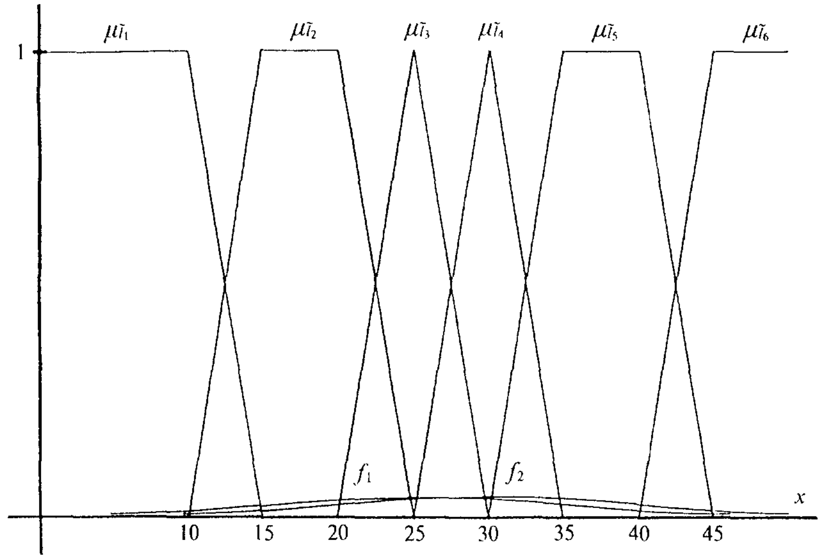

Fig. 4. Membership functions of $\tilde{l}_1$ = "length approximately lower than 10", $\tilde{l}_2$ = "approximately 15 to 20", $\tilde{l}_3$ = "approximately 25", $\tilde{l}_4$ = "approximately 30", $\tilde{l}_5$ = "approximately 35 to 40", $\tilde{l}_6$ = "approximately higher than 45", and normal density functions with means 25 and 30 and standard deviation 10.

standard deviation 10 and P_2 denotes the Normal distribution with mean 30 and standard deviation 10, the induced probability distributions on $\tilde{\boldsymbol{X}}$ are given by

$$\begin{aligned}
&\tilde{P}_1(\tilde{l}_1) = 0.1018, \quad \tilde{P}_1(\tilde{l}_2) = 0.2943, \quad \tilde{P}_1(\tilde{l}_3) = 0.1954\,,\\
&\tilde{P}_1(\tilde{l}_4) = 0.1733, \quad \tilde{P}_1(\tilde{l}_5) = 0.1873, \quad \tilde{P}_1(\tilde{l}_6) = 0.0416\,,\\
&\tilde{P}_2(\tilde{l}_1) = 0.0403, \quad \tilde{P}_2(\tilde{l}_2) = 0.1873, \quad \tilde{P}_2(\tilde{l}_3) = 0.1733\,,\\
&\tilde{P}_2(\tilde{l}_4) = 0.1954, \quad \tilde{P}_2(\tilde{l}_5) = 0.2943, \quad \tilde{P}_2(\tilde{l}_6) = 0.1080\,.
\end{aligned}$$

Therefore, the directed divergences of order $\alpha = 1/2$ in $\tilde{\boldsymbol{X}}$ are given by

$$I^*_{1/2}(P_1 | P_2; \tilde{\boldsymbol{X}}) = I^*_{1/2}(P_2 | P_1; \tilde{\boldsymbol{X}}) = 0.0674\,.$$

If the information from $\boldsymbol{X}$ were exact we could obtain the directed divergence of order $\alpha = 1/2$ in $\boldsymbol{X}$ (Matusita's measure of squared distance), where

$$I^*_{1/2}(P_1 | P_2; \boldsymbol{X}) = I^*_{1/2}(P_2 | P_1; \boldsymbol{X}) = 0.0732\,.$$

Consequently, if in order to draw posterior conclusions a sample of size $m = 65$ from $\boldsymbol{X}$ were originally necessary, the adequate size of the sample fuzzy information from $\tilde{X}$ in order to remove the loss of information caused by the fuzziness is $n(\alpha) = 71$, since 71 is the minimum integer higher than 65[log 1.0732]/[log 1.0674].

3. Alternative suggestions

Consider that the family of probability measures $\boldsymbol{P}$ on the experiment X can be expressed by $\{P_\theta, \theta \in \Theta\}$, θ being a numerical or vector-valued parameter, and the parameter space Θ being an interval in a Euclidean space.

According to Ferentinos and Papaioannou (1981), we could alternatively construct a family of *parametric measures of information* from the non-parametric ones in Definition 2.2 by using the following method (θ real)

$$I_\alpha^*(\theta;\tilde{X}) = \liminf_{\Delta\theta \to 0} [I_\alpha^*(P_\theta | P_{\theta+\Delta\theta};\tilde{X})]/(\Delta\theta)^2, \qquad \alpha \neq 1 .$$

It should be emphasized that, following ideas like those in Ferentinos and Papaioannou (1981), it can be verified that

$$I_\alpha^*(\theta;\tilde{X}) = [\alpha(\alpha-1)I^F(\theta;\tilde{X})]/[2(2^{\alpha-1}-1)]$$

(where $I^F(\theta;\tilde{X})$ is the Fisher information measure (1925) in the f.i.s. $\tilde{X}$, extended as measures in Definition 2.2).

On the other hand, when we wish to look for the suitable size of the sample fuzzy information from $\tilde{X}$ in order to eliminate the loss of information due to the fuzziness, we often find that it becomes practically impossible to determine a size suitable for all pairs of probability measures in $\boldsymbol{P}$ (or, for all $\theta \in \Theta$, in accordance with the alternative plan). This inconvenience could be avoided by means of the knowledge of the prior probability distribution on the family $\boldsymbol{P}$ (or, on the parameter space Θ). Thus, with such a knowledge the *expected directed divergence of order α with respect to the prior distribution* may be defined and used with the purposes in Section 2.

4. Concluding remarks

The study in this paper could also be developed for the extension of the Kullback-Leibler measure ($\alpha \to 1$) and the gain of information of order α of Rényi (1961). Obviously, this last study would lead to results equivalent to those herein examined. As we have just remarked, it is not operative to accomplish this kind of analysis for Shannon's amount of information, because of the lack of exact relations between the amounts corresponding to $\tilde{X}^{(n)}$ and $\tilde{X}$.

On the other hand, the notions in Section 2 could be used to define a criterion like that in Ferentinos and Papaioannou (1982), and the following ideas in our previous papers (Gil *et al.* (1984, 1985*b*) and Gil (1988)), to compare fuzzy information systems.

Finally, the results we have just achieved may be directly applied to the problem of the loss of information due to grouping of observations (see Kale (1964) and Ferentinos and Papaioannou (1979)), since a class of exact observations can be considered as a special example of fuzzy information (with the membership function given by the indicator function of that class).

REFERENCES

Casals, M. R., Gil, M. A. and Gil, P. (1986). The fuzzy decision problem: an approach to the problem of testing statistical hypotheses with fuzzy information, *European J. Oper. Res.*, **27**, 371–382.

Ferentinos, K. and Papaioannou, T. (1979). Loss of information due to groupings, *Trans. 8th Prague Conf. on Inf. Theo., Stat. Dec. Func. and Ran. Proc.*, C, 87–94.

Ferentinos, K. and Papaioannou, T. (1981). New parametric measures of information, *Inform. and Control*, **51**, 193–208.

Ferentinos, K. and Papaioannou, T. (1982). Information in experiments and sufficiency, *J. Statist. Plann. Inference*, **6**, 309–317.

Fisher, R. A. (1925). Theory of statistical estimation, *Proc. Cambridge Philos. Soc.*, **22**, 700–725.

Gil, M. A. (1988). Probabilistic-possibilistic approach to some statistical problems with fuzzy experimental observations, *Combining Fuzzy Imprecision and Probabilistic Uncertainty in Decision-Making*, (eds. J. Kacprzyk and M. Fedrizzi), pp. 286–306, Lecture Notes in Economics, No. 310, Springer-Verlag, Berlin.

Gil, M. A., López, M. T. and Gil, P. (1984). Comparison between fuzzy information systems, *Kybernetes*, **13**, 245–251.

Gil, M. A., Corral, N. and Gil, P. (1985*a*). The fuzzy decision problem: an approach to the point estimation problem with fuzzy information, *European J. Oper. Res.*, **22**, 26–34.

Gil, M. A., López, M. T. and Gil, P. (1985*b*). Quantity of information; comparison between information systems: 1. Nonfuzzy states, 2. Fuzzy states, *Fuzzy Sets and Systems*, **15**, 65–78, 129–145.

Kagan, A. M. (1963). On the theory of Fisher's amount of information, *Soviet Math. Dokl.*, **4**, 991–993.

Kale, K. (1964). A note on the loss of information due to grouping of observations, *Biometrika*, **51**, 495–497.

Kullback, S. and Leibler, A. (1951). On information and sufficiency, *Ann. Math. Statist.*, **22**, 79–86.

Mathai, A. M. and Rathie, P. N. (1975). *Basic Concepts in Information Theory and Statistics*, Wiley, New Delhi.

Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications, *Ann. Inst. Statist. Math.*, **19**, 181–192.

Okuda, T., Tanaka, H. and Asai, K. (1978). A formulation of fuzzy decision problems with fuzzy information, using probability measures of fuzzy events, *Inform. and Control*, **38**, 135–147.

Rathie, P. N. (1973). Some characterization theorems for generalized measures of uncertainty and information, *Metrika*, **20**, 122–130.

Rényi, A. (1961). On measures of entropy and information, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 547–561, Univ. California Press, Berkeley.

Tanaka, H., Okuda, T. and Asai, K. (1979). Fuzzy information and decision in statistical model, *Advances in Fuzzy Sets Theory and Applications*, 303–320, North-Holland, Amsterdam.

Zadeh, L. A. (1965). Fuzzy sets, *Inform. and Control*, **8**, 338–353.

Zadeh, L. A. (1968). Probability measures of fuzzy events, *J. Math. Anal. Appl.*, **23**, 421–427.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems*, **1**, 3–28.